Applied Language Sciences

RESEARCH Open Access

Using generative AI to revise articles: How much is too much?

Richard Watson Todd[1]*

*Correspondence:
richard.tod@kmutt.ac.th

[1] School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand

Abstract

The use of generative AI (GenAI) to aid in the writing of research articles is potentially highly beneficial, yet the practice runs the risk of producing texts that are not the author's original contribution. Based on distributed cognition, this research focuses on using GenAI prompts of different strengths to revise an article. Eight extracts of early drafts of human-written articles were revised using four prompts ranging from light editing to heavy editing. The resulting texts were evaluated on five criteria. Three criteria showed similar patterns (tagging by an AI detector, proportion of original wording retained, and use of rare words) with the lightest edited texts being around 90% original contribution, and the heaviest edited texts around 20% original contribution. Offloading the processes of revising articles to GenAI can result in loss of authorial ownership. The implications of the findings for journals and for GenAI training for researchers are discussed.

Keywords: Generative AI, research articles, editing, revising, distributed cognition, authorial ownership

1. Introduction

A key principle of research integrity is that authors must take full responsibility for the content of their publications (Krumholz, 2025). For writing up research, traditionally a major concern was plagiarism where authors would engage in imposture with a potential risk of being found out. Since the advent of generative AI (GenAI) and its massive explosion in use, imposture is more likely to involve representing GenAI text, rather than the text of other authors, as one's own. There are now numerous grey areas in what is acceptable practice. Can AI-generated text be included in an article as long as the author has curated it? How much GenAI editing of a text is acceptable? Does GenAI content generation imply a lack of author responsibility? At the same time the extent to which GenAI is used in writing an article is very difficult to discover.

The changes brought about by the rise of GenAI have impacts on the training of research writing. Increasingly, courses in English for research publication purposes (ERPP) are addressing the potential roles of GenAI in the writing process and attempting to balance the benefits and dangers. For example, the course on the use of GenAI in research writing by Nature Masterclasses (https://www.nature.com/masterclasses/generative-ai-in-publishing/51470194) highlights the supposed improvements in productivity while emphasising the need for research integrity and responsibility. The use of GenAI in research writing is complex, context- and discipline-specific, and in constant flux as new tools and new versions become available. Consequently, most GenAI in ERPP courses rely on interacting with GenAI tools and reflecting on these experiences (e.g., Gabriel, 2024; Ou et al., 2024). Yet, studies into AI literacy for researchers (e.g., Parker & Becker, 2026) show that researchers consistently ask for guidance in areas such as evaluating GenAI output for bias, ensuring accountability and maintaining authorial voice.

The issue of authorial voice is particularly important as GenAI revisions to a text, when extensive, can result in substantial stylistic and argumentative changes, leading to a loss of authorial voice with a consequent impact on feelings of authorial ownership (Altalib, 2026). Despite such potential consequences, using GenAI to edit or revise texts is the most widely used application of GenAI in the research process (Andersen et al., 2025) and is also broadly accepted. In a survey, nearly 80% of PhD students viewed light editing, akin to proofreading, as appropriate (Wu et al., 2026). For authors whose first language is not English (and even for English native speakers with little experience in academic writing, Lei & Yang, 2020), being able to use GenAI to edit their drafts to meet the expected standards is empowering, which, to some extent, reduces the inequalities inherent in an English-dominated environment (Mohebbi, 2026). However, 70% of the PhD students rated extensive revisions as inappropriate (Wu et al., 2026), presumably because allowing GenAI free rein in revising a text "inevitably results in a complete rewrite that leaves little trace of the author's original style" (Baron, 2024, p. 80). Indeed, text editing can blur into text generation, and "the content generated by NLP systems [i.e., GenAI] is not the author's original content" (Tang, 2024, p. 1242). With light editing or supportive GenAI use, writers can feel that their voice has, at least partially, been retained and so feel some sense of ownership, perceptions which do not necessarily hold for heavy editing (Altalib, 2026). While there is fairly wide agreement on the appropriacy of light editing and the inappropriacy of heavy editing (Cheng et al., 2025), there is surprisingly little research into

how the wording of prompts affects the level of editing. There is, therefore, a need for research investigating how different GenAI prompts change texts and affect authorial voice and ownership.

The use of GenAI as a tool to revise writing can be viewed from the perspective of distributed cognition, which argues that cognitive processes are not isolated in an individual but are distributed in an ecology. There are two main types of distributed cognition (Hutchins, 2000; Salomon, 1997). Shared cognition is where interaction with other people is key to the cognitive processes; cognitive offloading is where external structures, such as materials and tools, become part of the cognition. For writing, traditionally some cognitive processes would be offloaded to pen and paper. For example, when planning, writers would not try to organise all of their thinking in their brain, but would offload aspects of memory to written notes. GenAI is a powerful tool for offloading cognitive processes, such as structuring an argument, checking for inconsistencies, and revising texts. In using GenAI in these ways, it becomes a collaborator where human-GenAI interaction is viewed as a coupled cognitive system (Zhao & Han, 2026). In any collaboration, the collaborators take on different roles and have different levels of input into the final product. From a distributed cognition perspective, human-GenAI interaction falls on a continuum of human agency with different proportions of human and GenAI contributions to the final product. As greater proportions of cognitive processes are offloaded to GenAI, issues of authorial ownership, a concern in research integrity, become apparent. This paper views human-GenAI interaction as distributed cognition, where editing or revising a draft of a research article is offloaded to GenAI. The study investigates how prompts of different strengths affect the extent to which the revised version can be considered the author's original contribution.

1.1 GenAI use to support writing

Nearly all journals now require declarations concerning how AI was used in the research, and often these include guides for how GenAI can and cannot be used to help in writing the article. From recent surveys (Charbonneau & Zhang, 2026; Ganjavi et al., 2024), only 1% of journals explicitly prohibited the use of GenAI in any form, and the majority permitted the use of GenAI to improve the language of the manuscript. These guidelines are policed through self-regulation and self-disclosure (Zhaksylyk et al., 2023). Examples include:

> Authors preparing a manuscript for this journal can use AI Tools to support them … Ultimately, authors are responsible and accountable for the contents of their work. This includes accountability for: … Editing and adapting all material thoroughly to ensure the manuscript represents the author's authentic and original contribution and reflects their own analysis, interpretation, insights and ideas. (Elsevier, 2026, https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals)

> We distinguish various uses for AI and related technologies as: assistive (and no longer requiring disclosure), generative (requiring disclosure), and prohibitive … AI tools that make suggestions to improve or enhance your own work, such as tools to improve language, grammar, or structure, are considered

> assistive AI tools and do not require disclosure by authors or reviewers … Inappropriate use of GenAI includes: Generation of incorrect text or content. (Sage Publications, 2026, https://www.sagepub.com/journals/publication-ethics-policies/artificial-intelligence-policy)

Under the Elsevier guidelines, cases where GenAI has been used to either generate or heavily revise text without the product being curated by the author(s) would be prohibited. Supposedly, such cases can be found by searching for phrases such as "As an AI language model" (Gulumbe, 2024). There are 1,860 instances of this phrase returned in Google Scholar but none of the top 100 returns show uncurated GenAI output (nearly all concern interacting with GenAI). Other phrases, such as "If you want, I can also" (the standard opening to the final paragraph in a ChatGPT response), while rarer, provide clearer evidence of a lack of curation of GenAI output. Of the 118 returns in Google Scholar, 108 appear to be blatant copying from GenAI. For example:

> Accordingly, the projection is used to motivate discussion of design safeguards (e.g., risk-based pricing, portfolio caps, and monitoring discipline) under conditions of potential scheme expansion. If you want, I can also produce a 2.3. Variable Definitions and Measurement subsection (top-journal standard) that defines "default", "growth", and "coverage" precisely and makes the paper more "reviewer-proof". (Xhafa, 2025, p. 189)

Most researchers, however, are not so careless, but authors who generate whole sections through GenAI are unlikely to pay much attention to the journal guidelines and would probably not self-declare such use of GenAI.

The publishers appear to be aiming at authors with at least some integrity, for example, those who write a rough first draft before turning to GenAI for writing support, and these authors are also the target group for GenAI in ERPP courses. This language mediation is one of the three main uses of GenAI to aid writing by university students (Khanh-Quynh & Cong-Lem, 2026). Using GenAI as an assistive tool for light editing of text is clearly beneficial for academia, and is supported in the Sage guidelines by not requiring disclosure. However, GenAI can also be used to revise text to such an extent that the finished product does not represent the author's authentic and original contribution.

In using GenAI to edit texts, the wording of the prompt affects the extent to which the revised text is either lightly edited (retaining authorial ownership) or AI-generated. This research uses five criteria to measure the effects of the prompts on the revised texts:

1. The extent to which the revised text is tagged as human-authored or AI-generated by an AI detector.
2. The length of the revised text.
3. The percentage of original wording retained in the revised text.
4. The use of rare words in the revised text.
5. The presence of noun + preposition phrases in the revised text.

Using these criteria, the study aims to answer the research question: How do editing prompts of different strengths affect the authorial ownership of a text? Conducting such an analysis should allow us to make suggestions about the borderline between acceptable text editing and unacceptable text generation with implications for the design of GenAI in ERPP courses.

2. Research methodology

2.1 The data

Starting with eight human-authored texts, each text was revised using four different GenAI prompts to produce 40 texts for analysis (8 original and 32 revised). Each of the 40 texts was analyzed for the five criteria to provide evidence of the extent to which the text could be considered the author's original contribution.

2.1.1 The original texts

The eight original texts consisted of extracts several paragraphs long taken from the literature reviews of early drafts of research articles. All were written by PhD students studying applied linguistics at a Thai university whose first language was not English. The research reported generally falls within the areas of corpus linguistics and discourse analysis. All drafts were written before 2022 and so could be guaranteed to be human-authored. The texts all fall under broad consent allowing work completed as part of study to be used for research. The texts range from 299 to 449 words in length.

2.1.2 The revised texts

Assuming that authors using GenAI to revise their articles are not AI specialists, the free version of ChatGPT (GPT 5.3 in March 2026), the most widely used GenAI application, was used in this study. Each of the eight texts was entered into ChatGPT four times to be revised following four prompts. Each use was started as a new conversation.

The first two prompts assume a low level of AI literacy and are the most basic prompts possible. To see if the directions edit and revise produce different responses, these prompts are:

1.Please edit the following text:

2.Please revise the following text:

These are termed 'Plain Edit' and 'Plain Revise' respectively. Intuitively, and based on advice for student writers (e.g. https://slc.berkeley.edu/writing-worksheets-and-other-writing-resources/editing-vs-revision), editing focuses on the surface features of text at the sentence level, whereas revising involves more holistic features such as argumentation. We may therefore expect Plain Revise texts to have made more changes to the original text than Plain Edit.

The third prompt reflects both the spirit and the wording of the journal AI use guidelines, aiming to limit changes to surface features and retain as much of the original author-written text as possible. This prompt mirrors the human written, light AI editing condition of Cheng et al. (2025) with little processing offloaded to GenAI. Termed 'Full Edit', the prompt is:

3.Please edit the following text retaining as much of the original content and wording as possible with the only changes being those related to infelicitous phrasing.

The fourth prompt may follow the wording of the journal AI use guidelines but does not manifest the spirit. Termed 'Full Revise', the prompt aims to maximise the likelihood of the article being accepted while paying lip-service to research integrity. This prompt mirrors the human written, heavy AI editing condition of Cheng et al. (2025) where large proportions of processing have been offloaded to GenAI with potential implications for authorial ownership. The prompt is:

4.Please revise the following text to make it seem as if it were written by a leading

researcher for a high-level academic journal.

To illustrate how these prompts change the text, a short paragraph from one of the original texts is provided together with the four revised texts. Using the side-by-side text comparison tool, https://www.diffchecker.com/, words changed in the revised texts are highlighted in bold.

Original text

ESP genre analysis appears to involve two broad approaches: first, a genre-based approach involving the identification of a genre by investigating the operation of texts within a context in terms of communicative purposes, and second, identification of linguistic features which commonly occur within texts sharing communicative purposes (Bruce, 2008). Among ESP researchers and theorists, communicative purpose is the primary criterion for the identification of genres. This view of genre analysis is based on the principle that several texts can be categorised into a single genre by their shared communicative purpose.

1. Plain Edit text

ESP genre analysis generally encompasses two broad approaches. The first is a genre-based approach, which involves identifying a genre by examining how texts operate within a context, particularly in terms of their communicative purposes. The second approach focuses on identifying linguistic features that commonly occur in texts sharing similar communicative purposes (Bruce, 2008). Among ESP researchers and theorists, communicative purpose is regarded as the primary criterion for genre identification. This perspective is grounded in the principle that multiple texts can be categorised within a single genre based on their shared communicative purpose.

2. Plain Revise text

ESP genre analysis is generally understood to encompass two broad approaches. The first is a genre-based approach, which involves identifying a genre by examining how texts operate within a particular context, especially in terms of their communicative purposes. The second focuses on identifying the linguistic features that commonly occur in texts sharing similar communicative purposes (Bruce, 2008). Within ESP research, communicative purpose is widely regarded as the primary criterion for genre identification. This perspective is grounded in the principle that texts can be grouped into a single genre based on their shared communicative purposes.

3. Full Edit text

ESP genre analysis appears to involve two broad approaches: first, a genre-based approach involving the identification of a genre by investigating how texts operate within a context in terms of communicative purposes; and second, the identification of linguistic features that commonly occur within texts sharing communicative purposes (Bruce, 2008). Among ESP researchers and theorists, communicative purpose is the primary criterion for the identification of genres. This view of genre analysis is based on the principle that several texts can be categorised into a single genre based on their shared communicative purpose.

4. Full Revise text

ESP genre analysis has conventionally been conceptualised in terms of two complementary, though analytically distinct, orientations. The first adopts a genre-based perspective, in which genres are identified through the examination of how texts function within specific socio-rhetorical contexts, particularly with respect to their communicative purposes. The second foregrounds the recurrent linguistic features that typify texts sharing such purposes (Bruce, 2008). Within ESP scholarship, communicative purpose is widely regarded as the primary criterion for genre identification, reflecting the assumption that

texts may be grouped into a genre insofar as they realise a shared set of communicative intentions.

2.2 Data analysis

The 40 texts (5 versions of 8 texts) were evaluated based on 5 criteria. The results show the findings for the 8 texts combined.

Criterion 1: Proportion of AI-generated text

The increased use of GenAI to aid writing has led to the development of numerous online AI detection tools. While their reliability varies, high-quality AI detectors generally perform better at distinguishing between human-authored and AI-generated text than human informants (Cheng et al., 2025; Elkhatat et al., 2023). Seven online detectors were piloted using 4 original texts and 4 Full Revise texts. Only one of the detectors consistently identified the original texts as 100% human-written and the Full Revise texts as at least 50% AI-generated.

The Quillbot AI detector v7.1.0 (https://quillbot.com/ai-content-detector) calculates three values for any inputted text: % of AI-generated text, % of human-written and AI-refined text, and % of human-written text. After inputting all 40 texts, only two returned values for human-written and AI-refined text; these percentages were therefore added to the human-written percentages for those two texts.

Criterion 2: Length of text

The length of each of the 40 texts was calculated using the Word Count function in Microsoft Word.

Criterion 3: Proportion of original wording retained

Using the diffchecker side-by-side comparison tool (https://www.diffchecker.com/), each of the 32 revised texts were compared against their respective original text. The number of words retained from the original text in the revised text (normal font in the examples above) and the number of changed words (bold font in the examples) were counted and calculated as percentages of the total number of words in the revised text.

Criterion 4: Use of rare words

GenAI output has a reputation for including a higher than usual frequency of certain relatively rare words, such as *delve* and *underscore* (Hirosawa et al., 2025; Juzek & Ward, 2025). These are well-known examples of a tendency for GenAI output to include a relatively high proportion of words outside the top 5,000 most frequent words in English (Watson Todd, 2024), which, in turn, is a symptom of the formal nature of GenAI-generated texts (Rodrigues et al., 2026). To see if the revised texts included a higher frequency of rare words than the original texts, all 40 texts were analyzed using the Vocabulary Profiler at https://www.eapfoundation.com/vocab/profiler/. BNC/COCA was used as the comparative database since this generates the frequencies of the number of words in a text at each of the top 1,000 word ranges up to 26,000. The most frequent 1,000 words are termed 1K words; the next most frequent 1,000 words (i.e., words ranked 1,001 to 2,000) are termed 2K words, and so on. The frequencies of word types for the ranges 3-5K (to account for formal language, such as the replacement of phrasal verbs with Latinate verbs), 6-10K (which includes *delve* and *underscore*), and 11-25K (for rarer words) were calculated and represented as percentages of the number of words in each text.

Criterion 5: Frequency of noun + preposition phrases

One characteristic of academic writing is frequent use of modified noun phrases, such as those post-modified by prepositional phrases (Parkinson & Musgrave, 2014). When writing academic texts, GenAI may overuse features distinguishing the register from other uses of language. Jiang and Hyland (2025) investigated the frequency of types of 3-word bundles in human-authored and AI-generated academic texts, finding that the type most

associated with GenAI texts was noun + preposition bundles (e.g. *the potential for*). To examine this, corpora were compiled for each of the 5 types of text and these were tagged for parts of speech using the Multidimensional Analysis Tagger at https://sites.google.com/site/multidimensionaltagger (Nini, 2019). Using AntConc 3.5.9 (Anthony, 2022), the frequencies of first, noun + *of* phrases, and, second, noun + preposition phrases were counted for each corpus and the strength of collocation measured using T-scores (where higher scores indicate stronger associations).

3. Results

To present the findings, the results from the 8 texts are collated allowing the different versions of the texts to be compared using the five criteria.

3.1 Proportion of AI-generated text

Table 1 presents the average percentages assigned to humans and GenAI in the texts using the Quillbot AI detector, together with minimum and maximum values.

Table 1. Proportions of human and AI text in the five versions

	Human-written				AI-generated			
Text version	Average	SD	Max	Min	Average	SD	Max	Min
Original	100.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0
Plain Edit	67.4	16.7	91.0	43.0	32.6	16.7	57.0	9.0
Plain Revise	50.9	11.8	67.0	31.0	49.1	11.8	69.0	33.0
Full Edit	94.5	12.3	100.0	65.0	5.5	12.3	35.0	0.0
Full Revise	23.3	21.4	57.0	0.0	76.8	21.4	100.0	43.0

Table 1 shows that the Full Edit texts are very similar to the original text in the level of human-written content, while the Full Revise texts are difficult to distinguish from fully AI-generated texts. In other words, the Full Edit texts appear to retain authorial voice and ownership, whereas, in the Full revise texts, cognitive control over the text has been relinquished to GenAI. The use of 'Edit' in the prompt retains more of the human-written elements than using 'Revise', but the difference is not clear.

3.2 Length of texts

Table 2 shows the average lengths of the five versions of the texts.

Table 2. Lengths of the five versions of the texts

Text version	Average	SD	Max	Min
Original	384	49.6	449	299
Plain Edit	354	28.5	400	317
Plain Revise	359	34.5	404	319
Full Edit	377	45.5	436	295
Full Revise	437	73.7	555	356

The Full Revise texts are on average roughly 50 words longer than the original texts suggesting that GenAI may have added additional content to the original. In this case, offloading processes to GenAI has impacted authorial ownership of the finished product.

Such additions can be seen in the first sentence of the example paragraphs where the Full Revise text adds information that the two approaches are "complementary, though analytically distinct". The other versions are generally, but not markedly, shorter than the original texts.

3.3 Proportion of original wording retained

Table 3 shows the percentages of the wording in the original text which are retained in the revised texts.

Table 3. Proportions of original wording retained in the revised texts

Text version	Average	SD	Max	Min
Plain Edit	56.7	6.0	65.0	47.9
Plain Revise	47.3	8.7	60.0	32.2
Full Edit	83.6	5.3	94.1	77.1
Full Revise	17.7	5.8	28.9	11.5

From Table 3, the majority of the Full Edit texts are the same as the original texts suggesting that they reflect the author's authentic and original contribution and maintain authorial voice. Both the Plain Edit and the Plain Revise texts retain around half of the original texts with the Plain Edit texts retaining slightly, but not markedly, more than the Plain Revise texts. These can probably be considered borderline cases in terms of the extent to which they reflect the original contribution. The Full Revise texts retain very little of the original with GenAI contributing the majority of the final text. It is difficult to see how an author with integrity could claim authorial ownership and self-report such texts as being their authentic contribution.

3.4 Use of rare words

Table 4 shows the percentages of word types in the 3-5K, 6-10 K and 11-25K ranges for the five versions of the texts.

Table 4. Percentages of rare words in the five versions of the texts

	3-5K words				6-10K words				11-25K words			
Version	Av	SD	Max	Min	Av	SD	Max	Min	Av	SD	Max	Min
Original	18.6	3.4	26.0	15.7	3.0	1.3	4.5	1.0	0.7	0.9	2.5	0.0
P Edit	22.7	3.7	29.0	17.9	2.8	1.2	4.7	1.0	0.8	0.7	2.1	0.0
P Revise	22.6	4.5	31.4	18.1	2.9	1.0	3.8	1.2	0.8	0.7	2.1	0.0
F Edit	19.5	3.2	26.1	15.2	2.9	1.3	4.5	0.9	0.7	0.9	2.5	0.0
F Revise	30.0	3.1	33.4	23.9	4.7	1.3	7.4	3.3	1.4	0.9	2.9	0.4

To illustrate what these numbers mean, for the original text from which the paragraph used as an example was extracted, there are 36 word types in the 3-5K range representing 19% of all word types (e.g., academic, analysis, classification, genre) 2 word types (1%) in the 6-10K range (grammatical, lexical), and none in the 11-25K range.

The differences between the original text and the first three revised texts are negligible with a small rise in the proportions of 3-5K word types suggesting a slight increase in the level of formality (e.g., generates in the original is replaced by activates, and in terms of is replaced by in the domain of). The 6-25K word types in the original texts are mostly technical terms which are retained in the first three revised texts with no additional rare words. The Full Revise texts are noticeably different to the others with sharp increases in the use of less frequent words across all levels. The increase in the use of 3-5K word types indicates a sharp rise in the level of formality. The rarer word types in the Full Revise texts include some technical terms (e.g., anaphoric, psycholinguistic, semiotic), some words broadly associated with academic writing (e.g., discursive, longitudinal, operationalize), and some infrequent replacements of common concepts (e.g. elucidate, pertains, recasting). Taken together, these additions mean that the Full Revise texts are written in a style very different to that of the author's original texts resulting in a loss of authorial voice.

3.5 Frequency of noun + preposition phrases

Table 5 shows the frequencies and strengths of collocation of two forms of postmodified noun phrases.

Table 5. Frequency of postmodified noun phrases

	Noun + of		Noun + preposition	
Text version	Frequency	T-score	Frequency	T-score
Original	77	7.7	180	10.9
Plain Edit	49	6.0	125	8.7
Plain Revise	56	6.5	139	9.3
Full Edit	63	6.8	165	10.3
Full Revise	78	7.7	183	10.8

From Table 5, the original texts and the Full Revise texts make similar use of postmodified noun phrases, while these noun phrases are less common in the other three types of text. These findings suggest that, contrary to Jiang and Hyland (2025), for the data in this study postmodified noun phrases are not indicators of GenAI-written texts.

4. Discussion

Of the five criteria used to analyze texts in this study, three show a similar pattern across types of text, while two do not clearly distinguish between the texts. Given the minor differences in length of text and frequency of noun + preposition phrases, these criteria will not be considered further, and the discussion will focus on proportion of AI-generated text, proportion of original wording retained, and use of rare words.

The five types of text can be positioned on a continuum from human-written to AI-generated as shown in Figure 1 with similar placements on three criteria (proportion of AI-generated text, proportion of original wording retained, and proportion of 11-25K words). Generally, on these criteria, Full Edit texts are very similar to the original human-written texts, while Full Revise texts are close to AI-generated texts. The Plain Edit and Plain Revise fall in the middle.

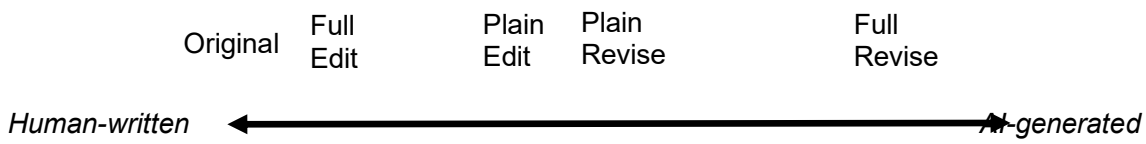

Figure 1 Levels of humanness in five types of text

Figure 1 supports the idea of a continuum of levels of collaboration between authors and GenAI as posited by distributed cognition theory. From left to right, cognitive processes are increasingly offloaded to GenAI, the extent to which the final text represents GenAI contributions increases, and the levels of human agency and control fall. This continuum is similar to a categorisation of GenAI users proposed by Poquet et al. (2026). Analysing GenAI use by students completing a reading synthesis task in an experimental setting, they found five types of GenAI users: no-offloading, clarifying, strategic offloading, comprehension offloading, and full offloading. These types differed in the extent to which comprehension and synthesis processes were offloaded to GenAI. Clarifying and strategic offloading primarily asked for GenAI support at the local level, whereas comprehension and full offloading applied to whole texts (such as asking for an easy-to-read summary) with little or no evaluation of the GenAI output. There are parallels between Poquet et al.'s (2026) user types and the continuum in Figure 1. The original human-written text is equivalent to no-offloading; Full Edit is similar to clarifying; Plain Edit and Plain Revise mirror strategic offloading; and Full Revise is akin to comprehension offloading. A fully GenAI-generated text would be full offloading. Based on these close parallels, from a distributed cognition perspective, the use of GenAI should not be seen as a dichotomy (whether to use GenAI or not), but as a continuum of human agency, proportions of contributions, and authorial ownership.

5. Implications

5.1 Study Contributions

Tang (2024) argues for the need for journals to develop policies about how much GenAI-generated content is acceptable, since "there remains a lack of clarity about the limits of permissible use" (Mohebbi, 2026, p. 225). Some journals already use AI detection tools as a matter of course. For example, the Journal of Population and Social Studies states:

> JPSS employs AI detection tools to assess the presence of AI-generated content … If an AI detection report suggests substantial AI involvement beyond what has been disclosed, authors may be required to provide further clarification (https://so03.tci-thaijo.org/index.php/jpss/AI_Use_Policy)

In such cases, it is unclear what tool is being used, how it works, how reliable it is, and what cutoff is viewed as sufficient for editorial concern.

Choice of tool is crucial, since a couple of the detectors piloted in this study produced seemingly random results. Even for more reliable detectors, some AI detection tools use probabilistic models (e.g., Miralles-González et al., 2025), while others are based on text features, including lexical diversity (e.g., Yamashita & Meier, 2025), yet none of the providers of readily available tools document how they work. The technology is being continuously updated, and there is an arms race with tools available for humanising AI-

generated texts to beat AI detectors (Zhou et al., 2024), some of which may remove the rare words favoured by GenAI. While the more reliable AI detection tools can provide input for human decision making, it is very difficult to see how they can be relied on as the technology shifts.

If AI detection tools cannot be used, journals will have to rely on author self-disclosure, which comes with its own problems. First, it is easy to envisage cases where an author might submit a Full Revise text to a journal and, following the Sage journal guidelines, not declare use of GenAI since editing supposedly falls under assistive GenAI use. Yet the submitted article is far closer to an AI-generated text than it is to the original article. Claiming that a Full Revise text is still the author's authentic and original contribution is highly dubious. Second, despite many journals requiring AI use statements, disclosure rates are very low as only 4.2% of articles include any AI disclosure statement, with a third of these explicitly reporting non-use (Moorhouse et al., 2026). The remaining 2.8% include any form of AI use, including light editing, yet an analysis of conference papers showed that 7-15% of texts were "substantially modified by AI beyond a simple grammar check" (Liang et al., 2024, p. 10), in other words, changed at the level of at least Plain Edit in this study. Even more worryingly, any GenAI revisions to a text might not be curated by the author. Albeit in an inauthentic experimental setting, Noy and Zhang (2023) found that 33% of authors accepted AI output without any checking, while another 53% conducted minimal checks. The overall picture, then, is that self-disclosure does not appear to be working with very low disclosure rates and, even when GenAI use is reported, the extent of human oversight is unclear.

In contrast to the AI detection tools, the proportion of original wording retained in the revised text cannot be gamed and is unaffected by technological developments. This criterion has not been widely used in previous research even though it provides a reliable measure. While it is impractical for journals to ask authors to report how much of their text is their own original work, using a side-by-side text comparison tool provides a clear picture for authors to see the extent to which their text is affected by GenAI revisions.

The problems in identifying and reporting GenAI use in research writing risk leading to declines in academic integrity and reinforce calls for greater researcher training in the use of GenAI (Drechsler, 2026). While this study does not directly address GenAI ERPP training, the three criteria (proportion of AI-generated text, proportion of original wording retained, and proportion of rare words) which can be used to identify the extent of GenAI revisions to a text may prove to be useful in GenAI ERPP courses. A key goal in many of these courses is to raise authors' awareness of how the use of GenAI impacts their writing (Ou et al., 2024), and the three criteria provide easily comprehensible evidence for such impacts. In addition, all three use easily accessible and usable free online applications facilitating their incorporation into courses. While the reliability problems of AI detection tools need to be highlighted, they provide a straightforward intuitive summary of the extent of GenAI influence on a text. Being able to identify and change rare words added by GenAI to a text allows authors to partially regain their authorial voice. Comparing pre- and post-GenAI revision versions of a text highlights, in many cases, how little of the author's original contribution remains after GenAI changes, and, depending on the context, trainers may even suggest certain proportions of retained text as a rule of thumb to guide author

decision making. Applying the easily implemented methods used in this study in GenAI ERPP training may help authors gain clear understandings of how the use of GenAI as a revision tool can result in texts which are not the author's authentic original contribution.

6. Conclusion

The benefits of using GenAI for light editing of human-written texts are tangible, yet the research community runs the risk of drowning in a mass of relatively worthless AI-generated research studies (Hu, 2026). The ideal is for authors, especially those without English as a first language, to write original articles and use GenAI lightly to increase readability (as with the Full Edit texts). Doing this retains human agency and authorial ownership while collaborating beneficially with GenAI. However, without careful prompt design when using GenAI as a revision tool, a substantial proportion of cognitive processing can be offloaded to GenAI running the risk of generating texts that are not the author's own work.

Declaration of GenAI Use in the Writing Process

During the preparation of this work the author used ChatGPT to generate texts for use as data and the Quillbot AI detector for data analysis as described in the Methodology section. No AI technologies were used in the preparation of the manuscript. The author takes full responsibility for the content of the published article.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The author reports there are no competing interests to declare.

References

Altalib, A. (2026). Voice and ownership in AI-assisted L2 writing. *International Journal of TESOL Studies*, 260702. https://doi.org/10.58304/ijts.260702

Andersen, J. P., Degn, L., Fishberg, R., Graversen, E. K., Horbach, S. P., Schmidt, E. K., ... & Sørensen, M. P. (2025). Generative Artificial Intelligence (GenAI) in the research process–A survey of researchers' practices and perceptions. *Technology in Society,* 81, 102813. https://doi.org/10.1016/j.techsoc.2025.102813

Anthony, L. (2022). *AntConc v3.5.9* [Computer software]. https://www.laurenceanthony.net/software/antconc/

Baron, R. (2024). AI editing: are we there yet?. *Science Editor, 47*(3), 78-82. https://doi.org/10.36591/SE-4703-18

Bruce, I. (2008). *Academic Writing and Genre: A Systematic Analysis.* Bloomsbury.

Charbonneau, D. H., & Zhang, M. (2026). Artificial Intelligence (AI) guidance for authors, peer reviewers, and editors: A content analysis of journal policies. *Accountability in Research,* 2632083. https://doi.org/10.1080/08989621.2026.2632083

Cheng, A., Lin, Y., Reedy, G., Joseph, C., Wirkowski, S., Mallette, V., Nagesh, V., Krieser, D. & Calhoun, A. (2025). Ability of AI detection tools and humans to accurately identify different forms of AI-generated written content. *Advances in Simulation, 10*(1), 66. https://doi.org/10.1186/s41077-025-00396-6

Drechsler, S. (2026). *The Responsible AI Literacy (RAIL) Competency Model for Researchers* – (Validation Version 0.9). Paderborn University, Germany. https://doi.org/10.17619/UNIPB/1-2504

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity, 19*(1), 17. https://doi.org/10.1007/s40979-023-00140-5

Gabriel, S. (2024). Generative Al in writing workshops: A path to Al literacy. In *Proceedings of the 4th International Conference on AI Research. Academic Conferences International.*

Ganjavi, C., Eppler, M. B., Pekcan, A., Biedermann, B., Abreu, A., Collins, G. S., ... & Cacciamani, G. E. (2024). Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: bibliometric analysis. *BMJ,* 384. https://doi.org/10.1136/bmj-2023-077192

Gulumbe, B. H. (2024). Obvious artificial intelligence-generated anomalies in published journal articles: A call for enhanced editorial diligence. *Learned Publishing, 37*(4). https://doi.org/10.1002/leap.1626

Hirosawa, T., Harada, Y., & Shimizu, T. (2025). Transforming world and word: a longitudinal analysis of verb usage within English abstracts from medical articles. *AI & Society,* 1-14. https://doi.org/10.1007/s00146-025-02750-8

Hu, G. (2026). The crushing weight of fake papers. *Journal of English for Academic Purposes,* 81, 101688. https://doi.org/10.1016/j.jeap.2026.101688

Hutchins, E. (2000). Distributed cognition. In *Smelser, N. J. & Baltes, P. B. (eds) International encyclopedia of the social and behavioral sciences,* 2068-2072. Elsevier. https://doi.org/10.1016/B0-08-043076-7/01636-3

Jiang, F., & Hyland, K. (2025). Does ChatGPT argue like students? Bundles in argumentative essays. *Applied Linguistics, 46*(3), 375-391. https://doi.org/10.1093/applin/amae052

Juzek, T. S., & Ward, Z. B. (2025, January). Why does ChatGPT "Delve" so much? Exploring the sources of lexical overrepresentation in Large Language Models. In *Proceedings of the 31st international conference on computational linguistics* (pp. 6397-6411).

Khanh-Quynh, T. T., & Cong-Lem, N. (2026). The new writing partner? EFL students' real-time use of GenAI tools in academic composing. *Applied Language Sciences,* 2, 260107. https://doi.org/10.65553/ALS.260107

Krumholz, H. M. (2025). Tools, not ghosts: artificial intelligence, writing, and responsibility. *Journal of the American College of Cardiology, 86*(14), 1015-1016. https://doi.org/10.1016/j.jacc.2025.08.064

Lei, S., & Yang, R. (2020). Lexical richness in research articles: Corpus-based comparative study among advanced Chinese learners of English, English native beginner students and experts. *Journal of English for Academic Purposes,* 47, 100894. https://doi.org/10.1016/j.jeap.2020.100894

Liang, W., Izzo, Z., Zhang, Y., Lepp, H., Cao, H., Zhao, X., ... & Zou, J. Y. (2024). Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews. *arXiv preprint arXiv:*2403.07183 https://doi.org/10.48550/arXiv.2403.07183

Miralles-González, P., Huertas-Tato, J., Martín, A., & Camacho, D. (2025). Not all tokens are created equal: Perplexity Attention Weighted Networks for AI generated text detection. *arXiv preprint arXiv:*2501.03940. https://doi.org/10.48550/arXiv.2501.03940

Mohebbi, A. (2026). Ensuring ethical standards in scholarly publishing: The future of AI-driven knowledge production. *Journal of English for Research Publication Purposes, 6*(2), 220-247. https://doi.org/10.1075/jerpp.00033.moh

Moorhouse, B. L., Nejadghanbar, H., Wu, C., Kaur, H., & Yeo, M. A. (2026). Emerging conventions in GenAI disclosure: how applied linguistics scholars disclose AI use. *Applied Linguistics Review.* https://doi.org/10.1515/applirev-2025-0291

Nini, A. (2019). The Multi-Dimensional Analysis Tagger. In Berber Sardinha, T. & Veirano Pinto M. (eds), MultiDimensional Analysis: Research Methods and Current Issues, 67-94, Bloomsbury Academic.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187-192. https://doi.org/10.1126/science.adh2586

Ou, A. W., Khuder, B., Franzetti, S., & Negretti, R. (2024). Conceptualising and cultivating Critical GAI Literacy in doctoral academic writing. *Journal of Second Language Writing,* 66, 101156. https://doi.org/10.1016/j.jslw.2024.101156

Parker, J. L., & Becker, K. P. (2026, April). Defining and assessing AI literacy for researchers across the research lifecycle. *Frontiers in Education,* 11, 1827603. https://doi.org/10.3389/feduc.2026.1827603

Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes,* 14, 48-59. https://doi.org/10.1016/j.jeap.2013.12.001

Poquet, O., Nanduri, M. S., Loyer, M. X. S., Stadler, M., Sailer, M., & Jovanovic, J. (2026). Profiling cognitive offloading in LLM-mediated synthesis writing: Volume vs. content. *arXiv preprint arXiv:*2606.10434. https://doi.org/10.48550/arXiv.2606.10434

Rodrigues, F. A., Sturm, N. F., & Pinheiro, F. L. (2026). A linguistic comparison between human-and AI-generated content. *Iscience, 29*(3). https://doi.org/10.1016/j.isci.2026.114976

Salomon, G. (1997). *Distributed Cognitions: Psychological and Educational Considerations.* Cambridge University Press.

Tang, G. (2024). Letter to editor: Academic journals should clarify the proportion of NLP-generated content in papers. *Accountability in Research, 31*(8), 1242-1243. https://doi.org/10.1080/08989621.2023.2180359

Xhafa, H. (2025). SME Financing and Credit Guarantees: Evidence and Design Lessons from Albania and the Western Balkans. *Transnational Academic Journal of Economics, 2*(3), 187-196. https://doi.org/10.5281/zenodo.18442763

Watson Todd, R. (2024) Generative AI in research: Uses and abuses. The 16th National and International Academic - Research Network Conference in Humanities and Social Sciences (HUSOC 2024), May 17, 2024. Available at: https://sola.pr.kmutt.ac.th/crs/wp-content/uploads/2025/07/Generative-AI-in-research.pptx

Wu, C., Moorhouse, B. L., Wan, Y., & Wu, M. (2026). Exploring PhD students' utilization of generative AI in academic writing for publication purposes: Insights for EAP. *Journal of English for Academic Purposes,* 79, 101612. https://doi.org/10.1016/j.jeap.2025.101612

Yamashita, H. & Meier, L. (2025). Detecting AI-generated text in student submissions using multi-modal classification. *International Journal of Innovative Science and Research Technology, 10*(10), 2302-2313. https://doi.org/10.38124/ijisrt/25oct1328

Zhaksylyk, A., Zimba, O., Yessirkepov, M., & Kocyigit, B. F. (2023). Research integrity: where we are and where we are heading. *Journal of Korean Medical Science, 38*(47). https://doi.org/10.3346/jkms.2023.38.e405

Zhao, H., & Han, Z. (2026). Understanding the mechanism of human–AI interaction: a distributed cognition perspective. *Journal of Documentation, 82*(4): 1042–1063. https://doi.org/10.1108/JD-01-2026-0003

Zhou, Y., He, B., & Sun, L. (2024, August). Navigating the shadows: Unveiling effective disturbances for modern ai content detectors. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (pp. 10847-10861). https://doi.org/10.18653/v1/2024.acl-long.584

Author biodata

Richard Watson Todd is a professor at School of Liberal Arts, King Mongkut's University of Technology Thonburi
E-mail: richard.tod@kmutt.ac.th