

ACADEMIC WRITING WITH GPT-3.5: REFLECTIONS ON PRACTICES, EFFICACY AND TRANSPARENCY

Oğuz 'Oz' Buruk
Gamification Group
Tampere University
oguz.buruk@tuni.fi

ABSTRACT

The debate around the use of GPT-3.5 has been a popular topic among academics since the release of ChatGPT. Whilst some have argued for the advantages of GPT-3.5 in enhancing academic writing, others have raised concerns such as plagiarism, the spread of false information, and ecological issues. The need for finding ways to use GPT-3.5 models transparently has been voiced, and suggestions have been made on social media as to how to use GPT-3.5 models in a smart way. Nevertheless, to date, there is a lack of literature which clearly outlines how to use GPT-3.5 models in academic writing, how effective they are, and how to use them transparently. To address this, I conducted a personal experience experiment with GPT-3.5, specifically by using text-davinci-003 model of OpenAI, for writing this article. I identified five ways of using GPT-3.5: Chunk Stylist, Bullet-to-Paragraph, Talk Textualizer, Research Buddy, and Polisher. I reflected on their efficacy, and commented on their potential impact on writing ethics. Additionally, I provided a comprehensive document which shows the prompts I used, results I got from GPT-3.5, the final edits and visually compares those by showing the differences in percentage.

Author Keywords

GPT, ChatGPT, AI, Language Learning Models, LLM, Academic Writing, Ethics of Writing

INTRODUCTION

In recent months, GPT-3.5 models has become very popular among writers and researchers, especially with the launch of ChatGPT. There has been a surge of interest in using this tool for academic writing for a variety of purposes. Scientists have even gone as far as to add ChatGPT as a co-author to papers.

The ability to communicate scientific material in a paper is essential for successful writing. In some cases, the lack of skill, hesitation, or being slow in or disliking writing process can present significant barriers for researchers independent from their experience [14]. The potential of GPT-3.5 therefore to revolutionize the way science is communicated cannot be underestimated. In addition, it offers a way for those with limited writing skills, or whose mother tongue is not English, to flowingly put their ideas into words, which

can help to reduce the advantage gap between researchers who do and do not have English as their first language [8].

Notwithstanding the excitement surrounding GPT-3.5, it has also raised serious concerns among scientists and journals. Notable outlets such as Nature and Science have published statements expressing the unease with its utilization [17,22], as well as the main principles for its use. Institutions have also published guidelines for employing GPT-3.5 in research and education [13]. These are mainly focused on ethical issues, such as content ownership, plagiarism and the potential for misinformation due to careless use without a proper check of a human-author.

In order for GPT-3.5 to be employed in an efficient yet responsible manner, further studies must be conducted to demonstrate how it can best be used. This piece details a personal experience experiment where I used GPT-3.5 in various ways in the production of this article. It reflects on the efficacy of the tool, outlines different ways of incorporating it into the writing process, and discusses methods for ethical and responsible use. I did not use ChatGPT, and instead used the Open AI Playground Interface which provides more control such as Temperature settings for changing the randomness level of the outcome. Both ChatGPT and Playground are based on the same training models and I specifically used text-davinci-003 in this experiment.

Through this experiment, I hope to provide researchers with:

- A better understanding of how GPT-3.5 can facilitate effective academic writing
- Possible ways to use it
- Reflections on ethical use and transparency

GPT-3 IN ACADEMIC WRITING

GPT-3.5 is a natural language processing (NLP) model developed by OpenAI, based on a deep learning technique called transformers. It is a large-scale language model that is trained on a massive amount of text data, allowing it to generate human-like text. GPT-3.5 is capable of generating text that is coherent and consistent with the context it is provided. It can be used for a variety of tasks, including text summarization, question answering, and text generation. Currently, code-davinci-002, text-davinci-002 and text-davinci-003 (which is also used in the generation of this

paper) models are referred to as GPT 3.5. The popular ChatGPT application, and other services of OpenAI Playground such as text completion, incorporates models that can be referred to GPT-3.5 [23].

GPT-3.5, and its predecessor like GPT-3 and GPT-2, have recently become popular amongst academics for their utilization in different scientific communication purposes such as addressing challenges in writing, navigating extensive literature and providing definitions of concepts [12]. Twitter threads have been released demonstrating the potential uses of ChatGPT in a “smart” way [1]. Studies conducted by independent experts to assess the quality of content found that it produces high-quality results, which are hard to distinguish from human-generated content [19]. GPT models has also been proposed as an aid for academic writing for students [15].

Despite the potential uses of GPT in academic writing, there are other potentially problematic results that have been surfaced. For example, some researchers have added ChatGPT as a co-author in research papers [9,10]. Such actions have been advised against by Science and Nature, who have released editorials [17,22] to dispel the notion of GPT authorship, citing concerns such as assigning responsibility, agency and ownership of content produced by language learning models, plagiarism and transparency. Debouche have also raised similar concerns for the utilization of GPT-3 and recommended authors to openly share the prompts and outcomes used [5]. Other issues include embedded biases (such as hate speech towards race, sexism) [3,11], exploitation of workers for data labeling, permissions on the data used for training and environmental concerns due to the energy used in algorithm training [16]. Moreover, some studies have proposed that AI-supported NLP models should be open-source and developed together with stakeholders, rather than being a product of a private company [4].

In conclusion, although GPT models has been proposed as a tool for academic writing, there are still debates focusing on the potential positives and negatives of its utilization. We also lack studies that would reflect on its efficacy in supporting writing, different ways of utilizing them and demonstrate a workflow that can be considered transparent by reviewers.

PROCEDURE FOLLOWED

In this personal experience study, I am exploring the possible implications of using GPT-3.5 to create an academic article in a time and effort efficient way while still maintaining academic integrity and transparency. I am also reflecting on the implications of different ways of using GPT-3.5 on my writing speed, style and motivation. While this study is not intended to be a generalizable study examining all potential applications and ethical considerations of GPT-3.5 in academic writing, I intend to create a reputable source which

explains the writing process with GPT-3.5 and demonstrates it in a transparent fashion to help other researchers use the tool responsibly.

In order to create the content of this paper, I utilized the text completion model (text-davinci-003) of OpenAI Playground text completion as a tool, in all sections of the paper. I, first, outlined the structure of my paper and took notes about the content of each section. Then, I read the guide [24] for creating text completion prompts to better understand the model's capabilities. For each section, I created input prompts with different methods such as rough paragraphs, questions, bullet points or voice recordings. These prompts were then fed into the GPT-3.5 model and the output was saved. I presented all results in a supplementary material by showing the original prompt, the output from GPT-3, as well as how I edited it, allowing the process to be transparently seen afterwards (as done in [16]). I used countwordsfree web-site for visualization [21]. Additionally, I took notes to a separate document regarding my experience while running the trials with GPT-3.5. Lastly, I organized my data into an affinity diagram to highlight the pertinent parts for discussion. I also did a final manual-copy edit to the paper and made minor changes and additions to support the narrative flow.

WAYS OF USING GPT-3.5 FOR ACADEMIC WRITING

In this section, I have outlined the various methods I employed in my writing process, providing examples of the prompts I used, the outcomes of the prompts, and the final version of the paragraph. I also included the visual representation of differences between each in the supplementary material. Each of these sections has been crafted using the method described.

Chunk Stylist

I have used GPT-3.5 as an effective writing tool by typing chunks of text without worrying too much about the style, grammar, or spelling. This approach has enabled me to create text quickly by pouring my train of thought to text, as I am not restrained by of scrutinizing if the content is engaging or readable. This strategy can be advantageous to both experienced and novice writers alike, because temperature settings of text completion tool can be adjusted to produce either more predictable results closer to original text or more improvised and creative outcomes. I used this Chunk Stylist method during the writing of my introduction, background, and this section. To demonstrate the process, I have provided the prompt and result below, and final version of the paragraph.

Prompt:

Rewrite the following text in first person, rick, clear and in academic language:

Original Text:

One effective way of using GPT-3 is to write a chunk of text without caring too much about the style, grammar or spelling. This makes the writing process quite fast, freed me from thinking of if the style is engaging or readable and basically let me to reflect my train of thought to text in a quite fast way. This might be effective both for fluent writers and for others who struggle to put the words into the paper in a flowing manner. Because, by playing with the temperature settings of GPT-3, authors either might get less random results which is closer to the text they put, or more randomized ones that allow GPT-3 to be a bit more improvisational on the given text. I have written this section, introduction and the background by using the chunk stylist method. Below, I also demonstrate the prompt I have used, the result I got and the final version is this paragraph.

Result (difference to original: 82.61%):

I have personally experimented with GPT-3 as an effective writing tool by typing chunks of text without worrying too much about the style, grammar, or spelling. This approach has enabled me to create text quickly, as I am not restrained by considering if the content is engaging or readable. This strategy is advantageous to both experienced and novice writers alike, because temperature settings can be adjusted to produce either more predictable results or enhanced randomness. I used this chunk stylist method during the writing of my introduction, background, and this section. To demonstrate the process, I have provided the prompt, result, and final version of this paragraph.

Final (difference to result: 25.54%):

I have used GPT-3.5 as an effective writing tool by typing chunks of text without worrying too much about the style, grammar, or spelling. This approach has enabled me to create text quickly by pouring my train of thought to text, as I am not restrained by scrutinizing if the content is engaging or readable. This strategy can be advantageous to both experienced and novice writers alike, because temperature settings of text completion tool can be adjusted to produce either more predictable results closer to original text or more improvised and creative outcomes. I used this Chunk Stylist method during the writing of my

introduction, background, and this section. To demonstrate the process, I have provided the prompt and result below, and final version of the paragraph.

Research Buddy

GPT-3.5 can be a useful tool for exchanging research ideas and starting points for academic writing. Its natural language processing capabilities can allow us to explore ideas quickly and easily, while its predictive algorithm can suggest ideas and approaches that we may not have thought of. This can help us to develop effective arguments and explore diverse topics more efficiently. However, due to its reliance on machine learning, GPT-3.5 can also present some risks, such as producing inaccurate or random results, and its predictions may not always be in line with the academic rigor required of the project. Thus, it is important to always keep critical thinking at the forefront of the research process, and to carefully evaluate any output given by GPT-3.5. For example, the first paragraph of Section 2 were written with the research buddy method, by specifically prompting to give a definition of GPT-3.5 from the cited website. However, the results were too broad and were not describing what GPT-3.5 accurately, so I needed to add critical information manually. The first paragraph of the Section 2 and this section were written by using the Research Buddy approach. Below are the prompts:

Prompt:

How to use GPT-3 as a research buddy for exchanging research ideas for academic writing? Mention research advantages and risks. Write only one paragraph in rich, clear, first person and academic language:

Result:

As a researcher, I have found that GPT-3 can be an extremely useful tool for exchanging research ideas and starting points for academic writing. Its natural language processing capabilities can allow us to explore ideas quickly and easily, while its predictive algorithm can suggest ideas and approaches that we may not have thought of. This can help us to develop effective arguments and explore diverse topics more efficiently. However, due to its reliance on machine learning, GPT-3 can also present some risks, such as producing inaccurate or random results, and its predictions may not always be in line with the academic rigor required of the project. Thus, it is important to always keep critical thinking at the forefront of the research process, and to carefully evaluate any output given by GPT-3.

Final (difference to result: 39.02%):

GPT-3.5 can be a useful tool for exchanging research ideas and starting points for academic writing. Its natural language processing capabilities can allow us to explore ideas quickly and easily, while its predictive algorithm can suggest ideas and approaches that we may not have thought of. This can help us to develop effective arguments and explore diverse topics more efficiently. However, due to its reliance on machine learning, GPT-3.5 can also present some risks, such as producing inaccurate or random results, and its predictions may not always be in line with the academic rigor required of the project. Thus, it is important to always keep critical thinking at the forefront of the research process, and to carefully evaluate any output given by GPT-3.5. For example, the first paragraph of Section 2 were written with the research buddy method, by specifically prompting to give a definition of GPT-3.5 from the cited website. However, the results were too broad and were not describing what GPT-3.5 accurately, so I needed to add critical information manually. The first paragraph of the Section 2 and this section were written by using the Research Buddy approach. Below are the prompts:

Bullet-to-Paragraph

As part of the experiment, I utilized the bullet-to-paragraph method to produce a paragraph for the Method section. This method is similar to the Chunk Stylist of writing, as it involves listing the primary principles or steps of a procedure before instructing GPT-3.5 to generate a flowing paragraph in a specific language style. By doing so, I was able to create both this section and the second paragraph of the Method section.

Prompt:

Write one paragraph from the following bullets in rich, clear, first person and academic language:

Original Text:

- Bullet to paragraph is also similar to chunk styler but it's writing in bullets
- A researcher can list the main principle or steps of a procedure
- Then give a prompt to make a paragraph out of this bullets in specific language styles

- GPT-3 will produce a flowing paragraph
- This section and the second paragraph of method section were created with this method

Result (difference to original: 78.25%):

As part of an experiment, I utilized the bullet-to-paragraph method to produce a paragraph for the Method section. This method is similar to the chunk-style of writing, as it involves listing the primary principles or steps of a procedure before instructing GPT-3 to generate a flowing paragraph in a specific language style. By doing so, I was able to create both this section and the second paragraph of the Method section.

Final (difference to result: 4.12%):

As part of the experiment, I utilized the bullet-to-paragraph method to produce a paragraph for the Method section. This method is similar to the Chunk Stylist of writing, as it involves listing the primary principles or steps of a procedure before instructing GPT-3.5 to generate a flowing paragraph in a specific language style. By doing so, I was able to create both this section and the second paragraph of the Method section.

Talk Textualizer

Talk Textualizer is providing a monologue to GPT-3.5 and let it turn it into a readable text. Playground's sophisticated transcription feature enabled converting speech recordings into a readable text, even when there were long pauses or stuttering. To further refine this text, I supplemented it with a prompt designed to transform it into academic writing. This approach formed the foundation for this section, discussion, limitations, and conclusion.

Prompt:

Turn the below text into academic writing, clear, first person, and rich academic language:

Original Text:

[transcribed from the spoken audio]

Talktextualizer is almost like having a conversation with ChatGPT. ChatGPT has an advanced transcription function, speech to text function, that gets your speech

recording and even if you have long pauses or even if you stutter and tell the same word a few times, it still turns it into a somewhat legible text. Later, for this method, we take this text and add on top of it a prompt that will turn it into an academic writing. This section is written in that way And the whole of discussion and limitations and conclusion also heavily relied on this method.

Result (difference to original: 84.26%):

I utilized Talktextualizer to have a conversation with ChatGPT. This platform has an advanced transcription function which converts speech recordings into a legible text, even in cases of long pauses or stuttering. To further refine this text, a prompt was added to turn it into an academic writing. This method formed the basis of my discussion, limitations, and conclusion.

Prompt 2:

Copy-edit the following text, academic stlye, first person, clear, rich and flowing language:

Original Text 2:

I utilized Talk Textualizer by giving a monologue to GPT-3. GPT 3's advanced transcription function converts speech recordings into a legible text, even in cases of long pauses or stuttering. To further refine this text, I added a prompt to turn it into an academic writing. This method formed the basis of my discussion, limitations, and conclusion.

Results 2 (difference to original text 2: 61.24%):

I employed Talk Textualizer by providing a monologue to GPT-3. GPT-3's sophisticated transcription feature enabled it to convert speech recordings into a readable text, even when there were long pauses or stuttering. To further refine this text, I supplemented it with a prompt designed to transform it into academic writing. This approach formed the foundation for my discussion, limitations, and conclusion.

Final (difference to result 2: 20.79%):

Talk Textualizer is providing a monologue to GPT-3.5 and let it turn it into a readable text. Playground's sophisticated transcription feature enabled converting speech recordings into a readable text, even when there were long pauses or stuttering. To further refine this text, I supplemented it with a prompt designed to transform it into academic writing. This approach formed the foundation for this section, discussion, limitations, and conclusion.

Polisher

Polisher is using GPT-3.5 for copy editing purposes. Material given to GPT-3 should already be in a complete form, with a clear narrative flow. By providing a prompt like "copy-edit this paragraph in an academic style", GPT-3.5 can be used to correct grammatical mistakes, spelling errors and increase the clarity of the text. This method can produce results that are close to the original text, and temperature settings of GPT-3.5 may be lowered if authors want minimal changes. This method has been applied in this section, and also for polishing edited results of prompts throughtout this piece (for example, the first paragraph of Talk Textualizer was obtained using the Polisher method on the result of the same section).

Prompt:

Copy-edit the following text, academic style, first person, clear, rich and flowing language:

Original Text:

Polisher is using GPT-3 almost like a copy editor. In this method, based material that would be put into GPT-3 should already be in a complete form and in a proper narrative flow. By giving a prompt like "copy-edit this paragraph in an academic language," GPT-3 can be used to correct the gramattical mistakes, spelling error and for increaing the clarity of the text. This method can give results closer results to the original text, and the temperature settings might be lowered if the athors want minimal changes. This method has been used in this section, and also for polishing the many of the edited results (for example the first paragraph of Talk Textualizer was obtained by using Polisher method on the indicated result).

Result (difference to original: 34.47%):

Polisher is using GPT-3 in a similar way to a copy editor. Material given to GPT-3 should already be in a complete form, with a clear narrative flow. By providing a prompt like “copy-edit this paragraph in an academic style”, GPT-3 can be used to correct grammatical mistakes, spelling errors and increase clarity of the text. This method can produce results that are close to the original text, with temperature settings that may be lowered if authors want minimal changes. This method has been applied in this section, and also for polishing edited results (for example, the first paragraph of Talk Textualizer was obtained using the Polisher method on the indicated result).

Final (difference to result: 19.25%):

Polisher is using GPT-3 for copy editing purposes. Material given to GPT-3 should already be in a complete form, with a clear narrative flow. By providing a prompt like “copy-edit this paragraph in an academic style”, GPT-3 can be used to correct grammatical mistakes, spelling errors and increase the clarity of the text. This method can produce results that are close to the original text, and temperature settings of GPT-3 may be lowered if authors want minimal changes. This method has been applied in this section, and also for polishing edited results of prompts throughout this piece (for example, the first paragraph of Talk Textualizer was obtained using the Polisher method on the result of the same section).

DISCUSSION*Reflections on using GPT-3.5 for Writing Articles*

The use of GPT-3 to aid my writing process had a largely positive effect on my motivation and productivity. Even though the article I wrote was not particularly complicated or lengthy, my writing process was much smoother than usual. Before starting to write, I spent about three weeks mulling over the project, allowing the idea to mature, following developments, and finding two free days for writing and experimenting with GPT-3.5. What I realized was that during this incubation period and the writing stage, I became less preoccupied with the mechanics of writing and more focused on accurately conveying my ideas. This heightened my motivation and enabled me to compose the text faster, compared to my previous experiences. Furthermore, my thinking and writing processes were both more efficient, streamlining the whole writing experience.

Academic writing is an essential skill for scholars from all disciplines, though the level of training, experience, and receptivity to writing can vary depending on the field of study. For instance, those in the social sciences and humanities may have more practice with and exposure writing, while academics in more applied fields such as engineering or design may have less proficiency and experience with written expression. Drawing on my own experience, in design research, crafting compelling narratives is almost as important as in other social sciences in order to communicate the value and position of artifacts created or the design knowledge produced. Personally, I find myself more adept at envisioning artifacts and bringing them to life (applied part of the design research) than at imagining stories and writing, and the laborious nature of writing process which incorporates computational tools only for visual styling can be a bottleneck in my research. However, upon exploring the use of GPT-3.5 to textualize my ideas, I recognized a similarity to my design practice. With my background in design, I am accustomed to utilizing tools such as CAD software, 3D printers, electronic boards, and other prototyping tools for creating tangible artifacts, and these tools are integral to my thinking process of designing things and makes it easier to conceptualize the process of thinking about artifacts and their production. Similarly, GPT-3.5 has provided me with a similar freedom as I have leaned on it in the same way I do with computational design tools; it has enabled me to focus on the ideas I have rather than the daunting task of manual writing. It also allowed me to work on my writing without distractions and interruptions. Although basic writing skills are still necessary for creating a narrative, GPT-3.5 has helped me to lighten the burden of writing and streamline the process.

I have also found that I have not been as time-efficient with my writing as I had anticipated. Writing a section of text and then making multiple revisions until I reach a satisfactory outcome has taken quite some time. For instance, it took me approximately 10 minutes to write the primary material of an introduction while the subsequent revisions took around 90 minutes which is quite long for such a short section. Although I believe I was able to write more quickly due to the lack of distractions which happens in my usual writing practice because of the interruptions in my thought process, I still believe that authors need to dedicate a substantial amount of time in the process. For creating the report which shows the comparisons between the original text, outcomes of GPT-3.5 and final edited text, I spent around 5 more hours. In addition, the current slow speed of GPT-3.5 due to the overwhelming demand on its servers means it cannot be considered a quick solution for writing, but rather a tool that makes writing process smoother and introduces new writing techniques suitable for different skill levels.

I believe that the quality of this text is variable across different sections. As I made the last round of revisions, I noticed this especially when reading through the entire piece.

Initially, I edited the text immediately after receiving the results from GPT-3.5, and then I made further revisions after the entire piece was written. Some sections were plain and unengaging, while others were easier to read and more compelling. I could have done a few more rounds of revisions to ensure a consistent tone and a more flowing narrative, but I left the manuscript as it is to demonstrate what can be produced with GPT-3.5 in a short time (around a day) and with minimal editing. However, I believe that several more rounds of revision would be necessary to create an academic article with a consistent language style and a solid narrative flow.

Ways of Using Methods

I utilized the methods of Chunk Stylist and Talk Textualizer predominantly when writing this article. I was inclined to use them because of their ability to formalize my own ideas, rather than relying on spontaneously generated material by GPT-3.5. When using Bullet to Paragraph or Research Buddy, I had to make several attempts and perform substantial editing to avoid sentences not based on facts. Thus, I believe the Chunk Stylist and Talk Textualizer are more suitable for writing a formal article than methods relying on big portion of generated text such as Research Buddy. For example, when I prompted GPT-3.5 to give me the description of GPT-3 based on [2], I found that the summary generated was not reflective of the source material (the document was mostly technical and required good understanding of NLP, so I could not assess the accuracy of information). As I was unable to identify enough supporting information to back up GPT-3.5's output, I had to remove most of the material and leave only the core points that I knew were accurate. I used Polisher mainly to refine the results that I had edited; however, it can be employed by proficient writers for copy-editing their text.

Transparency, Agency and Biases

One of the primary concerns of incorporating GPT-3.5 into the writing process is the potential for plagiarism and a lack of transparency [20]. In my own trials, some instances, the results I obtained were not dissimilar from those produced by tools like Grammarly or the spell and grammar checker of Microsoft Word. In other cases, however, I was unsure of my agency over the written text. This experience was arguably similar to using a professional copy-editing service where I need to carefully check the text and make sure that the intended meanings are retained. However, the speed of the process and the lack of knowledge about where the words and sentences derived from caused me to feel uneasy about using the content. Overall, looking at the full text, the difference between the original texts I prompted to GPT-3.5 and the outcomes I got was 70.27%, while the difference between outcomes and the final edited version was 29.13%. The biggest difference between an original text chunk and an outcome was 98.54% (almost the whole text were changed), the biggest difference between the outcome and the final

edited part was 71.44% and the smallest was 3.3% (copy-pasted to this manuscript almost without a change). You can see the detailed report in the supplemented document.

This variability in author's agency over the result produced GPT-3.5 require to develop practices for transparency for articles where GPT-3.5 (or other LLM) incorporated. In this article, I have documented all the prompts, results, and final edited versions I used. In some cases, this is essential; for example, when using methods such as Talk Textualizer, Bullet to Paragraph, or Research Buddy, the heavy influence of GPT-3.5 on the content, language, and tone of the writing is readily apparent. In such situations, it is of critical importance to be transparent and clearly demonstrate the process of the writing and how it has been transformed.

In their recent editorial, Nature suggested that authors must mention their use of GPT-3.5 in their writing [22], but I believe this may not be enough. If GPT-3.5 has only been used for minor copyediting purposes, then a note in the acknowledgement may suffice. However, if other methods have been used, such as Talk Textualizer or Bullet to Paragraph, a more substantial reporting—perhaps in the form of an external link or an appendix—is necessary to ensure transparency, which also has been used in other contexts such as classroom assignments [6]. This would also ensure that the authors would thoroughly check the content making sure that their content do not consolidate racial biases or any other radicalized political ideas unintentionally. Additionally, better tools may be required to demonstrate the extent of changes and highlight where heavy modifications have been made, as well as to provide transparency by showing the source from which the text was generated.

Concerns of Ecological Sustainability

The use of GPT-3.5 and similar tools raises ecological sustainability concerns [7]. In writing this article, I created approximately 500 requests and often carelessly reprompted GPT-3.5 for the same paragraph until reaching a satisfactory result. According to OpenAI's calculations, this article costed around \$2.5, but I am unaware of the carbon footprint created in doing so.

On the one hand, the integration and normalization of GPT-3.5 and similar language learning models into academia could have a substantial negative effect on energy consumption in comparison to the thought process facilitated by the brain and body, which could have resulted in less resource expenditure. On the other hand, my writing process has been more efficient with the use of GPT-3.5 - not because it is time-efficient as a tool but makes me more motivated and less prone to distraction while using it - potentially reducing electricity costs spent by my computer or office space. Additionally, my writing journeys often extend to the middle of the night, which is arguably not beneficial for my physical and mental health in the long-term, lowering the consequences of mental burden caused by the pressure and

stress of writing preventing overworking, a known problem in academia [18].

Using tools like GPT-3.5 in academia could have a significant negative impact on ecology and as academics we are responsible of regulating our behavior accordingly, such as using it with least prompts as possible. Still, we should consider the positive impacts of using these tools beyond just being more efficient in writing tasks, such as its impact on more efficient utilization of resources and improving the wellbeing of academics.

Additional tools that can help with GPT-based academic writing

Throughout my trials and reflections, I have come to recognize that further tools developed in the future may not only help increase the efficiency of writing, but also increase transparency and make us more aware of our responsibility - particularly with regard to ecological sustainability - and make our writing less prone to the dissemination of fake information. Currently, there are attempts to create watermark tools [16] that would enable people to easily identify if the text is generated through LLM models such as GPT-3.5. Although this would be useful in certain contexts, including classroom assignments that focus on teaching content to students, its utility may not be as significant in other scenarios, particularly for academic writing. A binary solution such as a watermark which only indicates whether GPT is used or not may not be the best approach; instead, we might need text editors that are supported by GPT and can clearly document prompts, the outcomes of those prompts, and make clear, visible, and easily understandable visualizations of the comparisons between the prompt, the result, and the edited version (as I have manually done in this piece).

Additionally, a tool that gives information of the carbon footprint of the writing piece and compare it to potential carbon footprint if the GPT was not used might be useful. Through such a comparison, scientists would be more aware of their responsibility in using GPT in a mindful way to the environment.

Another toolset which would be beneficial is one that prevents the dissemination of faulty information, integrating measures which detect authoritative tone in the written text and comprehend whether it refers to a fact. Such tools should provide reliable facts, scrape and suggest real references, and create correct references based on the names of the papers or the links; something that current GPTs are not capable of doing.

LIMITATIONS

Drawing upon my own experience, I sought to reflect on the use of GPT-3.5 in the creation of an academic article in this perspective piece. It is important to note that my reflections have been shaped by my perspective and positionality as a design researcher. Thus, while I believe that the ways of

using of GPT-3.5 in academic writing as discussed in this paper can be beneficial for scholars from a variety of disciplines, my reflections and opinions may not be relevant to all and should not be considered as generalizable across the board. In my opinion, my reflections may be of greater relevance to those working in fields where applied science plays an important role, yet expression of ideas is equally valued. Also, it must be noted that this trial was limited in duration, as one of its primary aims was to assess efficacy, and longer engagement with the tool might allow the generation of new ideas, practices and opinions. Moreover, the content of the writing in this paper was relatively practical and did not require much philosophical scrutiny. Thus, GPT-3.5 may not be as useful in fields where complex thoughts need to be expressed using precise or abstract language. Similar experiments might be conducted to understand and demonstrate its efficacy across disciplines.

CONCLUSION

In this paper, I have reflected on my experience of using GPT-3.5 as a tool for academic writing and discussed some basic methods for providing transparency when using it. I have shared my reflections on the efficacy of the tool and discussed the ethical considerations of using GPT-3.5 for academic writing around the issues raised by previous studies I have identified five ways of using GPT-3.5 for academic writing: Chunk Stylist, Research Buddy, Talk Textualizer, Bullet-to-Paragraph and Polisher. Of these, I have predominantly used Chunk Stylist and Talk Textualizer to great effect, enabling me to write more effectively while still retaining my own tone and ideas. Caution should be exercised with the other methods, however, as they introduce more spontaneity and randomness, which could result in plagiarism, the dissemination of false information, or even the exacerbation of hateful rhetoric if not used responsibly. To ensure fidelity and transparency, I documented all prompts, outcomes, and final edits in a separate supplementary document, which can serve as an example of using GPT-3.5 in a transparent way.

I hope that my reflections will be of benefit to researchers who are considering using GPT-3.5 and will inspire the development of tools for more responsible practices and methods for using AI-supported natural language processing in academia. I believe that this work will contribute to the debate surrounding the incorporation of this technology into our scholarly work, by increasing understanding of its efficacy and possible ways of responsible usage.

ACKNOWLEDGEMENTS

This manuscript was written by heavily benefiting from GPT-3.5. A detailed report of used prompts, outcomes and the final edited versions can be seen in Appendix.

While giving prompts, I used the word “GPT-3” instead of “GPT-3.5” since I realized the difference between two after a while. Still, all parts that might have been affected by this

difference (e.g., asking what GPT-3 is, or ways of using it in research) has been corrected.

REFERENCES

1. Mushtaq Bilal. 2023. Chatgpt is everywhere and everyone is using it... *Twitter*. Retrieved from <https://twitter.com/MushtaqBilalPhD/status/1621379333943083009>
2. Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, and others. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33: 1877–1901.
3. Ke-Li Chiu, Annie Collins, and Rohan Alexander. 2021. Detecting hate speech with gpt-3. *arXiv preprint arXiv:2103.12407*.
4. Vanya Cohen and Aaron Gokaslan. 2020. OpenGPT-2: Open Language Models and Implications of Generated Text. *XRDS* 27, 1: 26–30. <https://doi.org/10.1145/3416063>
5. Nassim Dehouche. 2021. Plagiarism in the age of massive Generative Pre-trained Transformers (GPT-3). *Ethics in Science and Environmental Politics* 21: 17–23.
6. Paul Fyfe. 2022. How to cheat on your final paper: Assigning AI for student writing. *AI & SOCIETY*: 1–11.
7. Van Rooij Iris. 2023. Stop feeding the hype and start resisting — irisvanrooijcogsci.com. Retrieved from <https://irisvanrooijcogsci.com/2023/01/14/stop-feeding-the-hype-and-start-resisting/>
8. Vijay Kumar Jain, Karthikeyan P Iyengar, and Raju Vaishya. 2022. Is the English language a barrier to the non-English-speaking authors in academic publishing? *Postgraduate Medical Journal* 98, 1157: 234–235. <https://doi.org/10.1136/postgradmedj-2020-139243>
9. Michael R King and chatGPT. 2023. A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. *Cellular and Molecular Bioengineering*: 1–2.
10. Tiffany H Kung, Morgan Cheatham, Arielle Medinilla, ChatGPT, Czarina Sillos, Lorie De Leon, Camille Elepano, Marie Madriaga, Rimel Aggabao, Giezel Diaz-Candido, and others. 2022. Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models. *medRxiv*: 2022–12.
11. Kris McGuffie and Alex Newhouse. 2020. The radicalization risks of GPT-3 and advanced neural language models. *arXiv preprint arXiv:2009.06807*.
12. Albert Meroño-Peñuela, Dayana Spagnuolo, and GPT-2. 2020. Can a Transformer Assist in Scientific Writing? Generating Semantic Web Paper Snippets with GPT-2. In *The Semantic Web: ESWC 2020 Satellite Events: ESWC 2020 Satellite Events, Heraklion, Crete, Greece, May 31–June 4, 2020, Revised Selected Papers* 17, 158–163.
13. Carmo Moniz. 2023. ChatGPT is here — what’s NYU doing about it? — nyunews.com. Retrieved from <https://nyunews.com/news/2023/02/02/chatgpt-nyu-academic-integrity-policy/>
14. June Oshiro, Suzanne L. Caubet, Kelly E. Viola, and Jill M. Huber. 2020. Going Beyond “Not Enough Time”: Barriers to Preparing Manuscripts for Academic Medical Journals. *Teaching and Learning in Medicine* 32, 1: 71–81. <https://doi.org/10.1080/10401334.2019.1659144>
15. Tobias Schmohl, Alice Watanabe, Nadine Fröhlich, Dominikus Herzberg, and others. 2020. How Artificial Intelligence can improve the Academic Writing of Students. In *Conference Proceedings. The Future of Education 2020*.
16. Chris Stokel-Walker and Richard Van Noorden. 2023. What ChatGPT and generative AI mean for science. *Nature* 614, 7947: 214–216. <https://doi.org/10.1038/d41586-023-00340-6>
17. H. Holden Thorp. 2023. ChatGPT is fun, but not an author. *Science* 379, 6630: 313–313. <https://doi.org/10.1126/science.adg7879>
18. Angel Urbina-Garcia. 2020. What do we know about university academics’ mental health? A systematic literature review. *Stress and Health* 36, 5: 563–585. <https://doi.org/10.1002/smi.2956>
19. Jan Philip Wahle, Terry Ruas, Frederic Kirstein, and Bela Gipp. 2022. How Large Language Models are Transforming Machine-Paraphrase Plagiarism. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 952–963. Retrieved from <https://aclanthology.org/2022.emnlp-main.62>

20. 2020. Next chapter in artificial writing. *Nature Machine Intelligence* 2, 8: 419–419. <https://doi.org/10.1038/s42256-020-0223-0>
21. 2021. Compare text online - Show Text Difference — countwordsfree.com. Retrieved from <https://countwordsfree.com/comparetexts>
22. 2023. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature* 613, 7945: 612–612. <https://doi.org/10.1038/d41586-023-00191-1>
23. OpenAI API — Model Index. Retrieved from <https://platform.openai.com/docs/model-index-for-researchers>
24. OpenAI API — Text Completion Guide. Retrieved from <https://platform.openai.com/docs/guides/completion/introduction>