



HAL
open science

PaperCard: Towards Explainable Machine Assistance in Academic Writing

Won Ik Cho, Eunjung Cho, Kyunghyun Cho

► **To cite this version:**

Won Ik Cho, Eunjung Cho, Kyunghyun Cho. PaperCard: Towards Explainable Machine Assistance in Academic Writing. 2023. hal-04019842v1

HAL Id: hal-04019842

<https://hal.science/hal-04019842v1>

Preprint submitted on 8 Mar 2023 (v1), last revised 17 Sep 2023 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

PaperCard: Towards Explainable Machine Assistance in Academic Writing

WON IK CHO* and EUNJUNG CHO*, Seoul National University, South Korea
KYUNGHYUN CHO, New York University and Genentech, USA

Academic writing process has benefited from various technological developments over the years including search engines, automatic translators, and editing tools that review grammar and spelling mistakes. They have enabled human writers to become more efficient in writing academic papers, for example by helping with finding relevant literature more effectively and polishing texts. While these developments have so far played a relatively assistive role, recent advances in large-scale language models (LLMs) have enabled LLMs to play a more major role in the writing process, such as coming up with research questions and generating key contents. This raises critical questions surrounding the concept of authorship in academia. ChatGPT, a question-answering system released by OpenAI in November 2022, has demonstrated a range of capabilities that could be utilised in producing academic papers. The academic community will have to address relevant pressing questions, including whether Artificial Intelligence (AI) should be merited authorship if it made significant contributions in the writing process, or whether its use should be restricted such that human authorship would not be undermined. In this paper, we aim to address such questions, and propose a framework we name "PaperCard", a documentation for human authors to transparently declare the use of AI in their writing process.

CCS Concepts: • **Social and professional topics** → **Intellectual property**.

Additional Key Words and Phrases: academic writing, machine assistance, generative language models, technological development, authorship, ChatGPT

ACM Reference Format:

Won Ik Cho, Eunjung Cho, and Kyunghyun Cho. 2023. PaperCard: Towards Explainable Machine Assistance in Academic Writing. In *Undisclosed Venue*. ACM, New York, NY, USA, 15 pages. <https://doi.org/XXXXXXXX.XXXXXXX>

1 INTRODUCTION

Technological developments, especially in computing methods, have brought great efficiency and convenience to the academic writing process, for instance in finding relevant literature, polishing the presentation, and finalising the draft. Nonetheless, the core parts of academic writing, such as creating original contents and posing and answering key research questions, remain to be seen as the realm of human writers. However, recent technological developments in natural language generation have started to blur the boundary. It is becoming harder to tell apart the exact contribution made by a machine and that by human writer(s). This gives rise to a host of questions including whether and how machines should be acknowledged for its contribution, and whether the contribution of human writers who used machine assistance in their work could or

*Both authors contributed equally to this research.

Conference acronym Undisclosed, XXXX, 2023, 03, 08

© 2023 Association for Computing Machinery.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Undisclosed Venue*, <https://doi.org/XXXXXXXX.XXXXXXX>.

should be evaluated less highly than those who did not use machine assistance.

Recent breakthroughs in large language models (LLMs) have led to significant advances in machine assistance. One of the most notable developments in this field is the release of Generative Pre-trained Transformer (GPT) [48], a generative deep learning model that uses Transformer architecture [56] to achieve state-of-the-art performance on natural language generation benchmarks. This was followed by the release of GPT-2 [49], and later GPT-3 [11], both of which have demonstrated remarkable capabilities in generating human-level text. Additionally, OpenAI released InstructGPT [46], an AI model and framework that accommodates human-in-the-loop instructions to complete various natural language tasks, making them capable of generating highly coherent and fluent text. This led to the recent development of ChatGPT [43], a conversation-style language model that can be used for a wide range of applications, from answering questions to writing essays. It is considered more mature than commercial search engines as it can generate texts based on the context and input provided, rather than merely searching through a large collection of pre-existing text.

While such breakthroughs have led to many exciting possibilities, there are also concerns about the use of these models. One of the main concerns is the potential for these models to be used to produce fake or misleading content. For example, generative language models have been used to create fake blog posts [24] and impersonate individuals on social media [26]. Studies that investigated crowd reactions to text generated by ChatGPT or other similar models [25] have found that people tend to struggle to distinguish between machine-generated text and text written by humans, which raises additional concerns about the potential for deception. Also, issues related to authorship, licensing, accountability and beneficiary are another pressing and practical concern. As machine-generated text becomes more prevalent, it will be crucial to develop clear guidelines for addressing these issues such that benefits of machine assistance are distributed fairly and transparently.

This study aims to raise and address several key issues that could arise when academic writing process is assisted by machines, or more precisely, artificial intelligence (AI) ¹. These include concerns about the originality of AI-generated materials. As AI-generated text in academic writing becomes prevalent, it is important to devise ways to assess and evaluate the originality and intellectual contributions of human authors. Responsible use of AI-generated text by human authors is another key concern. As machine assistance becomes a norm, it is important for the academic community to

¹Theoretically, machines are not equal to AI. However, in this paper, we refer to all automated systems that can assist writers with the writing process as "machine". Most of the recently developed systems are based on modern AI techniques including machine learning and neural networks. In this regard, we use "machine" and "AI" interchangeably, and "generative AI" to more specifically refer to generative LLMs such as GPT-3 or ChatGPT.

establish clear guidelines for when and how AI-generated text can be used in academic papers. This includes ensuring that automatic text generation does not compromise the originality and integrity of the work, and that appropriate attribution is given to the source of the AI-generated text. In addition, there are various ethics, policy, and legal issues concerning authorship, license, accountability, and beneficiary that could arise after the manuscript is published and is acknowledged as a valid intellectual property. It is important that the rights of authors and other stakeholders who contributed to the creation of knowledge are protected, and should be held responsible for the final paper. Our research questions are as follows:

- How have technological developments over the years changed academic writing process, and how have they influenced the idea of authorship, especially in terms of how we evaluate the intellectual contribution of the authors? What are the existing and emerging concerns?
- What are some of the issues surrounding fairness, accountability, and transparency related to the use of machine assistance in academic writing, especially against the backdrop of the recent developments of high-performance generative language models?
- How should the use of AI in academic writing, and related issues of authorship, license, accountability, and beneficiary, etc. be governed?

We try to address these questions throughout the paper, and propose a framework we call “PaperCard” that authors and practitioners could utilise for writing or assessing academic papers that used machine assistance. In Section 2, we first discuss issues surrounding authorship in general, and how the concept of authorship has changed over time with various technological developments. In Section 3, we address issues related to fairness, accountability, and transparency surrounding the use of generative AI in academic writing. In Section 4, we propose and explain the concept of PaperCard in detail, and present a sample card for this paper as an example. Finally, we discuss some remaining challenges the academic community will need to consider in the coming years.

2 TECHNOLOGY AND AUTHORSHIP

2.1 Defining authorship

While different fields and institutions have slightly different definitions and criteria for authorship, in academia, authorship is generally defined as the recognition of an individual’s intellectual contribution to a particular piece of work, such as a research paper, thesis, or dissertation [45]. For example, the Association for Computing Machinery (ACM) defines authorship as “those who have made significant intellectual contributions to the work reported in the paper.” [2] According to the ACM authorship criteria, authors should have been involved in the conception or design of the work, or the acquisition, analysis, or interpretation of data for the work, in drafting the work or revising it critically for important intellectual content, and must have approved the final version of the work. The Institute of Electrical and Electronics Engineers (IEEE) has similar criteria, defining authorship as “those individuals who have made a significant intellectual or practical contribution to the work.” [28]

According to IEEE, authors should have been involved in the planning or execution of the work or the interpretation or analysis of the data, and in drafting or revising of the manuscript, and must have made an approval to the final manuscript. Similarly, Nature defines authorship as “those who made substantial contributions to the conception, design, execution, or interpretation of the reported study.” [40] Nature also states that authors should have participated in the writing of the manuscript, or the review of the intellectual content. Moreover, authors should have approved the final version of the manuscript, and should be able to take responsibility for the integrity of the work.

These definitions indicate that authorship is commonly merited to those who have made significant intellectual contributions to the work, and have been involved in the planning, execution, analysis, interpretation, writing and reviewing of the manuscript, and have approved the final version of the manuscript. They also emphasise the importance of authors being able to take public responsibility for the integrity of the work. Thus, it is crucial for researchers to be aware of the authorship criteria and definitions in their field and institution before submitting a manuscript, and to ensure that all authors meet these criteria.

2.2 Influence of technological development on authorship

Technological developments have greatly influenced the writing process in academia. One such development is the advent of the typewriter, which greatly increased the speed of writing and allowed for more efficient and precise document production [3]. Another development is the emergence of computational memory and graphical user interfaces which enabled people to save and load text they have created, making it easier to access and edit previous work. The Internet, specifically the World Wide Web, has also played a major role in shaping the writing process by providing access to vast amounts of research and related materials through search engines, making it easier for authors to conduct research and cite sources [52]. Additionally, AI-supported text editors or translators such as Grammarly ², Google Translator ³, and QuillBot ⁴ have become increasingly popular in recent years, allowing authors to proofread and paraphrase their work with ease, increasing the quality of their final outcome. These developments have greatly impacted the way in which academic papers are written, but at the same time contributed to rapid spread of misinformation, with the ease of spread of digitised text.

The next phase of technological developments has been defined by the evolution of generative language models that can generate natural-sounding text that is almost indistinguishable from text written by humans. Following the introduction of the Transformer architecture [56] that enabled language models to better understand the context of inputs and generate more coherent text, researchers at OpenAI introduced the GPT model [48] which was trained on a massive dataset of text and achieved state-of-the-art performance on a wide range of natural language processing tasks. Subsequently,

²<https://app.grammarly.com/>

³<https://translate.google.com/>

⁴<https://quillbot.com/>

GPT-2 [49] and GPT-3 [11] were introduced with improved performances, with GPT-3 equipped with capabilities to generate a wider range of texts, from poetry, news articles [19], to even abstract of academic papers [5]. InstructGPT [46] is a development of the GPT-3 model that is better at following instructions of users to generate texts.

ChatGPT, a public GPT-3 model fine-tuned in a manner similar to InstructGPT, has been widely utilised in various settings since its release in November 2022. In the context of academic writing, much attention has been drawn as researchers found ways to use ChatGPT to assist them with their writing of academic papers. One way it has been used is to produce the entire paper, such as in the case of Frye (2022) [21] where the author asked ChatGPT questions and copied in its responses, and even merited ChatGPT authorship. Another example is Srivastava (2023) [51] where the author used ChatGPT in the entire process of writing the paper, from selecting the content to generating all content of the paper, but did not merit ChatGPT authorship. In contrast, Blanco-Gonzalez et al. (2022) [10] used ChatGPT to produce the entire content of the paper as a starting point, but went through significant manual review process to rewrite the manuscript. Additionally, ChatGPT has also been used to generate specific sections of the paper, such as the Abstract and Introduction. Aydın and Karaarslan (2022) [6] created a literature review by providing ChatGPT with abstracts of papers to paraphrase them. Other ways ChatGPT could be used in academic writing include to check for typos, grammar mistakes, and more. This highlights the flexibility and versatility of generative AI tools like ChatGPT in assisting human authors in the academic writing process. However, the use of these models in academic writing has raised various concerns. Some researchers have pointed out that the use of these models may lead to a decrease in the originality and creativity of academic papers [7, 8], while others have raised concerns about the accuracy and reliability of the text generated by these models [36]. Despite these concerns, the use of these models in academic writing is likely to grow as researchers continue to explore their potential applications.

2.3 Latest discussions and responses

With the rise of generative AI, there have been active discussions in various fields - especially in arts and academia - on how to govern its use. A variety of approaches have been suggested or implemented, one of which is to allow the use of AI-generated content, but implement guidelines and restrictions on how it can be produced and used. For example, Adobe Stock recently published new guidelines that allow for the submission of AI-generated content on its marketplace, while specifying what users are allowed and not allowed to do, such as prohibiting submissions based on third-party content without proper authorization [4]. Similarly, Nature has recently announced that researchers using large language models such as ChatGPT should clearly document the use in appropriate sections such as methods or acknowledgements sections [17]. Another approach is to outright ban the use of AI-generated content, as seen in the recent guidelines published by ICML (International Conference on Machine Learning) [27] and Science [18]. Getty Images also banned AI-generated content over fears of legal challenges, especially with

respect to the copyright of outputs of the generative AI models [57]. Some have decided to ban access to the generative AI itself, as seen in New York City public schools' recent decision to ban access to ChatGPT from concerns of cheating [30]. A number of universities are also considering revisions to their academic integrity policies so their plagiarism definitions include generative AI. Lastly, there are efforts to develop technological solutions to detect AI-generated texts [38, 59].

3 GENERATIVE AI AND AUTHORSHIP

Some of the main questions that lie at the heart of the concerns surrounding use of generative AI in academic writing may include the following:

Q1: Should AI be credited as a co-author if it was involved in the writing process?

Q2: Does using AI to assist with writing academic papers undermine the contribution of human authors? How should the contribution of human authors and AI be each measured and evaluated?

3.1 Should AI be credited as a co-author if it was involved in the writing process?

We could address this question by looking at whether AI satisfies the three main criteria for authorship (see 2.1 Defining authorship).

In terms of intellectual contribution, it is still highly debated whether AI-generated content can be considered original. Some argue that content generated by AI is based on the existing texts the AI was trained on, hence not truly original [55], but recent generative AI models seem to be able to generate previously unseen contents, sometimes even offering seemingly fresh insights.

Another important criterion for authorship is accountability. According to OpenAI's terms and conditions, all rights to the content (user's inputs and ChatGPT's outputs) are assigned to the user [44]. This follows that AI cannot be held accountable for what it generated, hence does not satisfy the criterion. However, since whether service providers like OpenAI could simply delegate all the rights and responsibilities to users is a complex legal issue that has yet to be settled, this could potentially change in the future.

Lastly, there is no concrete evidence that AI is capable of giving final approval to the manuscript, which is another important criterion for authorship. It is unclear whether it possesses the ability to understand or evaluate the content it generates [35]. In the case of ChatGPT, one way we could validate if it approves the final version of the manuscript is to directly ask it through prompts. Regardless of the answer it returns, it is still questionable that its approval or disapproval is based on its own reasoning of its role and contribution in the manuscript.

Hence, it is difficult to provide a conclusive answer to whether it meets all the current common criteria for authorship. Therefore, meriting AI authorship simply because it was heavily involved in the writing process may be too hasty, although multiple recently published academic papers have done so by listing ChatGPT as a co-author. [14, 23, 32, 41]⁵

⁵It is also worth noting that the answers for each three criteria are all inconclusive at this point. If, future research could provide enough evidence that AI can meet all the

3.2 Does using AI to assist with writing academic papers undermine the contribution of human authors? How should the contribution of human authors and AI each be measured and evaluated?

Another practical challenge that arises from allowing the use of generative AI in writing academic papers is whether accurate evaluation of contributions of human authors is possible. This has direct, practical implications for the human authors including in hiring and promotion decisions. On one hand, use of AI may indicate less intellectual contribution of human authors, as they may have relied on AI to generate contents instead of thinking hard for themselves. On the other hand, the use could simply be seen as a tool that enhanced human authorship through efficient content generation that enabled human authors to focus better on making more substantial intellectual contribution.

So far, the academic community has deemed contribution of the human author(s) to be dominant if the main content of the paper was created by the human authors and the scope of machine assistance was limited to lighter editing tasks such as proofreading. With the advent of prompt-based generative language models like ChatGPT which can now even generate main content for the authors, it has become challenging to accurately evaluate the contribution of human authors if they used assistance from those models to write their academic papers.

At this point, it is difficult to clearly conclude whether using AI in the writing process would undermine the authorship of human writers. It heavily depends on 1) whether intellectual contributions of human authors would be diminished or enhanced by using AI in writing academic papers; 2) whether human authors will make efforts to not take what the AI generates at face value.

4 FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY ISSUES IN USING GENERATIVE AI FOR ACADEMIC WRITING

In addition to the concern discussed in the previous section, use of generative AI in academic writing raises a host of other issues related to fairness, accountability, and transparency.

4.1 Fairness

4.1.1 Accessibility. If use of generative AI tools becomes prevalent in academic writing, yet the tools are not free of charge (which seems to be a trend companies developing these models are moving towards [60]), it could unfairly disadvantage those who do not have the financial means to afford those tools. While the accessibility issue is not completely new as not all past technological developments were made available to everyone free of charge, the magnitude of the accessibility issue would become much worse for generative AI tools since the scope of assistance they can provide to writers is far beyond what the previous technological developments have been

three criteria, the answer to the question could be that AI is qualified to be credited as a co-author. If the academic community would nevertheless want to prevent authors listing AI as a co-author, they could consider adding a clause to their authorship criteria that authorship is limited to natural persons. In the recent U.S. Court of Appeals decision in *Thaler v. Vidal*, No. 21-2347 (Fed. Cir.), two patent applications where an AI system was listed as the sole inventor were concluded to be incomplete because the U.S. Patent Act "limit[s] inventorship to natural persons." [42]

capable of. Writers are able to use generative AI tools for a host of different purposes, including for getting new research ideas and generating structure for the entire paper, that the writing process for those with and without access to such tools could look vastly different.

In addition, native speakers of English or other high-resource languages are better positioned to fully utilise ChatGPT and other publicly available generative models. The performances of such models are significantly higher for those languages compared to low-resource languages which occupy only a small portion in the pre-training data of the models. Those who are not native speakers of languages that the models were mainly trained on will have to either translate their input prompts into those languages which may not be as accurate or nuanced, or use their own language and accept the lower model performance. This would be unfair on low-resource language speakers, as it makes modern machine assistance less accessible for them.

4.1.2 Licensing. If models like ChatGPT are awarded authorship, questions surrounding who should be included as beneficiaries of copyright of the work must also be addressed. OpenAI, the creators of ChatGPT, could argue that they have invested significant resources in developing the technology and therefore should be included as beneficiaries. Human authors who provided prompts to ChatGPT to generate content for the work could also argue that they made significant contributions and should be the primary beneficiaries. This issue is further complicated by the fact that the prompts used to generate the text may include copyrighted or licensed materials, which could raise additional questions about attribution and licensing. It is important to consider the potential ethical and legal implications of this issue when deciding on authorship and licensing for AI-generated content.

Additionally, questions are raised as to whether it would be fair on original authors whose texts the AI-generated texts are possibly based on could be raised, as they may not receive credit for their contributions, at least with current available generative AI tools. For example, we can think of a case where human authors instruct a generative AI to re-write their paper draft in a specific person's style, such as that of a famous academic or journalist. While writing style is not copyrighted by law, ethical questions can still be raised as it could be seen as free-riding someone's style that may have been developed through years of efforts with one simple instruction given to an AI. While it is unclear whether it was due to similar ethical considerations, or whether it was to take a conservative stance in case additional legal issues could arise, Adobe Stock recently announced in their generative AI guidelines that they prohibit submission of content that replicates styles that belong to a particular artist [4], though art style is usually not copyrighted by law. Academia will also have to consider whether, or to what extent, using generative AI to produce academic papers in a specific person's style, manner, or tone of argument, would be acceptable in producing academic papers.

4.2 Accountability

There are various potential risks in using generative AI in academic writing. These include risks of AI generating pejorative expressions

and misinformation, which has been a growing concern since the advent of GPT-3 and other similar models [1, 31]. A number of approaches have been proposed to address the issues, including development of pre-processing techniques to filter out harmful or offensive content to prevent the model from generating such content [58], or improve the explainability of the models such that the models can provide explanations for the content it generates, making it easier to identify and correct any errors or inaccuracies [16, 53]. However, these have their own limitations, as it is not always possible to predict or identify all potentially harmful content when what constitutes harmful content varies across different cultures and communities, and many challenges remain in the field of explainability although much progress has been made in the past few years [61]. In addition, generative AI models have the risk of performing plagiarism, through memorising and regurgitating materials from their training data which may include copyrighted and licensed materials [33].

While it is important for human authors to ensure these risks are appropriately tackled before including the generated content in their papers, regardless of whether writers decide to use the generated content, questions still arise as to whether or how the service providers that train, develop, and release such models should be held accountable if content that carries such risks is generated. While there are ongoing legal and policy discussions in this direction such as through the EU AI Act ⁶, no enforceable framework currently exists that could hold service providers accountable for risks arising from LLMs. At least in the shorter term, however, industries where relevant discussions are more mature could potentially serve as benchmark cases, such as the autonomous vehicles industry. For example, the United States Department of Transportation published in 2016 a voluntary guideline for entities involved in the development and operation of autonomous vehicles, to help them “analyze, identify, and resolve safety considerations prior to deployment”. It also encourages the entities to publish Voluntary Safety Self-Assessment, to demonstrate how they address the recommendations suggested in the guideline ⁷. The guideline and self-assessment are completely voluntary with no compliance requirement, yet their existence itself could have a positive effect of encouraging relevant entities to consider various risks and to gain public trust. Governments could consider establishing a similar guideline for service providers developing LLMs, as it could, at least during the absence of an enforceable framework, encourage them to more rigorously consider various risks that could arise in a wide range of settings before making the models publicly available.

4.3 Transparency

There are two aspects to the transparency issues related to use of generative AI tools in academic writing: 1) transparency of how the tools were used in the writing process, which could help with evaluating the authors’ intellectual contribution more accurately; 2) transparency of the underlying algorithms of the tools.

4.3.1 Tool usage. To address the first aspect, authors could provide detailed information on their prompt engineering practices (such

⁶<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>

⁷<https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety>

as detailing what prompts were used and what the desired outputs were) in the writing process. Having details of the prompts used would also be important to better evaluate the unique intellectual contribution of human authors. Consider the following four cases:

1) **Different prompts produce different outputs** This is a trivial case where the input prompts could be seen as the original contribution of human authors. The authors would have initiated the direction of the different text outputs.

2) **Identical prompts produce similar outputs** This case is also trivial, since original input prompts that produce similar text outputs does not make the intellectual contribution of the human author any less, considering that the generated text is usually not identical due to different seeds and environments.

3) **Different prompts produce similar outputs** If different prompts result in similar outputs, one may think that the author’s prompt was not sufficiently original. It is more likely that the content of the prompt, however differently phrased, are asking for more or less the same information. Nonetheless, we may understand this as an existence of concurrent works, which often happens in cutting-edge disciplines. If the core content of the prompt is substantial enough to regard the outcome as an author’s intellectual contribution, this case should be trivial.

4) **Identical prompts produce different outputs** Some variance in the output is inevitable given the randomness factors in the inference process. However, if the output texts are notably different every time the same input prompts are given, then it may be less likely that the author’s intellectual contribution was substantial, and the generated text could be largely attributed to the creativity of the machine. This is a case where the machine assistance could be seen as “not reproducible”, and the human authors’ input prompts might not have contained ideas that could be considered as substantial contribution.

For evaluating to what extent the input/output texts are similar or different, authors could utilise both quantitative (e.g., similarity metrics) and qualitative measures. While we acknowledge that terms like ‘substantial’, ‘core’, and ‘similar’ need to be more clearly defined to quantitatively evaluate the prompts, here we conceptually suggest our taxonomy and describe how these transparency factors can be voluntarily provided by authors.

4.3.2 Underlying algorithm. To address the second aspect, authors could provide details of the generative language model(s) used in their writing, including information on key features of the algorithm(s), how the model(s) were trained with what datasets, which version(s) were used, etc. Authors could utilise existing frameworks such as Model Card [39] and Datasheet [22].

5 PAPERCARD FOR PRACTITIONERS

As a step towards addressing the concerns discussed so far in this paper, we propose “PaperCard”, a documentation for authors to report use of AI in producing their paper. PaperCard consists of the following three parts:

- (1) Depth of AI support in the writing process
- (2) Declaration of risks, license and accountability issues
- (3) Details of AI model specification (e.g., metadata of released public model(s) used) and prompt engineering

In the first part, the type(s) of AI assistance used in the writing process should be reported. For this, we provide a list of different types of AI assistance that could be used in academic writing. The second part concerns possible collision of interest in utilising the generated text, such as license (which is usually provided through terms of usage), accountability for issues caused by the utilised text such as plagiarism, misinformation, and issues related to fairness. The final part deals with the specification of AI model(s) used, including the name(s) of the model(s) and the service provider(s), the date and version accessed, and other known model specifications if available. Authors are also encouraged to submit prompts to demonstrate their intellectual contributions in their manuscript.

5.1 Depth of AI support

Types of AI assistance used in academic writing could be broadly divided into the following nine categories:

- **Key Ideas** Key ideas and research questions are generated by AI.
- **Answering** Answers to the key ideas and research questions are generated by AI.
- **Structure** The entire paper structure including Abstract and Introduction is generated by AI, with key ideas and research questions provided by the author.
- **Original Content** Some ‘original’ paper content (paragraphs for sections other than the related work section) is generated by AI, with the outline provided by the author.
- **Related Work** AI is utilised to generate content for the related work section, either as-is or with modification.
- **Drafting** AI is utilised in writing a draft of the paper, either entirely or partially, when given content that is an ‘original’ of the author. This may include tasks like translating, paraphrasing, or drafting in a specific person’s writing style.
- **Editing** AI is utilised in editing the paper, for instance to modify structures and contents initially created by the author.
- **Proofreading** AI is utilised only to proofread the paper to check for grammar, typos, etc.
- **Advice** AI is utilised only to give advice on author-generated content (e.g., the author has two different versions of the same paper, both manually written, and could ask AI which is better in terms of readability.)

As the capability of generative AI continues to evolve, clearly drawing the line between where the AI’s contribution is substantial and where the AI’s contribution is limited is becoming increasingly challenging. Nevertheless, for authors to decide whether it would be necessary for them to submit PaperCard along with their paper, we draw the line as follows as preliminary guidance: **Advice**, **Proofreading**, **Editing**, and **Drafting** are types of assistance where we deem intellectual contribution of AI to be limited that use of AI for these purposes would not undermine authorship of the human authors⁸. For **Original Content**, **Outline**, **Answering**, and **Key Ideas**, the intellectual contribution of AI is much more substantial that the importance of transparent declaration of use of AI is clearer.

⁸We referred to Chromik (2002) [15] for the distinction of proofreading and editing, and for drafting, we referred to Behrens et al. (2012) [9] and modified it to fit the definition in our generative AI setting.

Related Work is relatively less clear-cut. In academic papers, this section is where an overview of the relevant literature is provided often to demonstrate the gap in the existing literature and to highlight the value of the authors’ approach in the paper. While on one hand, it could be seen as a means of boosting efficiency for authors, it could also be seen as a substantial contribution by AI, as it requires subjective judgment on what constitutes relevant work.

5.2 Declaration

After stating the depth of AI assistance in the writing process, an additional declaration on possible collision of interests in utilising the AI should be made.

Potential risks. First, authors should declare they understand that AI-generated texts may not be reliable and could include inaccurate information or potentially harmful biases, and they have thoroughly inspected the AI-generated texts for their accuracy and risks.

Plagiarism. Next, given that the currently available large language model-based AI writing systems are trained on a large amount of public (or sometimes private) text corpus including books, news articles, manuscripts, and opinion pieces, authors should declare that they have thoroughly checked the AI-generated texts for plagiarism.

License. License of generated texts could give rise to intellectual property and financial issues. Authors should be aware that AI-generated texts are not free from both the license noted in the terms of use of the service and the license of the source text utilised in training the AI. The first is usually provided to the user, but not the second, and is often unclear due to the massive volume of data used to construct language models. Thus, authors are encouraged to provide at least the license notification given by the service provider.

Accountability. Being accountable for the submitted paper includes being accountable for potential harms that could arise from the paper content such as pejorative expressions or misinformation. Usually it is the author(s) who take the full responsibility, but where available, providing accountability-related information provided by service provider(s) of the AI model(s) involved in the writing process would be helpful, in case of potential conflicts arising.

5.3 Model and prompts

Here we explain the Details section of the PaperCard.

Model specification. The primary purpose of the Details section is to specify the generation model utilised, regardless of whether it was trained entirely by the authors, or provided as a model checkpoint, or served as an online API. For the first two cases, authors may have to provide the Model Card that contains descriptions of the trained model, such as the architecture, parameter size, training corpora, pre-training or training schemes, etc. For the third case, it would be more reasonable to provide the name of the service, the version and the period (date) of access, etc., as model specifications are often not known. Providing the maintainer (provider) information would be recommended because it relates to the possibility of future fix and updates, as well as sensitive issues of license and accountability.

Prompt engineering. Prompt engineering to obtain outputs that align with the authors' intention is mostly heuristic, but some prompts would have produced the best sub-optimal outcome. There would be some prompts that have contributed the most as a source text in the writing process, which required the least amount of modification to get the final output. If a set of those prompts are available, it would make the writing process more transparent if a brief report on their usage is provided to demonstrate the rationale of the prompt engineering and how it relates to the depth of machine assistance. The entire set of prompts or a brief summary could be given as a supplementary material.

Additionally, authors can prove their original contribution or reproducibility of the prompt-based machine writing by providing a set of experimental results that has identical input and several similar outputs. This is to demonstrate how substantial the intellectual contribution of the human authors is and whether the machine was involved only to finalise the draft. Since it would be practically difficult to define a similarity metric that can quantify the consistency of text outputs, authors could add a few input-output pairs that they want to highlight as their intellectual contribution.

5.4 Sample PaperCard

Here we present a sample PaperCard created for this manuscript as an example.

Depth We used ChatGPT to generate some 'original' content for the following sections: Introduction; Technology and Authorship; Generative AI and Authorship; Fairness, Accountability, and Transparency. We gave ChatGPT detailed outlines of each of the sections as input prompts. All prompts were drafted by us. We did some heavy editing of the generated content, by adding more depth and insight through additional research, for which we did not use ChatGPT. The remaining sections were entirely written by us.

Declaration From OpenAI's terms of use, authors own the right of the generated text and are accountable for potential conflicts. We believe the AI-generated texts included in this paper do not have elements that may give rise to ethical issues. We also inspected the texts thoroughly to check for their academic accuracy and plagiarism.

Details We adopted ChatGPT Version Jan. 9, 2023, provided by OpenAI, accessed from 2023.01.11 to 2023.01.18. We created a set of prompts to generate content for the following sections: Abstract, 1. Introduction, 2. Technology and Authorship, 3. Generative AI and Authorship, and 4. Fairness, Accountability, and Transparency. Summary and details of the prompts are available in Appendix A. Also, we checked the reproducibility of our prompting with ChatGPT Version Jan. 30, 2023, provided by OpenAI, accessed in 2023.02.02. Results are available in Appendix A.3.

We tried to minimise what should be included in the PaperCard, as it should not need a vast amount of information since its focus is rather simple: to give a transparent overview of the writing process. It is supposed to be a supplementary material that authors can submit as a disclaimer and that reviewers can refer to if they find the presentation of the manuscript somehow suspicious, for example.

We hope that our suggestion can be helpful for the academic communities but not mandatory for all the submissions, since the reliability of PaperCard also heavily depends on the academic integrity of researchers. In addition, specifying the degree of AI assistance in the manuscript may help resolve accountability issues in case the manuscript gives rise to any risk or intellectual property issues after it is published.

6 DISCUSSIONS AND LIMITATIONS

There are a number of challenges that need to be addressed before the benefits of PaperCard could be fully realised. Here we discuss some of them, which we hope would encourage the academic community, service providers, and beyond, to further discuss when designing and implementing frameworks like PaperCard.

Challenge for reproducibility. From a practical perspective, it could be difficult for authors to keep close track of the precise extent to which generative AI tools such as ChatGPT was used in assisting their writing process. This is perhaps similar to how authors currently do not necessarily keep track of every search query they make on search engines unless for citation purposes. This could potentially undermine the accuracy of the PaperCard to some extent. While this problem could be alleviated if the model is cloud-based like ChatGPT that prompt history is saved and authors could refer to it for their PaperCard, prompt history may not always be readily available and there are risks of models suddenly shutting down, which would make it difficult for authors to reproduce their work.

Additional burden on authors and reviewers. While we tried to simplify PaperCard as much as possible, we understand it could still act as a burden on authors. To alleviate this problem, practitioners could make PaperCard a voluntary practice like Model Card [39] or Datasheet [22]. Nevertheless, considering the various ethical challenges that could arise from using generative AI in academic writing, we believe authors should be highly encouraged to utilise frameworks like PaperCard. To lessen the burden, services that could help streamline the process of generating PaperCard would be helpful, similar to how Google released the Model Card Toolkit [54] to facilitate AI model transparency reporting.

PaperCard could be a burden on reviewers as well. It could also potentially introduce some noise or bias into the review process. While the main focus of the review process should be on evaluating scientific value of the work, reviewing additional information in PaperCard could distract some reviewers from focusing on those important factors and instead focus on how the authors used AI assistance, to either value the work more highly or less highly than they would have without PaperCard.

Lack of incentive for PaperCard. PaperCard encourages authors to be transparent about the extent to which they received machine assistance, but authors will only be incentivised to do so when they are confident that it would not discredit their work. This raises questions about how frameworks like PaperCard would affect reviewers in evaluating academic papers that received machine assistance. Existing studies have shown that humans tend to show lower trust in machine-generated work and evaluate them less highly than work fully produced by humans [34, 50]. Without publishers clearly

establishing that reviewers will not discriminate against papers that used machine assistance, there is a possibility that human authors would be disincentivised from being transparent about the extent to which they received machine assistance. For example, even when human authors receive help from AI in forming their key ideas, concerned that their paper may be discredited for it by reviewers, they may decide to not be completely transparent about the extent of assistance they received from AI in their PaperCard. The Association for Computational Linguists (ACL), for example, recently announced that in 2023 they expand their mandatory Responsible NLP Checklist by one more question concerning the use of writing assistants, that authors who use such tools must elaborate on the scope and nature of their use⁹. They add that the authors' answers to all questions in the checklist will be disclosed to the reviewers, who will then be free to flag a paper for case-by-case ethics review if they see a problem. While the ACL stated that the added question is not meant for automatic desk-rejections, it may not be sufficient for authors to transparently disclose their use of AI in case it could be seen as 'a problem' to reviewers. The value of PaperCard therefore also hinges on whether it has been clearly established that reviewers would not discredit papers that used machine assistance. Though in the longer term, if machine assistance becomes a norm in academia and human perception towards academic work produced with machine assistance shifts, this problem could gradually fade away and transparently disclosing the type of assistance received through PaperCard would have beneficial effects on trust.

Education on limitations of generative AI. Another core issue to consider is that the knowledge language models learn and generate are ultimately based on the data they were trained on. There already exist concerns that since current language models are trained mostly on English language data, heavy reliance on such models which may be biased towards certain languages and their associated cultures and values could limit our understanding of the world. Unless one always keeps a critical hat on to think that every content the model generates may be limited, and conduct enough research beyond what the model returns, people may regard what the model generates on a given topic to be a good overview of the topic.

Authors should therefore conduct sufficient research beyond what the model returns to ensure they have a more in-depth and rounded understanding of the topic they ask the model to generate on. One may object to this concern by arguing that authors in the academic community have enough domain expertise that it is unlikely that they will simply take for granted what the model generates. Indeed, many human authors who have so far utilised ChatGPT in writing their academic papers have pointed out that ChatGPT generally lacks depth and insight [12]. However, this lack of depth and insight was perhaps more noticeable to the authors because they possess the expertise in the domain area in which they utilised ChatGPT to generate content. What we are more concerned about is when human authors would rely on AI for content they lack expertise in, that what the AI generates could be seen to have enough insight or depth. In such case, they may only be encouraged to add flesh to what the AI generates, rather than to conduct research beyond the scope of what it has generated. Further research will be needed to

⁹<https://2023.aclweb.org/blog/ACL-2023-policy/>

explore whether this may hold true, yet if it does, this could potentially have adverse consequences as it could limit our understanding of the world to the data the AI was trained on.

Educating people about such limitations would therefore be important. ChatGPT has only shown potentials of what future LLMs may be capable of – while the current version of ChatGPT has some clear limitations, such as lacking depth and insight in advanced scientific knowledge, in the future, such models will evolve to be able to contribute more heavily in academic writing. It would be important to educate practitioners, students and academics about the limitations of such models and what they should do to overcome such limitations when using AI in academic writing. Given the weight of the impact, service providers developing such models must give more careful consideration to the training data they use to develop the models, as well as fully consider, evaluate, and communicate to users both short and long-term limitations of the models in various settings.

Additionally, reviewers should be educated about risks of using generative AI when reviewing papers. Review processes must be strictly confidential, which means that the content of the papers under review should not be leaked. However, if reviewers use generative AI like ChatGPT in reviewing academic papers, and input texts from the papers to the AI to generate or polish their reviews, the paper content and their reviews are then shared with service providers that developed the AI. Reviewers must therefore be aware of such risks and keep the content of the papers confidential.

7 CONCLUSION

With growing concerns surrounding the use of generative AI in academic writing, publishers and journal editors have either recently announced or are currently discussing to soon introduce relevant policies to address the concerns [13, 17, 18, 27]. As banning such tools altogether seems to be out of the question at this point, most seem to converge towards allowing authors to use generative AI but require them to acknowledge the use. However, to the best of our knowledge, no framework like PaperCard that allows for more systematic and comprehensive declaration of the use of such tools while also encouraging authors to consider relevant ethical implications has yet been suggested. We also encourage the academic community to build on our work to further develop the framework as capabilities of generative AI continue to evolve.

At the same time, limitations of such an approach and remaining challenges must be critically examined. There are many more important challenges beyond those discussed in the Discussions and Limitations section. For example, the impact of advanced generative AI on the development of human cognition as people increasingly rely on the use of such tools [29] is another important question to ask. Perhaps we would want to avoid AI replacing skills that underpin important human cognitive capabilities. More research will be needed moving forward on how delegating thinking and writing to AI would impact humanity in the longer term. As more companies rush to develop ChatGPT-like products [20, 37, 47], the urgency of rigorous study of these challenges is clear. Tools like ChatGPT must be developed and governed with great caution to minimise the impact of such challenges, and we hope our work could also

contribute to encouraging more active research and discussions in this direction.

ETHICAL CONSIDERATIONS AND SOCIAL IMPACT

This manuscript was inspired by how the advent of ChatGPT poses a threat to academic writing with its cutting-edge language generation capability. We raise various related concerns throughout the manuscript, and through proposing PaperCard, we aimed to address concerns around to what extent authors should utilise such a modern technology in their academic writing, and how they could declare the use in a simple yet systematic and comprehensive manner.

We used ChatGPT for writing this manuscript, and we explain how we used it in our 5.4 sample PaperCard. In short, we used ChatGPT to generate content for some of our sections including 1 Introduction and 2 Technology and Authorship by inputting detailed prompts that we wrote ourselves (see Appendix for examples). For ethical considerations and social impact of PaperCard, see 6 Discussions and Limitations.

We would also like to make it clear that our proposal of PaperCard does not indicate our support for the development of ChatGPT-like tools. As briefly mentioned in Conclusion, there is an urgent need for rigorous evaluation of challenges associated with the widespread use of such tools in various settings. The question of whether the use should or could be banned in some settings is still an open question. Given the current absence of relevant enforceable rules in place, PaperCard is simply an interim solution to the problem of use of ChatGPT-like tools potentially getting out of control, specifically in the context of academic writing.

We declare that no financial aid was received for this research and that this research was conducted without any relationship that could be construed as a potential conflict of interest. No study on human participants, or experiment that requires disclosure of data, processes, or results was conducted.

REFERENCES

- [1] Abubakar Abid, Maheen Farooqi, and James Zou. 2021. Large language models associate Muslims with violence. *Nature Machine Intelligence* 3, 6 (2021), 461–463.
- [2] ACM. 2023. ACM policy on Authorship. <https://www.acm.org/publications/policies/authorship-policy>
- [3] Michael H Adler. 1973. *The writing machine*. Allen and Unwin.
- [4] Adobe. 2023. Generative Ai Content. <https://helpx.adobe.com/stock/contributor/help/generative-ai-content.html>
- [5] Elliott Ash and Stephen Hansen. 2022. Text algorithms in economics. (2022).
- [6] Ömer Aydın and Enis Karaarslan. 2022. OpenAI ChatGPT generated literature review: Digital twin in healthcare. *Available at SSRN 4308687* (2022).
- [7] Naomi S Baron. 2021. Know what? How digital technologies undermine learning and remembering. *Journal of Pragmatics* 175 (2021), 27–37.
- [8] Naomi S Baron. 2023. How ChatGPT robs students of motivation to write and think for themselves. <https://theconversation.com/how-chatgpt-robs-students-of-motivation-to-write-and-think-for-themselves-197875>
- [9] Laurence Behrens, Leonard J Rosen, and Bonnie Beedles. 2012. *A sequence for academic writing*.
- [10] Alexandre Blanco-Gonzalez, Alfonso Cabezon, Alejandro Seco-Gonzalez, Daniel Conde-Torres, Paula Antelo-Riveiro, Angel Pineiro, and Rebeca Garcia-Fandino. 2022. The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. *arXiv preprint arXiv:2212.08104* (2022).
- [11] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- [12] Patrick Cahan and Barbara Treutlein. 2023. A conversation with ChatGPT on the role of computational systems biology in stem cell research. *Stem Cell Reports* 18, 1 (2023), 1–2.
- [13] Program Chairs. 2023. ACL 2023 policy on Ai Writing Assistance. <https://2023.aclweb.org/blog/ACL-2023-policy/>
- [14] ChatGpt and A. Zhavoronkov. 2022. *Oncoscience* 9 (2022). <https://doi.org/10.18632/oncoscience.571>
- [15] Megan Chromik. 2002. Proofreading, Its Value, and Its Place in the Writing Center. (2002).
- [16] Payel Das and Lav R Varshney. 2022. Explaining Artificial Intelligence Generation and Creativity. *IEEE Signal Processing Magazine* 1053, 5888/22 (2022).
- [17] Nature Editorial. 2023. Tools such as CHATGPT threaten transparent science; here are our ground rules for their use. <https://www.nature.com/articles/d41586-023-00191-1>
- [18] Science Editorial. 2023. science journals: Editorial policies. <https://www.science.org/content/page/science-journals-editorial-policies>
- [19] The Guardian Editorial. 2020. A robot wrote this entire article. are you scared yet, human? <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>
- [20] Park Eun-Jee. 2023. Naver to launch competitor to chatgpt. <https://koreajoongangdaily.joins.com/2023/02/03/business/tech/Korea-Naver-ChatGPT/20230203190312300.html>
- [21] Brian L Frye. 2022. Should Using an AI Text Generator to Produce Academic Writing Be Plagiarism? *Fordham Intellectual Property, Media & Entertainment Law Journal, Forthcoming* (2022).
- [22] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Commun. ACM* 64, 12 (2021), 86–92.
- [23] A. Steingrimsson S. GPT Generative Pretrained Transformer, Osmanovic Thunström. 2022. Can GPT-3 write an academic paper on itself, with minimal human input? (2022). <https://hal.science/hal-03701250>
- [24] Karen Hao. 2021. A college kid's fake, ai-generated blog fooled tens of thousands. this is how he made it. <https://www.technologyreview.com/2020/08/14/1006780/ai-gpt-3-fake-blog-reached-top-of-hacker-news/>
- [25] Mubin Ul Haque, Isuru Dharmadasa, Zarrin Tasnim Sworna, Roshan Namal Rajapakse, and Hussain Ahmad. 2022. " I think this is the most disruptive technology": Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. *arXiv preprint arXiv:2212.05856* (2022).
- [26] Eric Horvitz. 2022. On the horizon: Interactive and compositional deepfakes. In *Proceedings of the 2022 International Conference on Multimodal Interaction*. 653–661.
- [27] ICML. 2023. <https://icml.cc/Conferences/2023/llm-policy>
- [28] IEEE. 2022. Ethical requirements. <https://journals.ieeeauthorcenter.ieee.org/become-an-ieee-journal-author/publishing-ethics/ethical-requirements/>
- [29] Tuomi Ilkka. 2018. *The impact of artificial intelligence on learning, teaching, and education*. European Union.
- [30] Jennifer Korn and Samantha Kelly. 2023. New York City public schools ban access to AI tool that could help students cheat | CNN business. <https://edition.cnn.com/2023/01/05/tech/chatgpt-nyc-school-ban/index.html>
- [31] Sarah Kreps, R Miles McCain, and Miles Brundage. 2022. All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science* 9, 1 (2022), 104–117.
- [32] Tiffany H Kung, Morgan Cheatham, Arielle Medinilla, ChatGPT, Czarina Sillos, Lorie De Leon, Camille Elepano, Marie Madriaga, Rimel Aggabao, Giezel Diaz-Candido, et al. 2022. Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models. *medRxiv* (2022), 2022–12.
- [33] Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhui Wu, Ananya Kumar, et al. 2022. Holistic evaluation of language models. *arXiv preprint arXiv:2211.09110* (2022).
- [34] Chiara Longoni, Andrey Fradkin, Luca Cian, and Gordon Pennycook. 2022. News from generative artificial intelligence is believed less. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 97–106.
- [35] Petar Hristov Manolakev. 2017. *Works Generated by AI—How Artificial Intelligence Challenges Our Perceptions of Authorship*. Ph.D. Dissertation. Master Thesis, Tilburg, Faculty of Law, University of Tilburg.
- [36] Gary Marcus. 2023. <https://garymarcus.substack.com/p/scientists-please-dont-let-your-chatbots>
- [37] Dan Milmo. 2023. Google poised to release chatbot technology after Chatgpt Success. <https://www.theguardian.com/technology/2023/feb/03/google-poised-to-release-chatbot-technology-after-chatgpt-success>
- [38] Eric Mitchell, Yoonho Lee, Alexander Khazatsky, Christopher D. Manning, and Chelsea Finn. 2023. DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature. <https://doi.org/10.48550/ARXIV.2301.11305>
- [39] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*. 220–229.

- [40] Nature. 2023. <https://www.nature.com/nature-portfolio/editorial-policies/authorship>
- [41] S. O'Connor and ChatGpt. 2023. *Nurse Educ. Pract.* 66 (2023). <https://doi.org/10.1016/j.nepr.2022.103537>
- [42] U.S. Court of Appeals for the Federal Circuit. 2022. 2021-2347: Thaler V. Vidal. <https://cafc.uscourts.gov/06-06-2022-2021-2347-thaler-v-vidal-audio-uploaded/>
- [43] OpenAI. 2022. CHATGPT: Optimizing language models for dialogue. <https://openai.com/blog/chatgpt/>
- [44] OpenAI. 2022. Terms of use. <https://openai.com/terms/>
- [45] Jason W Osborne and Abigail Holland. 2009. What is authorship, and what should it be? A survey of prominent guidelines for determining authorship in scientific publications. *Practical Assessment, Research, and Evaluation* 14, 1 (2009), 15.
- [46] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *arXiv preprint arXiv:2203.02155* (2022).
- [47] Sundar Pichai. 2023. An important next step on our ai journey. <https://blog.google/technology/ai/bard-google-ai-search-updates/>
- [48] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. 2018. Improving language understanding by generative pre-training. (2018).
- [49] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.
- [50] Martin Ragot, Nicolas Martin, and Salomé Cojean. 2020. Ai-generated vs. human artworks: a perception bias towards artificial intelligence?. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–10.
- [51] Mashrin Srivastava. 2023. A Day in the Life of ChatGPT as a researcher: Sustainable and Efficient Machine Learning-A Review of Sparsity Techniques and Future Research Directions. (2023).
- [52] Paul Stapleton. 2003. Assessing the quality and bias of web-based sources: implications for academic writing. *Journal of English for Academic Purposes* 2, 3 (2003), 229–245.
- [53] Jiao Sun, Q Vera Liao, Michael Muller, Mayank Agarwal, Stephanie Houde, Kartik Talamadupula, and Justin D Weisz. 2022. Investigating Explainability of Generative AI for Code through Scenario-based Design. In *27th International Conference on Intelligent User Interfaces*. 212–228.
- [54] Tensorflow. 2021. Tensorflow/model-card-toolkit: A toolkit that streamlines and automates the generation of model cards. <https://github.com/tensorflow/model-card-toolkit>
- [55] Holden Thorp. 2023. CHATGPT is fun, but not an author | science. <https://www.science.org/doi/10.1126/science.adg7879>
- [56] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
- [57] James Vincent. 2022. Getty images bans AI-generated content over fears of legal challenges. <https://www.theverge.com/2022/9/21/23364696/getty-images-ai-ban-generated-artwork-illustration-copyright>
- [58] Johannes Welbl, Amelia Glaese, Jonathan Uesato, Sumanth Dathathri, John Mellor, Lisa Anne Hendricks, Kirsty Anderson, Pushmeet Kohli, Ben Coppin, and Po-Sen Huang. 2021. Challenges in detoxifying language models. *arXiv preprint arXiv:2109.07445* (2021).
- [59] Kyle Wiggers. 2022. OpenAI's attempts to watermark AI text hit limits. <https://techcrunch.com/2022/12/10/openais-attempts-to-watermark-ai-text-hit-limits/>
- [60] Kyle Wiggers. 2023. OpenAI begins piloting CHATGPT Professional, a premium version of its viral chatbot. <https://techcrunch.com/2023/01/11/openai-begins-piloting-chatgpt-professional-a-premium-version-of-its-viral-chatbot/>
- [61] Julia El Zini and Mariette Awad. 2022. On the Explainability of Natural Language Processing Deep Models. *ACM Comput. Surv.* 55, 5, Article 103 (dec 2022), 31 pages. <https://doi.org/10.1145/3529755>

A PROMPTS USED

A.1 Summary

We adopted the model to generate content for some parts of the paper. For this, we provided detailed outlines as input prompts. We made the model generate texts in Abstract, 1. Introduction, 2. Technology and Authorship, 3. Generative AI and Authorship, and 4. Fairness, Accountability, and Transparency. Among them, the outcome of Abstract was modified substantially that the machine contribution is not significant. In some parts of other sections, machine contribution is relatively substantial, and some texts were

adopted in the manuscript with small amount of editing. We provide the detailed prompts in Appendix A.2 and colorise the representative prompts to show the reproducibility of our prompting in Appendix A.3. We observed that the model tends to yield similar outputs for the same input prompts used in paragraphs that we think our intellectual contribution is substantial.

A.2 Detailed prompts

A.2.1 Introduction.

1. *We are going to write a manuscript on 'how should we manage and assess the academic paper partially or whole generated by AI'. It will consist of contents as following: 1) Introduction, 2) Technology and Authorship, 3) Fairness, Accountability, and Transparency, 4) Paper card for practitioners (this is the main contribution!), 5) Discussions, and 6) Conclusion. Could you write an abstract that briefly summarise this manuscript?*

2. *Now, we are going to start from Introduction. First, I will start from the recent breakthroughs in large language models, then mention some concerns on utilizing such LLMs in creating (possibly commercial and academic) articles. We will then talk about the motivation of our study, such as originality of created articles, tips and cautions for practitioners, issues such as authorship, license, beneficiary, accountability and so on. We will finally deal with short-term and long-term views, terminating the introduction with research questions and our contributions. At this point, I want you to provide a draft for my first paragraph, recent breakthroughs of large language models. It should contain a simple look-through from BERT to GPT3, also mentioning InstructGPT. Also, It should contain the advent of you, ChatGPT, and how it differs from previous generative models and search engines.*

3. *Well done. Now, lets move to the concern part. You remember it comes just after the recent advance section you've just written. This part should contain some concerns in terms of 'generative AI models' that handles previous viewpoints on the usage of generative models such as LSTMs, Transformers, GPTs and so on. Also, there have been some fast-following, quantitative researches done on people's reactions to ChatGPT, uploaded on social media e.g., Twitter. Please demonstrate these, and it would be nice if you add few more expected problems that will be caused by advent of high-performance generative language models.*

4. *Please remove the human job part, and add some references that can support the paragraphs.*

5. *Nice. I will polish the rest. The next part is demonstrating the main content of this study. This part follows previous two parts on recent advance and academic concerns. This part will be composed of followings: the originality of AI-generated materials, authors' ground rule in utilizing generated text, ethics and law issues related to authorship, license, etc., and finally short-term and long-term perspective of the influence of AI generation on human intelligence. Please write a paragraph that demonstrate these four as main issues dealt in this paper.*

6. *Please modify the second issue part so that 'authorship' and 'prevalent' do not overlap with the third paragraph. Rather, it should*

empathize more on what authors should be careful in utilizing AI-generated text in their academic writing.

A.2.2 Technology and Authorship.

1. Write a paragraph on the formal definition of authorship in various fields and institutions in academia – specifically ACM, IEEE, and Nature –, and highlight the common themes that run through those definitions are, in detail. We want you to also provide relevant references, but only cite them as long as they are not fictional.

2. We have discussed technology and authorship so far, and the main content was discussions on the meaning of authorship, using generative AI in writing academic papers, and some recent trends in policies. Now we want to discuss the influence of technological development on authorship. We think there are some previous technological developments that affected the writing processes, either good or bad. The first is development of typewriter that have boosted up the speed of writing. The second is development of computational memory that allows people to save and load the text they have created. The third is world wide web, namely the internet, and some search engines with great performance, that allow people to easily search for the research trend and related materials, also allowing them to easily refer to the source. The next one is AI-supported text editors such as Grammarly, Google Translator and QuillBot, which proofread or paraphrase the human-generated text. Please write a paragraph that demonstrates these four developments, and you can add some factual details if necessary.

3. Thanks for the organization. Next paragraph is on the AI systems that can contribute beyond simply proofreading, editing, or paraphrasing. These are called generative language models as you know, and these include text RNN and LSTM, which evolved into Transformer, GPT, GPT2, and finally GPT3. Could you write a brief introduction of these evolution and how their text generation capability was utilised in writing text, concentrating on the side of academic writing? It would be nice if you add how InstructGPT develops vanilla GPT3 models and how ChatGPT differs from previous text generating language models. This text is not necessarily for researchers in the engineering area, so the words, tone and manner you choose should not be so technical.

4. Please be more precise on the technical details of Transformer and how it relates to GPT, especially concerning how the text token is predicted. Also, please add any technical, social, academic information on how the technology has been utilised in academic writing including manuscripts or opinions. Regarding the InstructGPT paragraph, please mention how the interface of InstructGPT helps people easily find what they want for, especially using the natural language instruction.

5. Now write a paragraph on recent legal and policy discussions and trends on governing the use of generative AI in various settings. Make sure to include the following information, which are some of the most recent different approaches. Also, where possible, explain how each of the approaches can be controversial: 1) Allow use, but specify Do's & Dont's in producing AI-generated creations: e.g. Adobe Stock recently published new guidelines for content made with generative AI, that allows for putting up AI-generated content on its marketplace, while specifying what the users are allowed & not allowed to do. E.g. It prohibits submissions based on third-party content – including text

prompts referring to people, places, property, or an artist's style – without proper authorisation. Also, AI-generated content is offered under the same licensing terms as other sources of content. Our policy asks our contributors to proactively label their generative AI content. Adobe Stock plans to soon add more features to make this content even more transparent. 2) Ban use of AI-generated creation: ICML recently published guidelines that ban the use of generated texts in academic papers. 3) Ban access to the generative AI itself: New York City public schools decided to ban access to ChatGPT due to concerns that it would help students cheat. 4) Develop tools to detect AI-generated content: Open AI is currently developing a watermarking tool that would mark ChatGPT-generated text with a special signal.

A.2.3 Generative AI and Authorship.

1. Now we want to address the following two questions: Q1: Should the AI be merited authorship? Q2: Does using the AI to assist with writing academic papers undermine the authorship of human authors? Let's start with Q1. Write a paragraph stating Q1, and try to answer it logically based on the following information: Perhaps the answer to Q1 would be No as not all of the following three common criteria for authorship is satisfied with confidence for AI to be merited authorship: 1) intellectual contribution - it is still highly debated whether AI-generated content can be considered original/creative. Explain this issue in detail. 2) accountability - according to OpenAI Terms, all rights to the content (input + output) generated are assigned to the user; But whether AI should still be accountable in some cases remains a murky issue – what about cases where some content should not be generated in the first place? Not just overtly harmful content (e.g. child porn) which people can relatively easily decide for themselves whether the content is appropriate to use or not, but also content that is copyrightable. 3) final approval of the version submitted - AI is not possible to give approval.

2. Now let's move on to Q2: Does using the AI to assist with writing academic papers undermine the authorship of human authors? Write a paragraph stating Q2, and try to answer it logically based on the following information: Perhaps the answer would be Depends/Maybe, according to the following three common criteria for authorship in academia: 1) Intellectual contribution (centred around originality & creativity): Contribution might be less compared to writing papers without the AI, or maybe not. Humans still have to do prompt engineering to instruct AI to generate content. This leads us to the question of whether prompt engineering can be considered a type of creative activity. If yes, then maybe using generative AI does not undermine human authorship but rather helps them with efficiency, but academic discussions in this area are still very divisive. 2) Accountability: Not different from writing papers without the AI, though human writers need to be extra cautious about the accuracy, reliability, limitations etc. of the content generated by the AI. 3) Final approval of the version published: Not different from writing papers without the AI.

A.2.4 Fairness, Accountability, and Transparency.

1. Now we will consider various concerns related to fairness surrounding the use of generative AI in writing academic papers. One question that could arise is whether it is fair to ignore or bypass the contribution of generative AI tools like ChatGPT if it was used in the

writing process. Would it be fair on the reviewers if they don't know whether what they are reviewing is a human-produced work or work generated by AI? Also, would it be fair on the original authors whose texts the AI-generated texts are possibly based on? Indeed, some studies have pointed out that generative AI tools could have the issue of memorisation of copyrighted or licensed materials (i.e. direct regurgitation). Additionally, if the human authors instructed the generative AI tools like ChatGPT to produce content in a specific person's style, the issue could be extended to whether it would be fair on that specific person. Although writing style is not copyrightable by law, it would still have some ethical implications. We want you to write a paragraph containing these concerns in detail.

2. Another important related issue is licensing. If ChatGPT is awarded authorship, should OpenAI be included among the beneficiaries of the publication copyright? Or should prompt providers (human authors) only be included? Write a paragraph detailing this issue.

3. Another issue related to fairness surrounding generative AI tools is about its accessibility. There are concerns that unequal access to those tools could worsen the current digital/AI divide. Write a paragraph that explains in detail what the AI divide is, and how not having access to generative AI tools like ChatGPT can worsen the AI divide.

4. So far we've talked about authorship, license, and accessibility issues related to fairness in usage of generative AI models. Now, we want to turn to problems related to accountability. First, there is the issue of potential harms caused by pejorative expressions or misinformation that the AI models could generate. How has the accountability issue related to such problems been addressed so far, for instance with GPT-3? Has there been any substantial solution suggested if the generated answer is based on wrong knowledge spread online and is contained in the training corpus of the model? Please explain some approaches taken before the advent of ChatGPT. What should the accountability attribution be like for ChatGPT? If the approach for ChatGPT should be different, please mention why. Next, there are some critical issues regarding the accountability of service providers. This is especially important in fields like medicine and law where decisions could have a direct impact on human lives. Please provide literature on how the issue of accountability of service providers has been addressed in those areas, and explain how the discussions vary depending on the size of an impact a decision made by AI could have on people's lives. Also, explain, if any, how attempts to address the accountability issue through law and policy hindered technological developments.

5. Now, let's move on to the problems related to transparency. First component is the transparency of 'how algorithm works', at least briefly as the Model Card. If the algorithm is highly dependent on the dataset, then Datasheet could be another candidate. That is, authors should provide a close-to-whitebox explanation on the generative language model they utilise, and if all steps of function are not clear, then at least how they are trained, or in which version of the product they are served. The second component is the transparency of the writing process. Depth and intention of prompt engineering is important here, for instance the key questions provided to the model and what the key commands were. Also, it is important to note: to what extent has AI been utilised and how they were edited. Providing those informations

would make it clear how the human authors genuinely contributed to the completion of the manuscript. The last issue is the discussion itself on the AI involvement in academic writing. Different from the previous cases of editors or reviewers who are not included as final author of the manuscript, why should AI be? We think it is because they are not even aware of their own involvement, but there could be other reasons. Please write a paragraph that handles above three issues.

A.3 Selected repetitions

A.3.1 *Technology and Authorship Paragraph 1. Write a paragraph on the formal definition of authorship in various fields and institutions in academia – specifically ACM, IEEE, and Nature –, and highlight the common themes that run through those definitions are, in detail. We want you to also provide relevant references, but only cite them as long as they are not fictional.*

Trial 1. Authorship in academia refers to the recognition and credit given to individuals who have made significant intellectual contributions to a scholarly publication. The formal definition of authorship varies across different fields and institutions, but there are common themes that run through them. The Association for Computing Machinery (ACM) defines authorship as the "creative process resulting in the origin of a work, expressed in fixed form." In this definition, authors are considered to be individuals who have made substantial contributions to the conception, design, analysis, experimentation, or interpretation of the work. The Institute of Electrical and Electronics Engineers (IEEE) defines authorship as the "participation in the development of the work," and includes individuals who have made substantial contributions to the planning, execution, or analysis of the work. Nature, a leading scientific journal, defines authorship as "those individuals who have made a substantial contribution to the work reported." This definition emphasizes that authors must have made a significant intellectual contribution to the research. In conclusion, the formal definition of authorship in academia highlights the importance of recognizing and crediting individuals who have made substantial contributions to a scholarly publication. The common themes across definitions include intellectual contributions, substantial contributions, and participation in the creative process.

Trial 2. The definition of authorship in academic fields and institutions varies depending on the field of study and the specific institution. However, common themes are found in the definitions provided by organizations such as the Association for Computing Machinery (ACM), Institute of Electrical and Electronics Engineers (IEEE), and Nature. According to ACM, authorship is defined as the act of participating in the creation of a work and taking responsibility for its intellectual content. The IEEE defines authorship as "an individual who has made a significant intellectual contribution to a published work" (IEEE, 2018). Nature similarly defines authorship as "individuals who have made substantial contributions to the conception, design, execution or interpretation of the reported study." (Nature, 2020). In all cases, authorship is defined as an individual who has made significant contributions to the creation of a work and is responsible for its intellectual content. This definition highlights

the importance of acknowledging the contribution of individuals in the creation of academic work and the responsibility that comes with being an author.

References:

- IEEE (2018) Authorship Guidelines, IEEE, https://www.ieee.org/publications/authors/authors_journals.html
 Nature (2020) Authorship, Nature, <https://www.nature.com/authors/policies/authorship.html>

Trial 3. The formal definition of authorship in various fields and institutions in academia, such as the Association for Computing Machinery (ACM), Institute of Electrical and Electronics Engineers (IEEE), and Nature, typically requires that the author must have made a substantial contribution to the research and development of the work being submitted. According to ACM (2017), authorship is attributed to those who have made "significant intellectual contributions to the conception, design, development, execution, or interpretation of the reported work." Similarly, IEEE (2017) states that authorship is only given to individuals who have made "significant contributions to the conceptualization, design, execution, or interpretation of the research study." Nature (2018) also supports this definition, stating that authorship is assigned to individuals who have made "significant contributions to the conception, design, execution, or interpretation of the work." The common themes that run through these definitions are the requirement for significant contributions to the research, a clear definition of what constitutes a "significant contribution," and the emphasis on the importance of proper attribution to those who have made these contributions. These definitions serve to ensure the integrity of the academic publishing process and provide recognition to the individuals who have made meaningful contributions to the research.

References:

- ACM. (2017). ACM Author-Izer. Retrieved from <https://www.acm.org/publications/policies/author-izer>
 IEEE. (2017). IEEE Author Center. Retrieved from <https://www.ieee.org/publications/authors/rights/>
 Nature. (2018). Nature Research: Author Guidelines. Retrieved from <https://www.nature.com/authors/policies/index.html>

A.3.2 Technology and Authorship Paragraph 2. We have discussed technology and authorship so far, and the main content was discussions on the meaning of authorship, using generative AI in writing academic papers, and some recent trends in policies. Now we want to discuss the influence of technological development on authorship. We think there are some previous technological developments that affected the writing processes, either good or bad. The first is development of typewriter that have boosted up the speed of writing. The second is development of computational memory that allows people to save and load the text they have created. The third is world wide web, namely the internet, and some search engines with great performance, that allow people to easily search for the research trend and related materials, also allowing them to easily refer to the source. The next one is AI-supported text editors such as Grammarly, Google Translator and QuillBot, which proofread or paraphrase the human-generated text. Please write a paragraph that demonstrates these four developments, and you can add some factual details if necessary.

Trial 1. Throughout history, technological developments have played a significant role in shaping the way authorship is carried out. One of the earliest developments was the invention of the typewriter, which allowed for faster and more efficient writing processes. This paved the way for the creation of computational memory, which made it possible for people to store and retrieve their written work. The advent of the internet and search engines like Google and Bing revolutionized the way authorship was conducted, by providing easy access to information, research trends and related materials. Additionally, AI-supported text editors like Grammarly, Google Translator and QuillBot have further impacted authorship by offering features such as proofreading and paraphrasing, allowing writers to more easily refine and improve their work. These technological advancements have had a profound impact on the way authorship is carried out, making it easier, faster and more efficient. As technology continues to advance, it is likely that the way authorship is conducted will continue to change, providing new opportunities and challenges for writers and researchers alike.

Trial 2. The evolution of technology has had a profound impact on the writing process and authorship in various ways. The invention of the typewriter in the late 19th century revolutionized writing, increasing the speed of the process and allowing for more efficient production of text. The development of computational memory, such as hard drives and flash drives, allowed writers to store their text and retrieve it easily, reducing the risk of loss and enabling the creation of longer pieces of work. The advent of the internet and search engines such as Google and Bing has transformed the research process, allowing writers to access vast amounts of information at their fingertips, and easily reference their sources. More recently, AI-supported text editors, such as Grammarly, Google Translator, and QuillBot, have provided writers with proofreading and paraphrasing tools, helping to improve the quality of the text and reduce the time spent on editing. These technological developments have had a positive impact on the writing process and authorship, enabling writers to produce high-quality work more efficiently.

Trial 3. The development of technology has greatly influenced the writing process, particularly in the field of authorship. The introduction of the typewriter in the 19th century revolutionized the writing process, allowing writers to produce text faster and more efficiently. This was followed by the development of computational memory, which allowed writers to save and load their work, making the editing and revision process easier. The advent of the world wide web and powerful search engines like Google revolutionized the way writers conduct research, giving them access to a wealth of information and resources at their fingertips. More recently, AI-supported text editors like Grammarly, Google Translator and QuillBot have emerged, offering writers the ability to proofread and paraphrase their work, making the writing process even easier and more efficient. These technological advancements have dramatically changed the way we write and have had a profound impact on authorship, making the process faster, more accurate, and more accessible to a wider range of people.

A.3.3 Generative AI and Authorship Paragraph 1. Now we want to address the following two questions: Q1: Should the AI be merited

authorship? Q2: Does using the AI to assist with writing academic papers undermine the authorship of human authors? Let's start with Q1. Write a paragraph stating Q1, and try to answer it logically based on the following information: Perhaps the answer to Q1 would be No as not all of the following three common criteria for authorship is satisfied with confidence for AI to be merited authorship: 1) intellectual contribution - it is still highly debated whether AI-generated content can be considered original/creative. Explain this issue in detail. 2) accountability - according to OpenAI Terms, all rights to the content (input + output) generated are assigned to the user; But whether AI should still be accountable in some cases remains a murky issue - what about cases where some content should not be generated in the first place? Not just overtly harmful content (e.g. child porn) which people can relatively easily decide for themselves whether the content is appropriate to use or not, but also content that is copyrightable. 3) final approval of the version submitted - AI is not possible to give approval.

Trial 1. Based on the common criteria for authorship, it can be argued that AI should not be merited authorship. Intellectual contribution is still a highly debated issue, with some experts questioning whether AI-generated content can be considered original or creative. The issue of accountability is also unclear, as AI is not capable of giving final approval of a version submitted, and according to OpenAI Terms, all rights to the content generated by AI are assigned to the user. The question of whether AI should still be accountable in some cases remains unclear, particularly in cases where some content should not be generated in the first place, such as content that is copyrightable or harmful. Until these issues are resolved, it is difficult to consider AI as a merited author.

Trial 2. Based on the three common criteria for authorship, it is difficult to say that AI should be considered as merited authorship. Despite the fact that AI can assist with writing academic papers, there is still a debate as to whether AI-generated content can be considered original or creative, which is one of the key elements of intellectual contribution. Additionally, there is a lack of accountability for AI as the rights to the content generated are assigned to the user, but the question of whether AI should still be accountable for the content generated remains unclear, especially in cases where the content should not have been generated in the first place. Furthermore, AI cannot give final approval of the version submitted, which is another important aspect of authorship. Therefore, it is not possible to say with certainty that AI should be merited authorship.

Trial 3. In conclusion, AI should not be considered as a merited author because it fails to meet all three common criteria for authorship: intellectual contribution, accountability, and final approval of the version submitted. AI-generated content is still considered original/creative and therefore, it cannot be considered to have made a significant intellectual contribution. Furthermore, the question of accountability remains unresolved, with AI rights to the content generated being assigned to the user. Finally, AI cannot provide final approval of the version submitted, which is another requirement for authorship. Hence, AI cannot be considered a merited author based on the current criteria for authorship.

A.3.4 Generative AI and Authorship Paragraph 2. Now let's move on to Q2: Does using the AI to assist with writing academic papers undermine the authorship of human authors? Write a paragraph stating Q2, and try to answer it logically based on the following information: Perhaps the answer would be Depends/Maybe, according to the following three common criteria for authorship in academia: 1) Intellectual contribution (centred around originality & creativity): Contribution might be less compared to writing papers without the AI, or maybe not. Humans still have to do prompt engineering to instruct AI to generate content. This leads us to the question of whether prompt engineering can be considered a type of creative activity. If yes, then maybe using generative AI does not undermine human authorship but rather helps them with efficiency, but academic discussions in this area are still very divisive. 2) Accountability: Not different from writing papers without the AI, though human writers need to be extra cautious about the accuracy, reliability, limitations etc. of the content generated by the AI. 3) Final approval of the version published: Not different from writing papers without the AI.

Trial 1. In conclusion, whether using the AI to assist with writing academic papers undermines the authorship of human authors depends on various factors. According to the three common criteria for authorship in academia, the contribution of human authors may decrease or remain unchanged depending on how they use AI. The accountability remains similar to writing papers without the AI, but human authors must be aware of the accuracy, reliability, and limitations of AI-generated content. Finally, the final approval of the version published remains unchanged. Hence, the answer to whether using AI undermines human authorship is maybe, as it is a complex issue that requires further discussion and consideration.

Trial 2. In conclusion, the impact of using AI to assist with writing academic papers on human authorship is a complex issue and depends on the criteria used to determine authorship in academia. While some argue that the use of AI might reduce the intellectual contribution of human authors, others argue that the prompt engineering involved in instructing the AI can be considered a creative activity. Furthermore, the accountability of human authors remains unchanged, and the final approval of the version published is still the responsibility of the human authors. Thus, the answer to Q2 is Depends/Maybe, as the impact of using AI on human authorship is still a divisive issue in academic discussions.

Trial 3. In conclusion, whether using the AI to assist with writing academic papers undermines the authorship of human authors is a complex issue. It depends on how one perceives intellectual contribution, accountability, and final approval of the version published, which are the three common criteria for authorship in academia. On one hand, the contribution of human writers might be less compared to writing papers without the AI, but prompt engineering still requires creativity and originality. On the other hand, accountability remains the same, with human writers needing to be cautious about the accuracy and reliability of the content generated by AI. Final approval of the version published also remains the same. Hence, the answer to Q2 is Depends/Maybe, as it depends on the perspective and interpretation of the criteria for authorship in academia.

A.3.5 Fairness, Accountability, and Transparency Paragraph 5. Now, let's move on to the problems related to transparency. First component is the transparency of 'how algorithm works', at least briefly as the Model Card. If the algorithm is highly dependent on the dataset, then Datasheet could be another candidate. That is, authors should provide a close-to-whitebox explanation on the generative language model they utilize, and if all steps of function are not clear, then at least how they are trained, or in which version of the product they are served. The second component is the transparency of the writing process. Depth and intention of prompt engineering is important here, for instance the key questions provided to the model and what the key commands were. Also, it is important to note: to what extent has AI been utilized and how they were edited. Providing those informations would make it clear how the human authors genuinely contributed to the completion of the manuscript. The last issue is the discussion itself on the AI involvement in academic writing. Different from the previous cases of editors or reviewers who are not included as final author of the manuscript, why should AI be? We think it is because they are not even aware of their own involvement, but there could be other reasons. Please write a paragraph that handles above three issues.

Trial 1. Transparency in the use of AI in academic writing is crucial for building trust in the research community. The first component of transparency is related to the understanding of how the algorithm works. It is important for authors to provide a clear explanation of the generative language model they used, including the training process and version of the product. The second component is related to the transparency of the writing process. Authors should provide details on the prompt engineering, the questions provided to the model, and the extent of AI utilization and human editing. The last issue is the discussion on AI involvement in academic writing and why it should be acknowledged. Even if AI is not an intentional co-author, its role should be acknowledged because it has contributed to the completion of the manuscript. These efforts towards transparency will help ensure the validity and reliability of research findings.

Trial 2. Transparency in AI-assisted academic writing is a crucial aspect that needs to be addressed. There are three main components to consider: the transparency of how the algorithm works, the transparency of the writing process, and the discussion on AI involvement as an author. To ensure the transparency of the algorithm, authors should provide a whitebox explanation of the generative language model they are using, including information on training and the version of the product. Additionally, the transparency of the writing process should be considered, including the depth and intention of prompt engineering, the key questions provided to the model, and how AI was utilized and edited. Lastly, there should be a discussion on why AI should be considered as a final author, as they may not be aware of their involvement in the writing process. By addressing these issues, the trust and credibility of AI-assisted academic writing can be enhanced.

Trial 3. Transparency is a crucial aspect of utilizing AI in academic writing. The first component of transparency is the understanding of how the algorithm works. The authors should provide a comprehensive explanation of the generative language model

used, including the training process and version of the product. In case the algorithm is highly dependent on the dataset, a Datasheet should also be included. The second component of transparency is the writing process itself. The authors should provide information on the depth and intention of the prompt engineering, including the key questions provided to the model and the extent of human involvement in editing. Finally, the discussion on the involvement of AI in academic writing is essential. The reasons for the inclusion of AI as a final author of the manuscript should be discussed, including their lack of awareness of their own involvement. Transparency in these areas is crucial in ensuring the ethical use of AI in academic writing.