Guideline on the Boundaries of AIGC Usage in Academic Publishing

Institute of Scientific and Technical Information of China

ELSEVIER | SPRINGER NATURE | Wiley

I am happy to see that ISTIC is working with Elsevier, Wiley and Springer Nature to develop the Guidelines on the Boundaries of AIGC Use in Academic Publishing. These Guidelines provide suggestions for authors, research institutions and academic journal publishers on how to use AIGC in a responsible manner, including how to fully and accurately disclose and state the use of AIGC. As this is a very dynamic field, STM expects recommendations will need to be adapted in future. These suggestions are practical and a useful guide at this phase.

——Shuai Yan, STM China Consultant

Springer Nature is pleased to be involved in the production of guidance on this important topic as interest in and use of generative AI gathers pace. Because of its power we need to understand both intended and unintended outcomes and so proceed with caution and care. It's also important to note that, reflecting the iterative nature of technology, this is an iterative document that will evolve as the technology – and the challenges and benefits associated with it - itself evolves.

——Steven Inchcoombe, President, Research, Springer Nature

Generative AI represents a great deal of opportunity for the future of research and scholarly publishing. Given the pace of development, we need discussion and dialog addressing how all stakeholders in the research process implement this technology ethically and transparently. These guidelines are an important contribution toward raising awareness of issues surrounding the use of this new technology and emphasize the importance of accountability in its use.

——Michael Streeter, Director, Research Integrity & Publishing Ethics, WILEY

Elsevier welcomes this important initiative led by ISTIC to develop guidelines for the use of artificial intelligence generated content in scholarly and academic publishing. We appreciated the opportunity to contribute to the guidelines and to share our own experience.

As noted in the guidelines, the impact of AI within scholarly and academic publishing is growing, and it brings both opportunities and challenges. Our own policies focus on responsible use and the need for disclosure when authors use artificial intelligence to assist with the writing process of their articles, whereas our policies emphasize the protection of authors' confidentiality and data privacy rights when reviewers and editors use artificial intelligence. In all instances, we need to provide transparency and guidance to authors, readers, reviewers, editors and contributors. This guidance from ISTIC is a key step that will assist the community to use AI in a responsible, transparent and accountable way.

——Sarah Jenkins, Director, Research Integrity & Publishing Ethics

CONTENTS

1 Background

In recent years, artificial intelligence (AI) technology has been developing rapidly, especially with the release of ChatGPT, the AI chatbot in November 2022. Artificial Intelligence Generated Content (AIGC) has entered the public eye and is widely used. It is clear that AI is gaining the ability to generate fluent language, making it increasingly difficult to distinguish the mass of generated sentences from human-written text. Some scientists are already using chatbots as research assistants to sort through ideas, provide feedback on their work, help write code, and review the literature. The impact of AI on research paper writing, producing, and other aspects of research is growing, but it is also having a major impact on the transparency and integrity of scientific research, which has attracted enormous attention from the academic community.

The main concern of the research community is that scientists, researchers, and students may fraudulently present AI-generated text as their own or simply use AIGC to produce unreliable research results. Large Language Models (LLMs) work by learning statistical language patterns from large online text databases. However, it should be noticed that LLMs may generate false and misleading information, and they also fail to indicate the source of the information. Without output control, the utilization of AIGC may lead to the generation of inaccurate viewpoints or unreliable research results, damaging the integrity of the academic ecosystem. At the same time, AIGC can make it even more difficult to detect academic misconduct (such as plagiarism and image forgery).

Therefore, it is crucial to develop a guideline that clearly defines the boundaries of the AIGC usage in the academic community.

Currently, various national science and technology regulators, publishers, and other relevant organizations (the Committee on Publication Ethics (COPE), the International Committee of Medical Journal Editors (ICMJE), the International Association of Scientific, Technical and Medical Publishers (STM), Taylor & Francis, WILEY, Springer Nature, Elsevier, etc.) have engaged in the discussions of the AIGC usage in academic papers and have provided relevant rules and guidelines. Notably, the descriptions and requirements in the AI-related guidelines and normative documents issued by different organizations at different times often exhibit inconsistencies. Therefore, based on an extensive review and study of existing research and exploration in the industry, we are committed to establishing a framework and guideline that outlines the fundamental principles of best practices for AI technology in academic publishing. We aim to provide a comprehensive framework for the regulation of AIGC usage within the publishing industry, the scientific community, and science and technology regulators, with a further consensus on the appropriate application of AI technology. In the meantime, as AI is an emerging technology, its nature and usage will inevitably continue to evolve. Therefore, this guideline will be updated on an ongoing and timely basis as necessary.

2 Objectives

2.1 Prevent academic misconduct and enhance integrity governance

Taking the prevention of misuse of AIGC as the goal, strengthen education to raise the awareness of scientific integrity, and promote academic integrity governance, to make sure that research activities can be conducted in an orderly manner.

2.2 Guide relevant stakeholders to reach consensus on the use of AIGC

Specify the best practices that relevant parties should follow in the preparation, writing, submission, peer review, publication, and dissemination of academic journal articles, and provide detailed and standardized guidance on the appropriate use of AIGC.

3 Principles

3.1 Transparency and accountability

Transparency and accountability are the fundamental principles underlying the use of AIGC in academic publishing. In the process from academic research to publication and dissemination, all the participants (including researchers, authors, peer reviewers, and readers, etc.) should be aware of and explicitly disclose the use of AIGC. Transparency should include data transparency, which involve datasets, data sources, and data processing methods; in addition, the use of intellectual property and copyright information should be disclosed as well. Accountability is a shared responsibility of all the key stakeholders, including researchers, funders, policymakers, and publishers; the establishment and clarification of the accountability standards and related information is significant.

3.2 Quality and integrity

Quality and integrity assurance is fundamental to building trust in the application of AIGC in academic research. From the design and construction of algorithms, to the inputs for training AIGC, to the inputs used in practical applications, the principles of accountability and transparency should be followed, and the use of AIGC should be indicated through identifiers or feedback mechanisms to ensure that the quality and integrity of academic research is not compromised by the utilization of AIGC technology.

3.3 Privacy and security

Privacy and security are the fundamental legal principles for the use of AIGC. AIGC should be used with respect of privacy and data protection, including the assessment of data, privacy, and security impact. Data should be appropriately anonymized to protect privacy, and measures should be taken to protect data security.

3.4 Fairness

The utilization of AIGC should be under the principle of fairness to avoid bias. As AI has the risk of replicating

and amplifying bias, potential sources of bias should be carefully assessed and reviewed in the process of training data selection, algorithm design, model generation, optimization, and application. A feedback mechanism should be established in place to monitor, review, and correct potential biases in a timely manner. Meanwhile, AIGC can help provide services such as copyediting and language polishing to reduce such cultural or linguistic unfairness.

3.5 Sustainable development

The multidisciplinary nature of AI system makes it highly suitable for addressing global concerns, such as the United Nations Sustainable Development Goals, carbon neutrality, and so forth. It also offers opportunities for public and private organizations to improve efficiency to achieve greater environmental sustainability and responsibility. AI systems carry a promise to benefit humanity, including future generations. Funding and other incentives providers of high-quality data, such as publications and databases created by publishers, help to extract actionable knowledge.

Sustainable development should be a core principle of AIGC itself. To minimize duplication and waste, the utilization of AIGC should avoid over-reliance on data that may be temporarily or permanently unavailable, while the functional modules of the tool should be based on recognized standards and guidelines to ensure that data are searchable, accessible, interoperable, and reusable.

4 Behavioral framework/ practice guideline

AIGC can provide assistance (services) at various stages of research and academic publishing. In order to foster a conducive research environment, to address potential issues, and to prevent/reduce misuse of AIGC, this section provides a framework for code of conduct to guide authors, research institutions, academic periodical publishers, and so forth, on a compliant and responsible use of AIGC.

4.1 Research and writing

This section mainly provides guidance to researchers on the use of AIGC during the research and writing prior to submission.

4.1.1 Information collection

The data provided by AIGC are generated and extracted based on big data and language models. However, the accuracy and authenticity of which are not assessed or verified, and researchers need to confirm the reliability of the content.

Literature research: AIGC can be used to collect reference literature based on keywords or topics, classify and review the literature, summarize the conclusions, and provide references for researchers; moreover, help researchers identify new sources of information and keep track of the latest developments in the research field. It should be noted that the references provided by AIGC may be fictitious or outdated; researchers using AIGC to support their literature review must carefully reviewed and verify the authenticity of each suggestion and reference provided.

Concept clarification: AIGC can answer some basic conceptual questions to assist researchers in structuring their chapters. However, it should be noted that AIGC provides concept clarification based on existing research, researchers need to check the applicability of the concepts.

Research on viewpoint information: AIGC can collect information from the text on the viewpoints, emotions, and sentiment tendencies of the public or experts on

certain issues. Researchers need to monitor and control the viewpoint information provided by AIGC, and clean up the information provided by AIGC if necessary, to ensure that researchers use only valid, unbiased material and prevent the dissemination of false, biased, or discriminatory information.

4.1.2 Statistical analysis

In some cases, researchers have collected data but are uncertain about the best statistical analysis to test their hypotheses. Researchers can use AIGC to select the most appropriate method of analysis or statistical analysis; however, the data used should be collected from their own experiments, and the results of statistical analysis should be verified by the researchers to ensure the reliability.

Data analysis and interpretation: Researchers may use AIGC to interpret data, calculate statistical indicators, perform simple data analysis, and describe statistical results. However, AIGC cannot replace the researcher's own interpretation of the data.

Suggestions and guidance on statistical methods: AIGC can provide researchers with suggestions and guidance on statistical analysis based on the question and relevant knowledge. However, these suggestions and guidance are solely based on the language model and knowledge base it has learned, which may lead to omissions and inaccuracies. Therefore, researchers need to assess the feasibility of the statistical analysis suggestions provided by AIGC, evaluate them with other reliable statistical analysis and data mining tools, or seek guidance and assistance from the subject experts to finally determine whether to accept the suggestions provided by AIGC.

4.1.3 Charting

Assisted charting: Based on the characteristics of the data and the purpose of the graph, AIGC can recommend an appropriate type of statistical graph according to the application scenario, which helps present statistical results in a straightforward manner, effectively convey the message, thus saving researchers the effort in making graphs and thereby improving writing efficiency. However, images generated from experiments such as images of Western blot, cell technology analysis, tissue cell staining, etc., must be obtained through authentic experimental research and cannot be directly generated by AIGC.

Diagram format processing and optimization: Based on the data volume and graph requirements, researchers may use AIGC to assist in adjusting the style and format of the graph, such as font size, data labeling, adding legends, changing colors, and so forth, to make the graph clearer, more appealing, and easier to understand.

4.1.4 Text writing

In the writing process, AIGC can be used as a reference for researchers to improve the readability of the text, clarify the logic of the content, and recommend sentence patterns, and so forth; however, it should not be used to generate research hypotheses, write the entire text, interpret data, or draw scientific conclusions. All tasks related to scientific or intellectual contributions should be carried out by the researchers themselves, especially the writing of the critical parts of the paper. The purpose of using AIGC should be to focus on how to convey the scientific knowledge generated by the authors in the most readable way.

4.1.5 Language and copyediting

Academic language services: Language should not be a barrier to academic communication and scientific dissemination. AIGC can serve as a high-standard language reviewer, improving the readability and writing quality of manuscripts, and thus removing language barriers in the dissemination of research. At present, AIGC-supported academic language services can assist non-native English-speakers in copyediting their manuscripts to meet the submission requirements of international journals. However, researchers also need to be aware that when a manuscript is submitted (in whole or in part) to the public AIGC, it may become part of a large language model training corpus.

4.1.6 Citation organization

When dealing with content recommended by AIGC, "citation relevance" is crucial. Researchers must ensure that the cited content is relevant to the paper, including the authenticity of the citation and the cited content.

Citation format check: AIGC can verify whether the cited literature conforms to the citation format of academic papers and identify possible errors or deficiencies.

Automatic citation generation: AIGC can assist researchers in identifying sources of citations, and automatically generate citations that conform to the citation format of academic papers based on the literature information provided by the authors. However, AIGC should only be used as an auxiliary tool, authors still need to carefully check the format and content of the cited literature to ensure that the citation conforms to the requirements of academic papers.

Automatic sorting of references: AIGC can automatically sort and check the reference list according to the specified citation format, thus helping authors to perform the relevant tasks of standardized citation in academic writing, saving the effort of manual work, improving the quality of the paper and the efficiency of researchers.

4.2 Submission

AIGC can assist with the submission process, but this requires the professional judgement of researchers who should bear the ultimate responsibility.

4.2.1 Authorship

Without the guidance of human researchers, AIGC cannot independently initiate an original research, nor can it take the responsibility of a published work or research design. Similarly, in most countries, AIGC does not have legal status or the ability to hold or transfer copyrights, which are the basic requirements for authorship. Therefore, according to the COPE position statement on AI tools, AIGC cannot perform the role of authors and cannot be listed as authors.

4.2.2 Standardized citation

All content originating from other sources must be carefully reviewed and properly cited. Authors must verify the authenticity and accuracy of the information provided by AIGC, and make reference notes to the underlying data sources, tools, collection, processing, etc.

4.2.3 Disclosure and statement

The use of AIGC should be fully and accurately disclosed and stated. The following points should be clearly specified: the user; the AI technology or system (with version number stated); the time and date of use; the prompts and questions used to generate the text; the parts of the text written or co-written by AIGC; the ideas in the paper generated by AIGC. If any part of the manuscript was written using such tools, this must be described in the Methods or Acknowledgments section in an open, transparent, and detailed manner.

[Template]

Statement: In preparing this paper, the authors used [name of specific AIGC tool/service] for [purpose of use: such as literature review/data analysis/charting, etc.]. After using this tool/service, the authors have reviewed and edited the content as necessary and take full responsibility for the content of the publication.

4.2.4 Peer review

In the peer review process, authors may use AIGC to assist in responding to review comments; however, authors are responsible for addressing peer review comments, and should be prepared to demonstrate that their response is appropriate if questioned. In the peer review process, AIGC can categorize and label review comments, helping authors to quickly understand the comments, and recommending appropriate responses for authors efficiently address the review comments.

4.3 Post-publication/publishing

4.3.1 Data storage and sharing

On the premise of meeting relevant requirements, authors are encouraged to make their original data publicly available to control and prevent problems such as data contamination and falsification.

Authors should conduct a rigorous review of the data in their papers to ensure the accuracy, completeness, and reliability of the source data. The focus of the data review should include the methods of data collection and processing, the experiments, the accuracy and precision of measurements, and the methods of data storage, and so forth.

For research papers involving experimental process, researchers should record the experiments and the data collection process in a timely, accurate, and detailed manner to avoid errors or omissions, and submit the experimental data and process records together.

4.3.2 Submission and archiving of AIGC-related materials

Researchers are encouraged to share their research data (in certain situations), including but not limited to: original data, processed data, software, algorithms, protocols, methods, materials, etc. In particular, AIGC-generated content such as text, images, programs, and so forth, should be submitted and archived as supplementary material.

4.3.3 Detection and identification of AIGC-generated content

Journals and editors are encouraged to adopt new tools to detect and identify AIGC-generated content. It is recommended to establish a review process for AIGC tools, and to develop appropriate technologies for manual review and automated detection. The detection results should be used as an auxiliary supporting basis in a comprehensive evaluation together with the scope of the paper, the requirements of the journal, and the overall quality of the paper, and so forth.

Researchers must provide clear disclosure and statement when using AIGC to generate manuscript text and other materials, otherwise it constitutes academic misconduct. For example, newly generated text extracted from AIGC as part of the manuscript without attribution will be considered plagiarism.

5 Conclusion

AIGC is an emerging concept involving multiple actors with diverse application scenarios in academic publishing, and there are many gray areas regarding the boundaries of AIGC application. Institute of Scientific and Technical Information of China (ISTIC), in collaboration with international publishing groups such as Elsevier, Springer Nature, WILEY, and after seeking advice from relevant parties, proposes suggestions for the use of AIGC technology in the form of principles and behavioral framework/practice guideline to prevent academic misconduct, strengthen integrity governance, and guide stakeholders to reach a consensus on the use of AIGC.

It is worth noting that the purpose of this guideline is to provide an exploratory framework for the application of AIGC technology, and there are still some more specific and practical issues that require further research. For example, the application of AIGC technology in academic publishing involves many stakeholders, including not only the owners and users of AIGC technology development and application, but also various responsible parties in the academic publishing process such as authors, journals, editors, reviewers, disseminators, audiences, and research regulators. It is of great importance to clarify the relationships between these parties, to identify key responsibilities and constraints, and to clearly define their responsibilities. On the other hand, there is no consensus on how to deal with the misuse of AIGC technology, and it is crucial to propose suggestions for the proper handling of various misuses for the benefit of the implementation of responsibilities.

AIGC technology and tools are still under continuous innovation and development, and the scope and behavioral framework of this guideline will also need to be regularly adapted to meet new challenges and address emerging issues. We actively invite all parties to provide suggestions and feedback to update this guideline on the use of AIGC in order to provide stakeholders with a more specific and detailed practical framework for preventing academic misconduct.

6 Acknowledgements

With the support of Department of Supervision and Scientific Integrity of Ministry of Science and Technology of the People's Republic of China, this guideline is accomplished by the team jointly set up by Institute of Scientific and Technical Information of China (ISTIC), Elsevier, Springer Nature, and John Wiley & Sons, Inc. (Wiley). The key contributors to this project are Dr. Wenwen ZHENG, Professor Daiqing YANG, Professor Jiping GAO and Dr. Meiling WANG from ISTIC; Mr. Peter LEE, Ms. Xiaoling KANG, Ms. Jimeng LI and the Research Integrity and Publishing Ethics Centre of Expertise at Elsevier; Dr. Nick CAMPBELL, Mr. Chris GRAF, Mr. Arjan GROOTENBOER, Ms. Alice HENCHLEY, Ms. Zeying HU (Maggie), Mr. Steven INCHCOOMBE, Dr. Imogen ROSE, Mr. Henning SCHOENENBERGER, Ms. Katrin STIENEMEIER, Dr. Niels Peter THOMAS and Mr. Li ZHANG from Springer Nature; and Mr. Michael Streeter, Ms. Joyce GRIFFIN and Ms. Yu Wang (Shirley) from Wiley. We would also like to express our gratitude to the experts, scholars and experienced professionals from STM publishing and research management fields in China and abroad who have provided us with valuable suggestions and guidance.

中国科学技术信息研究所
INSTITUTE OF SCIENTIFIC AND TECHNICAL INFORMATION OF CHINA

Institute of Scientific and Technical Information of China (ISTIC)

Scientometrics and S&T Evaluation Center

Fuxing Road 15, Beijing, China, 100038

Contact: Dr. Wenwen Zheng

Tel: +86 10-58882604

Email: zhengww@istic.ac.cn