Reporting guideline for the use of Generative Artificial intelligence tools in MEdical Research: the GAMER Statement

Xufei Luo,[1] Yih Chung Tham,[2,3,4,5] Mauro Giuffrè [ID] ,[6,7] Robert Ranisch,[8] Mohammad Daher,[9] Kyle Lam [ID] ,[10] Alexander Viktor Eriksen,[11] Che-Wei Hsu [ID] ,[12,13] Akihiko Ozaki [ID] ,[14] Fabio Ynoe de Moraes,[15] Sahil Khanna,[16] Kuan-Pin Su,[17,18] Emir Begagić,[19] Zhaoxiang Bian,[20,21] Yaolong Chen [ID] ,[1,22,23] Janne Estill,[22,24] The GAMER Working Group

Abstract

Objectives Generative artificial intelligence (GAI) tools can enhance the quality and efficiency of medical research, but their improper use may result in plagiarism, academic fraud and unreliable findings. Transparent reporting of GAI use is essential, yet existing guidelines from journals and institutions are inconsistent, with no standardised principles.

Design and setting International online Delphi study.

Participants International experts in medicine and artificial intelligence.

Main outcome measures The primary outcome measure is the consensus level of the Delphi expert panel on the items of inclusion criteria for GAMER (Rreporting guideline for the use of Generative Artificial intelligence tools in MEdical Research).

Results The development process included a scoping review, two Delphi rounds and virtual meetings. 51 experts from 26 countries participated in the process (44 in the Delphi survey). The final checklist comprises nine reporting items: general declaration, GAI tool specifications, prompting techniques, tool's role in the study, declaration of new GAI model(s) developed, artificial intelligence-assisted sections in the manuscript, content verification, data privacy and impact on conclusions.

Conclusion GAMER provides universal and standardised guideline for GAI use in medical research, ensuring transparency, integrity and quality.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Generative artificial intelligence (GAI) tools, such as chatbots and large language models, are increasingly used in medical research to enhance efficiency and quality. However, their application lacks standardised reporting, leading to concerns about transparency, academic integrity and data reliability. Existing guidelines, like CONSORT (Consolidated Standards of Reporting Trials)-AI and STARD (Standards for Reporting of Diagnostic Accuracy Study)-AI, address GAI use in specific cases but do not tackle GAI's unique challenges in general, such as content verification and ethical implications.

WHAT THIS STUDY ADDS

⇒ This study introduces the **G**enerative **A**rtificial intelligence tools in **ME**dical **R**esearch (**GAMER**) checklist, a comprehensive reporting guideline developed through a rigorous international consensus process involving 51 experts from 26 countries. Comprising nine items, GAMER ensures transparent disclosure of GAI use in medical research, covering tool specifications, roles and impacts on findings. Unlike prior frameworks, it focuses exclusively on GAI and covers all steps of the research project and all study types, offering a standardised approach to improve the reproducibility and trustworthiness of medical research.

Introduction

An increasing number of generative artificial intelligence (GAI)-based tools have been developed in recent years. With the release of Chat Generative Pre-trained Transformers (ChatGPT) 3.5 at the end of 2022 GAI tools have become popular also with the general public, with promising

---

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ GAMER will provide the first universal guidance on how to report the use of GAI tools. We hope GAMER will be adopted as the minimum reporting standard by journals, which will promote the appropriate and transparent use of GAI in medical research.

---

applications in medicine. GAI represents a form of artificial intelligence (AI) that is trained on extensive multimodal data sets to generate new content and ideas, including articles, conversations, images, videos and music.[1] GAI tools can assist in a range of tasks in medical practice, from consulting on medical knowledge and drug discovery, to assisting with consent-taking interviews,[2] diagnosing and treating diseases using medical records and pathology images, as well as supporting medical research.[3] In medical research, GAI shows substantial potential for numerous applications such as structuring narrative sections of a manuscript article (eg, the introduction and discussion),[4] research code generation,[5] language editing assistance,[5] data extraction (eg, text from images), data structuring and transforming (eg, formatting text into a table) and data analysis.[6]

However, as an emerging technology, the application of GAI in medical research is accompanied by several issues and challenges, which are partly related to the lack of regulation and relevant guidelines.[7 8] First, the authenticity and reliability of content generated by GAI tools require thorough verification. For instance, while GAI tools can assist in writing scientific papers, the veracity of the generated content is not guaranteed, raising concerns about potential academic fraud.[9–11] Second, the use of GAI tools may pose risks related to data privacy breaches, as well as ethical issues.[12 13] Third, the quality of the data used to train the GAI tools may be unsatisfactory, which will in turn impact the tool's ability to produce appropriate outputs.[14]

The development of dedicated reporting guidelines for the use of GAI tools in medical research could help to address these issues. Reporting guidelines are simple, structured tools that assist health researchers to include necessary information in writing manuscripts.[15] In the field of AI, several reporting guidelines have already been established, such as Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD)-AI for prediction model evaluation,[16] Standards for Reporting of Diagnostic Accuracy Study (STARD)-AI for diagnostic accuracy studies,[17] Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT)-AI for randomised controlled trial protocols,[18] Consolidated Standards of Reporting Trials (CONSORT)-AI for randomised controlled trials,[19] Developmental and Exploratory Clinical Investigations of DEcision support systems driven by Artificial Intelligence (DECIDE-AI) for various study types,[20] ChatGPT and Artificial Intelligence Natural Large Language Models for Accountable Reporting and Use (CANGARU) for ChatGPT[21] and Checklist for Artificial Intelligence in Medical Imaging (CLAIM) for medical imaging.[22] However, these guidelines refer either to the use of AI in general or to specific tools. Guidelines that specifically focus on GAI or GAI-based tools are still lacking. Recent guidelines like TRIPOD-LLM[23] focus on specific large language model (LLM) applications, whereas instruments that provide broader guidance for all GAI tools across medical research are still lacking. GAI encompasses tools that autonomously generate new content, such as text, images or data, based on learnt patterns from vast data sets. Due to these unique

capacities, existing guidelines on the use of AI in general may not be sufficient to cover all essential aspects related to GAI tools. Although multiple journals and institutions have issued regulations or guidelines on the use of GAI tools,[24–32] the lack of a rigorous and transparent development process and the substantial inconsistencies between the different tools limit their international recognition and applicability.[8]

To address this gap, we convened an international, multidisciplinary group of experts to develop a comprehensive reporting guideline for the use of Generative Artificial intelligence tools in MEdical Research (GAMER). Drawing from a rigorous literature review and extensive consultations with key opinion leaders, the guideline was crafted through numerous iterations and discussions. GAMER complements study-specific guidelines like CONSORT-AI by stating reporting requirements for GAI use across all research phases, not limited to any specific study type. This guideline applies to all types of medical research (such as literature reviews, clinical trials or observational studies) and covers all forms of utilisation of GAI tools (eg, LLMs or image generators) in any phase of the study (such as study design or manuscript writing). It does not cover non-GAI tools or, for example, general web search engines, distinguishing its scope from broader AI applications.

Methods

We assembled an international multidisciplinary expert panel and followed the methodology recommended by the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) Network to develop the GAMER reporting guideline.[33] We analysed relevant reporting guidelines and literature,[24–31] conducted a Delphi survey with up to two rounds and held subsequent online meetings with the panellists to formulate the final version of the checklist.

Patient and public involvement

None.

Sponsor and supporting organisations

The GAMER reporting guideline was initiated by the Evidence-Based Medicine Center of Lanzhou University, Chinese EQUATOR Centre, WHO Collaborating Centre for Guideline Implementation and Knowledge Translation and Health Data and Digital Medicine Branch of the China International Exchange and Promotive Association for Medical and Health Care. This work was supported by the Research Unit of Evidence-Based Evaluation and Guidelines (2021RU017), Chinese Academy of Medical Sciences, School of Basic Medical Sciences, Lanzhou University, Lanzhou, Gansu, China. Informed consent was obtained from all participants in the Delphi process and online consensus meeting.

Protocols and registrations

The protocol for the GAMER reporting guideline has been published elsewhere.[34] We registered the project on the EQUATOR website on 3 November 2023 (https://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-other-study-designs/#CHEER).

Expert recruitment

We formed four expert groups: the Advisory Committee, the Core Team, the Delphi Expert Group and the Coordination Team. Their roles and responsibilities are detailed in the protocol.[34] For the Delphi Expert Group, we recruited members through two channels: first, by searching the

Table 1 GAMER checklist

No.	Item	Reported	Page
1	Did you use any GAI tools (such as large language models or large visual models) in any section or step of this manuscript or study?	□Yes □No □N/A	
2	Specify the GAI tool(s) used, their versions and/or release dates and the date(s)/period the tools were used.	□Yes □No □ N/A	
3	Describe whether a specific prompting technique was used to generate any content of the manuscript or to perform analyses during the study. Please also provide the unedited responses to the prompts.	□Yes □No □ N/A	
4	If a new GAI tool was developed or fine-tuned based on an existing AI model, report the name and version of the original model.	□Yes □No □ N/A	
5	Describe the role of GAI tools in all phases of this study where they were used (including manuscript writing).	□Yes □No □ N/A	
6	Report the specific section or paragraphs of the manuscript that GAI tools contributed to.	□Yes □No □ N/A	
7	Describe how the content generated by GAI tools was verified and (when necessary) modified.	□Yes □No □ N/A	
8	Describe how data privacy and confidentiality were ensured during the use of GAI tools.	□Yes □No □ N/A	
9	Describe whether and how the use of GAI tools may have influenced the interpretation of results, the study's overall accuracy, or conclusions.	□Yes □No □ N/A	

AI, artificial intelligence; GAI, generative artificial intelligence; GAMER, Generative Artificial intelligence tools in MEdical Research; N/A, not applicable.

Table 2 Glossary of terms relevant for the GAMER checklist

Terms	Explanation
Artificial intelligence (AI)	The science of developing computer systems that can perform tasks normally requiring human intelligence.[19]
Application programming interface	A set of rules or protocols that allows different software applications or components to communicate and interact with each other.
Chatbot	A computer programme designed to simulate human conversation, enabling users to interact with digital devices via text or voice.
ChatGPT	An AI chatbot programme developed by OpenAI, launched in November 2022.
Delphi survey	A research method that gathers the collective opinions of a group through staged consultations of surveys, questionnaires or interviews, aiming to reach consensus.[19] In this article, we mean by the Delphi method a modified version of the original approach, adapted for the needs of reporting guideline development.
Fine-tuning	Modifications or additional training of the AI model, aimed at enhancing its performance.[19]
Generative artificial intelligence (GAI)	A type of AI capable of generating new content and ideas, including conversations, stories, images, videos and music.
Generative pre-trained Transformer (GPT)	An AI model that uses a large-scale transformer neural network to understand and generate human-like text based on its training on vast amounts of language data.
Large language model	Deep learning model with a vast number of parameters, trained unsupervised on large volumes of text.[75]
Large visual models	A cutting-edge AI model designed for visual tasks, including image recognition, object detection, segmentation and image creation.
Prompt	Text input by a user into the chatbot for the purpose of communicating with the GAI tool.[76]
Prompt engineering	The process of refining a GAI prompt to enhance its accuracy and effectiveness.
Query	The process of asking questions by entering prompts into a chatbot.
Reporting guidelines	Recommendations for information authors should include in their manuscripts when writing about their research.
Temperature	A parameter that influences the randomness and creativity of the model's output.

GAMER, Generative Artificial intelligence tools in MEdical Research .

PubMed database on 3 October 2023, we identified 200 experts who had published research related to GAI tools in the medical field; and second, we invited experts through our previous collaborators and using a snowball sampling method. We paid attention to the diversity of the panellists to ensure GAMER's inclusivity and applicability across different medical research contexts and geographical regions, incorporating the various cultural and discipline-related perspectives. We informed the panellists about the GAMER project to obtain their consent and invited those who agreed to join.

Generation of the initial pool of items

We formed a pool of potential items for the GAMER checklist through the following methods: (1) retrieving published AI-related reporting guidelines such as DECIDE-AI[20] and CONSORT-AI[19] and collecting relevant items from them; (2) retrieving guidelines on the use of GAI tools from instructions for authors on websites of journals and publishers[8 35]; (3) conducting a scoping review of how GAI tools are reported in published literature; and (4) reviewing relevant literature recommended by the Advisory Committee. Using these methods, candidate items were gathered and discussed by the Core Team members to form the initial pool of items.

Delphi survey

We planned to conduct one to two rounds of the Delphi survey to gather the experts' opinions and suggestions and reach consensus. The decision to carry out a second round was based on the results of the first round of voting. The

participants were requested to rate each item on a scale ranging from 1 to 7 points, with 1 indicating strong disagreement and 7 indicating strong agreement for inclusion. The consensus to include or exclude the item was based on the following criteria:

1. If the median score was between 1 and 3, the item would be excluded.
2. If the median score was 4 or 5 (or higher with substantial comments on the content), the item would be discussed and entered into the next round of the Delphi survey or the consensus meeting.
3. If the median score was 6 or 7 without any substantial comments, the item would be included in the final checklist.

The Delphi survey was conducted using SurveyMonkey (http://surveymonkey.com). Detailed questionnaires for the

first and second rounds are shown in online supplemental appendices 1,2.

Online meeting

Online meetings were held after the Delphi survey to collect the experts' opinions on any remaining questions and optimise and formulate the final checklist. Considering that the experts are based in different time zones, the scheduling of online meetings was done via Doodle (https://doodle.com/), with the option to arrange more than one meeting if necessary. The online discussions were conducted through Zoom (V.6.0.11 (35001)) and recorded on video. The topics of the online meetings were to discuss items that had not achieved consensus and those with disputes, as well as to modify and refine the wording and order of the items. For experts who could not attend any of the online meetings, we sent the video recordings and summaries of the meetings, along with an online feedback form, to collect their views and suggestions. All opinions and suggestions were documented, and responses to each expert's suggestions were provided through email.

Approval of the final checklist

After the Delphi survey and online meetings, the core group discussed, modified and finalised the items of the GAMER checklist based on the experts' comments and suggestions. They also prepared a glossary and an explanations and elaborations (E&E) document. All documents were sent to the expert group members via email for approval before the final submission.

Results

Characteristics of the Delphi panel

A total of 44 experts from 26 countries or regions participated in the Delphi survey: 43 experts in the first round and 33 in the second round. The Delphi expert group included professionals from various medical specialties, epidemiology, computer science, medical ethics, AI and guideline methodology, as well as medical journal editors, policymakers and medical educators. Details on the GAMER working group members are presented in online supplemental appendix 3.

Results of the Delphi survey and online consensus meetings

A total of seven items were included in the first round of the Delphi survey, and all items met the pre-set threshold for acceptability (median score ≥6). However, one item, addressing the declaration of who was responsible for GAI use, was removed after discussion, as the panel deemed it redundant given authors' collective responsibility in academic publishing. In the first round of the survey, a total of 135 comments or suggestions were received, including four new items proposed by the panellists. These four items were taken to the second round of the Delphi survey, and consensus was reached on three items. During the second round of the survey, a total of 130 comments or suggestions were given. The questionnaires and details of the scores for both rounds of the Delphi survey are shown in online supplemental appendices 1,2, and the suggestions and comments from the experts with responses from the core group are provided in online supplemental appendices 4,5. We also requested the panellists to vote on which section of the manuscript the included items should be reported in online supplemental appendix 6. Online

meetings were held on 30 May and 31 May 2024 with 14 and 12 participants, respectively. The main issues discussed are outlined in online supplemental appendix 7. During the meetings, the participants reached consensus to definitely remove the two items dropped in the Delphi survey, as well as on the exact terminology to be used, and suggested some revisions and formatting to the existing items. A summary of the discussion is provided in online supplemental appendix 8. We also collected feedback from experts who did not participate in the online meetings through an online form (online supplemental appendix 9).

Final GAMER checklist and its explanation and elaboration

Based on the two rounds of Delphi survey, two online meetings and repeated revisions and optimisations by the core team, we developed the final GAMER checklist comprising nine reporting items, namely, general declaration, GAI tool's specifications, prompting technique, GAI tool's role in the study, declaration of new GAI model(s) developed, AI-assisted sections in manuscript, content verification, data privacy and impact on conclusion (table 1).

Item 1 Did you use any GAI tools (such as LLMs or large visual models) in any section or step of this manuscript or study?

Explanation and elaboration

Inappropriate, non-transparent or unverified use of GAI tools in manuscripts or studies can lead to untrustworthy research results. In severe cases, such use may be considered academic fraud, potentially resulting in retraction and undermining the reliability and integrity of academic work.[36 37]

If GAI tools were used in a medical research paper at any stage (ie, generating content ideas, structuring the manuscript, checking grammatical errors or improving clarity), our guideline suggests reporting several details to enhance the transparency of the study. Here, we focus solely on GAI tools, such as ChatGPT, Claude, Gemini or any other similar instruments. This checklist does not cover the use of tools intended solely for language translation (eg, Google Translate).

If the authors or investigators did not use any GAI tool in their study or writing the manuscript, the remaining items of this checklist do not need to be reported.

Examples from published studies

▶ During the preparation of the manuscript, the authors used ChatGPT and PaperPal to correct typographical and grammatical errors.[38]

▶ ChatGPT 3.5 designed by OpenAI was used to help with language editing.[39]

▶ This study investigates the use of ChatGPT-4 in identifying suitable candidates for bariatric surgery and providing surgical recommendations to improve decision-making in obesity treatment amid the global obesity epidemic.[40]

▶ Neither ChatGPT nor other generative language models were used for the ideation or writing process.[41]

Item 2 Specify the GAI tool(s) used, their versions and/or release dates and the date(s)/period the tools were used.

Explanation and elaboration

Authors should disclose in the relevant sections of the paper the name (eg, ChatGPT, Claude, Gemini) and the version or release date of the GAI tool(s) that were used. Since most GAI tools are being continuously trained with new data and

fine-tuned, it is also advisable to provide the exact date(s) when the tool was applied. Also, authors should disclose whether they are using the front-end interface or application programming interface (API). The temperature, token length, language, layers or other settings should also be reported if available.

Examples from published studies
► We used GPT-4 (OpenAI), an advanced LLM that was initially introduced in 2022. A single investigator (ASH) prompted GPT-4 (version dated 12 May 2023) for all queries.[42]
► All models used GPT-4 turbo (gpt-4–1106-preview), with temperature set at 0 to generate the most deterministic (ie, least random) results and context reset prior to each vignette.[43]
► From 10 May to 13 June 2023, responses to these queries were generated by using two versions of ChatGPT (version GPT-3.5 and GPT-4.0, OpenAI, California, USA) and Google Bard (Google, Alphabet, California, USA).[44]
► The March 2023 edition of GPT-4 (maximum determinism: temp=0) was provided in each case five times to assess reproducibility across repeated runs.[45]
► We queried GPT-4 (OpenAI model='gpt-4–0314'; role='user'; temperature=0; all other settings at default values) to consider the clinical history of each pair of ED presentations and return which patient had a higher-acuity presentation.[46]
► On 22 and 23 December 2022, the original full text of the question was put into a fresh chatbot session, in which the session was free of prior questions asked that could bias the results (version GPT-3.5, OpenAI) and the chatbot response was saved.[47]

Item 3 Describe whether a specific prompting technique was used to generate any content of the manuscript or to perform analyses during the study. Please also provide the unedited responses to the prompts.
Explanation and elaboration
Prompt engineering is the process of creating clear, concise and easily understandable prompts that help the machine or the AI model to generate or predict the content to its best capacity.[48] Prompt engineering has a significant impact on the responses from GAI tools, and high-quality prompts can effectively elicit high-quality answers.[49] Therefore, when using such tools, authors are advised to retain the dialogue records to facilitate reviewers, editors and readers in replicating and understanding the process. If possible, these dialogue records should be submitted together with the manuscript as .

Examples from published studies
► Appendix A presents the final prompts for each data element.[50]
► Summaries of example questions and the corresponding physician and chatbot responses are shown in the table.[47]
► Using the training data in the CDSA data set (n=78), we experimented and improved prompts iteratively and the final prompt is presented in figure 2.[51]
► The response data produced by GPT-3.5 and other remaining experimental data generated in this study are provided in the supplementary information /source data file. Source data is provided with this paper.[52]

► We developed an LLM-based workflow, using systems engineering methodology and spiral 'prompt engineering' process, leveraging OpenAI's API for batch querying ChatGPT.[51]

Item 4 If a new GAI tool was developed or fine-tuned based on an existing AI model, report the name and version of the original model.
Explanation and elaboration
Researchers often create, train or fine-tune tailored GAI tools based on existing LLMs to meet their research objectives and standards better. In such cases, we suggest disclosing detailed information about the original LLM that the tool is based on, including its name, release date, version and any other details. The details could be included as to keep the manuscript concise. This item applies only to those who have developed, trained or fine-tuned their own GAI tools. For widely recognised tools such as ChatGPT or Claude, this item is not applicable.

Examples from published studies
► We developed our model using LLaMA-65B. Leveraging low-rank adaptation, we performed supervised fine-tuning using a data set crafted for instruction-following tasks, including data generated by GPT-4 from 52 000 prompts in alpaca.[53]
► The purpose of this study was to create a baseline model for automated target word prediction of paraphasias within spoken discourse using the surrounding language alone. We fine-tuned the LLM BigBird to predict the intended target word of paraphasias within transcripts of the Cinderella story retell task using data from controls, people with aphasia and a combination.[54]
► The YOLOv7 model is trained using a clinical data set, with data augmentation techniques employed to enhance the data set to identify six types of pressure injury images. The established system features a front-end interface that includes responsive web design and a chatbot with ChatGPT, and it is integrated with a database for personal information management.[55]

Item 5 Describe the role of GAI tools in all phases of this study where they were used (including manuscript writing).
Explanation and elaboration
GAI tools can be used for diverse tasks in writing medical papers and conducting research, such as language polishing, outlining ideas, generating software code or improving the structure of the paper. These roles should be transparently disclosed and reported in the article.

Examples from published studies
► During the preparation of this work, the authors used ChatGPT in order to correct grammatical mistakes. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.[56]
► ChatGPT advanced data analysis (previously known as 'Code Interpreter') was used for analyses and may be accessed via https://chat.openai.com for ChatGPT Plus users.[57]
► The author generated this text in part with GPT-3, OpenAI's large-scale language-generation model. On generating draft language, the author reviewed, edited and revised the language to their own liking and takes ultimate responsibility for the content of this publication.[58]

► ChatGPT 3.5 designed by OpenAI was used to help with language editing.[39]

► The incorporation of ChatGPT was envisioned to enhance student learning experiences and assist in project planning, programming code generation, examination preparation, workflow exploration and technical interview preparation, thus advancing medical informatics education.[59]

Item 6 Report the specific section or paragraphs of the manuscript that GAI tools contributed to.
Explanation and elaboration
Authors should explicitly report in the article for which paragraphs or sections GAI tools were applied, to assist readers in better understanding and assessing the content and value of the paper. Mentioning the exact sections can facilitate and accelerate the peer-review process. If the GAI tools were used not only for text editing but also for research protocol design, content creation or generating new text, it is encouraged to report the section of the manuscript, stage or specific task. If the tool is used solely for language editing, it is commonly used for the whole text and listing the specific sections is not necessary. The distinction between this item and Item 5 is that the present item refers to the concrete location, part or paragraph in the manuscript for which the tools are used, while Item 5 details the specific roles or functions of the tools.

Examples from published studies
► This study used generative AI tools to analyse data, create preliminary themes, produce draft text and revise wording throughout the production of the manuscript.[60]

► The present work examines the cutting-edge advancements in the stages of 'Literature revision and analysis' and 'Write scientific reports and publications' (highlighted in red) using ChatGPT, a chatbot based on the GPT-3.5 language model.[61]

► ChatGPT 4.0 was used for grammar correction and ChatGPT image generator was used to draw figure 1B.[62]

Item 7 Describe how the content generated by GAI tools was verified and (when necessary) modified.
Explanation and elaboration
Content directly generated by GAI tools may contain false or exaggerated information, so it is recommended to manually proofread, verify and, if necessary, revise the generated content to ensure its accuracy and reliability. For example, if the tool was used for language refinement, the output must be checked by a human author to ensure that the revised text corresponds with the original intended meaning. If no verification was performed, the reason should be clearly stated.

Examples from published studies
► After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.[63]

► Next, each citation generated by GPT was fact-checked and replaced by the authors when the citation did not exist or when it did not match the content of the sentence.[64]

► The answers provided by ChatGPT were compared with the official answer key, which had been reviewed for any changes resulting from the advancement of medical knowledge.[65]

Item 8 Describe how data privacy and confidentiality were ensured during the use of GAI tools.
Explanation and elaboration
GAI tools like ChatGPT often interact with individuals who use personal information to complete the query to be asked. Such interactions may involve the exchange of highly sensitive information. Another concern is that chatbot models might store user data or use it for training purposes, meaning that the owners of the data no longer have control over it.[66] It is critical to maintain user privacy and data security during these communications. Standards like data anonymisation, end-to-end encryption and differential privacy should be incorporated to protect personal and patient data and forbid latent data breaches or misuse. Local data protection and ethical regulations also must be followed. For authors, it is advised to pay attention to the protection of privacy when giving inputs to the GAI tools, for example, by removing any sensitive patient information before the tool can access the data.

Examples from published studies
► Prior to inputting any information into the chatbots, each patient's data were anonymised and all personally identifiable information was removed according to data privacy standards.[67]

► To further protect patient privacy, any data sent to ChatGPT was de-identified or anonymised to remove personal information.[68]

► We also presented the prototypical anonymised case vignettes to two versions of ChatGPT, based on either GPT-3.5 or GPT-4 (ChatGPT version: 24 May 2023).[69]

Item 9 Describe whether and how the use of GAI tools may have influenced the interpretation of results, the study's overall accuracy or conclusions.
Explanation and elaboration
Authors should report whether and how the content generated by GAI tools may have influenced the results. If the tool was used for language editing correctly, it generally should not affect the content, assuming that the original text/prompt was clear. However, if the tool was used to directly generate content (such as software code), the results may not necessarily match the intention of the authors. The authors should therefore ensure the accuracy and integrity of all generated content. Moreover, as for any content of a scientific article, all authors should be responsible for the consequences of using GAI tools, as required by, for example, Elsevier for its journals.[70]

Examples from published studies
► Additionally, the study did not evaluate the impact of ChatGPT's use on actual clinical outcomes, patient satisfaction or healthcare provider workload, leaving the real-world implications of using ChatGPT in clinical practice uncertain.[71]

► During the preparation of this work, the author(s) used ChatGPT in order to improve readability and language. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.[72]

Discussion
The GAMER statement represents the culmination of an international consensus process that brought together a diverse group of experts with extensive AI and medical research backgrounds.

The high response rate from the invited experts highlights the recognised need and eagerness for a structured and standardised reporting guideline for the use of GAI tools in medical research. The development of the GAMER checklist followed a rigorous methodology, with standardised reporting and transparent disclosure of the process and results in each step and section.[33] We hope that this robust checklist will be widely adopted by a broad range of stakeholders and users including authors, reviewers and journal editors alike, fostering transparent reporting of the use of GAI tools, so as to enhance integrity and quality across the field.

To support the development of the GAMER checklist, we reviewed the methodologies and contents of the CONSORT-AI,[19] SPIRIT-AI,[18] DECIDE-AI,[20] STARD-AI,[17] TRIPOD+AI[16] and BePRE-CISE (Better Precision-data Reporting of Evidence from Clinical Intervention Studies & Epidemiology) checklists,[73] as well as a recently published systematic review.[74] We also published an editorial advocating for the development of a dedicated reporting checklist for the use of GAI tools early on in the project,[27] which was very beneficial for recruiting leading experts globally. The solid groundwork laid in the early stages was crucial for developing GAMER and is one of the effective guarantees for its subsequent dissemination and adoption.

To better assist stakeholders in understanding and using the GAMER checklist, we have provided a detailed list of E&E on each item and a glossary of terms related to the GAMER checklist (table 2). Moreover, the GAMER checklist itself is very brief, with only nine reporting items. The checklist, together with the E&E section, can efficiently help authors in reporting the use of GAI tools when writing their articles. This will also swiftly assist reviewers and editors in checking whether the use of GAI tools has been disclosed in the submission.

While journal instructions often focus on the use of GAI in manuscript preparation, GAMER extends this by providing a standardised approach for reporting its use also in study design, data collection and analysis. For example, if GAI is used to develop a research protocol, GAMER requires authors to disclose the tool's role, the prompts used and how the output was verified, which helps to ensure the transparency and reproducibility of the research process. If the authors instead reported only that a GAI tool was used but not the verification process or the prompts, the reader cannot necessarily be convinced or check the appropriateness of the selected methodology, which in turn can ultimately make the results of the study lose their value.

It is noteworthy that, like other reporting guidelines, the GAMER checklist can be used not only to guide researchers on how to disclose and report the use of GAI tools when writing articles, but also to assist reviewers and readers to evaluate whether the use of GAI tools has been properly and transparently reported in manuscripts under review and published articles (to further ascertain the quality). It is also important to note that the application of the GAMER checklist is not limited to specific types of research. If an article uses GAI tools, the checklist items should be followed to disclose all relevant information; however, if no GAI-based tool was used, it is also recommended to disclose this explicitly. We also recommend authors who used GAI tools in their manuscripts to submit the filled GAMER checklist during submission.

The Delphi expert panel decided to remove two items from the GAMER checklist due to lack of consensus after two rounds of survey. The first was a request to report who is responsible for the use of GAI tools. However, given the current principles of academic publishing which indicate that all authors are collectively responsible for all content of the manuscript, the expert panel concluded that such an item would not have substantial added value if included in the GAMER checklist. The second item removed was about reporting the date of GAI tools usage; its content was eventually integrated into item two (specification of the GAI tool). Additionally, regarding the terminology used in the GAMER checklist, the expert panel decided to use the term 'GAI tools', which is a broader term than, for example, 'large language models', which was also considered as an option. Due to the rapidly evolving nature of the field and the broader scope covered by GAI tools, we expect that other types of GAI tools, for example, large visual models and multimodal GAI tools, will become increasingly used in the future. Therefore, using a broad term such as GAI tool can also ensure the long-term relevancy of GAMER into the era of the next generation of GAI. The GAMER checklist thus aims to cover the use of all GAI tools instead of only LLMs.

Additionally, we also deliberated on the ideal reporting position for each item. Most of the required content was suggested to be reported in the methodology section. However, the panel did not see the need to make this a mandatory requirement, but rather a recommendation (online supplemental appendix 6).

Despite our rigorous approach in developing the GAMER reporting guideline, several limitations need to be considered. First, despite our best efforts to have a balanced gender distribution, the proportion of women in our expert panel was only 15%. This under-representation may be an indirect result of our recruitment strategy, which focused on screening first and last authors of relevant GAI-related papers, indicating that there may still be relatively few female researchers in this field. Addressing this imbalance is a priority for the future. Second, we did not include patient representatives in our checklist-making process because we considered that their understanding and knowledge of GAI tools might be limited. However, we will consider including them in future updates to ensure broader stakeholder engagement.

After the release of the GAMER reporting guideline, promotion and dissemination efforts are also crucial. First, we will disseminate the GAMER checklist via online platforms and academic conferences, including presenting and interpreting the GAMER checklist at various conferences and using social media platforms such as X (previously Twitter), WeChat and LinkedIn to increase awareness among researchers. Second, we will invite members of the Delphi expert group to translate the GAMER checklist into their local languages for broader dissemination. Our experts represent 26 countries and regions worldwide, offering broad geographical diversity and significant potential for promoting the checklist globally. Third, we will contact the editors of major journals and recommend integrating the GAMER checklist into the authors' guidelines of journals to better assist and enhance the disclosure of GAI tools usage. Fourth, we will create a dedicated website for the GAMER statement and organise members of the GAMER working group to promote and disseminate the checklist. These steps are designed to ensure wide-reaching awareness and implementation of the GAMER reporting guideline globally. Finally, we will establish a long-standing coordination group for GAMER so as to regularly discuss and review the statement. This group will meet on a yearly basis, conduct an assessment of the usability of the checklist considering the latest developments in AI technology and collectively decide whether a revision is necessary.

Conclusions

The GAMER reporting guideline have been developed through a comprehensive and structured consensus process to enhance the transparency and rigour of medical research involving GAI tools.

With the rapid proliferation of studies using GAI tools, there is a critical need for stringent reporting of the use of such tools. The GAMER guideline represents a comprehensive, universal and standardised reporting guideline for GAI tools in medical research. This guideline fills a major gap in existing reporting practices and can substantially benefit authors and reviewers of scientific articles, as well as journal editors, by improving the transparency of medical research. We hope for the widespread adoption of this robust guideline to further enhance the integrity and quality of GAI-based research across the field of medicine.

Author affiliations

[1]Research Unit of Evidence-Based Evaluation and Guidelines, Chinese Academy of Medical Sciences (2021RU017), School of Basic Medical Sciences, Lanzhou University, Lanzhou, Gansu, China

[2]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

[3]Centre for Innovation and Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

[4]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore

[5]Ophthalmology and Visual Science Academic Clinical Program, Duke-NUS Medical School, Singapore

[6]Department of Internal Medicine (Digestive Diseases), Yale School of Medicine, Yale University, New Haven, Connecticut, USA

[7]Department of Medical, Surgical, and Health Sciences, University of Trieste, Trieste, Italy

[8]Faculty of Health Sciences Brandenburg, University of Potsdam, Potsdam, Brandenburg, Germany

[9]Orthopedic department, Hôtel Dieu de France, Beirut, Lebanon

[10]Department of Surgery and Cancer, Imperial College London, London, UK

[11]Department of Geriatric Medicine, Odense University Hospital, Odense, Denmark

[12]Department of Psychological Medicine, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand

[13]Bachelor of Social Services, College of Community Development and Personal Wellbeing, Otago Polytechnic, Dunedin, New Zealand

[14]Jyoban Hospital of Tokiwa Foundation, Iwaki, Fukushima, Japan

[15]Department of Oncology, Queen's University, Kingston, Ontario, Canada

[16]Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota, USA

[17]Mind-Body Interface Research Center (MBI-Lab), China Medical University Hospital, Taichung, Taiwan

[18]An-Nan Hospital, China Medical University, Tainan, Taiwan

[19]Department of Neurosurgery, Cantonal Hospital Zenica, Zenica, Bosnia and Herzegovina

[20]Vincent V.C. Woo Chinese Medicine Clinical Research Institute, School of Chinese Medicine, Hong Kong Baptist University, Hong Kong, China

[21]Chinese EQUATOR Centre, Hong Kong, China

[22]Evidence-based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou, China

[23]WHO Collaborating Centre for Guideline Implementation and Knowledge Translation, Lanzhou, China

[24]Institute of Global Health, University of Geneva, Geneve, Switzerland

Collaborators The GAMER Working Group: Susan L Norris (Oregon Health and Science University, Portland, Oregon, USA), Jean-Christophe Bélisle-Pipon (Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada), Qingyu Chen (Biomedical Informatics and Data Science, School of Medicine, Yale University, New Haven, Connecticut, USA), Brian D Earp (Centre for Biomedical Ethics, National University of Singapore, Singapore), Lorenzo Righetto (Nature Medicine, Springer Nature), Renne Rodrigues (Postgraduate Program in Public Health, Universidade Estadual de Londrina (UEL), Brazil, Universidade Federal da Fronteira Sul, Campus Chapecó, Brazil), Yousif Subhi (Rigshospitalet, Copenhagen, Denmark, Department of Clinical Research, University of Southern Denmark, Odense, Denmark, Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark), Egor Chumakov (Department of Psychiatry and Addiction, St. Petersburg State University, St. Petersburg, Russia), Sophie Curbo (Department of Laboratory Medicine, Division of Clinical Microbiology, Karolinska Institutet, Sweden), Aybars Kivrak (Department of Orthopaedics and Traumatology, Avrupa Hospital, Adana, Turkey), Ery Ayelen Ko (Universidad Favaloro, Ciudad Autónoma de Buenos Aires, Argentina), Myeong Soo Lee (KM Science Research Division, Korea Institute of Oriental Medicine, Daejeon, South Korea), Dengxiong Li (Department of Urology, Institute of Urology, West China Hospital, Sichuan University, Chengdu, China), Andrey Litvin (Gomel University Clinic, Gomel State Medical University, Belarus), Peng Liu (Center for Psychological Sciences, Zhejiang University, Hangzhou, China), Sebastian Porsdam Mann (Center for Advanced Studies in Bioscience Innovation Law (CeBIL), Faculty of Law, University of Copenhagen, Denmark, Faculty of Law, University of Oxford, Oxford, UK), José Darío Martínez-Ezquerro (Unidad de Investigación Epidemiológica y en Servicios de Salud, Área Envejecimiento (UIESSAE), Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social (IMSS), Mexico City, Mexico, Centro de Ciencias de la Complejidad (C3), Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico), Surapaneni Krishna Mohan (Department of Biochemistry, Panimalar Medical College Hospital and Research Institute, Varadharajapuram, Poonamallee, Chennai - 600 123, Tamil Nadu, India), Philip Moons (KU Leuven Department of Public Health and Primary Care, KU Leuven - University of Leuven, Kapucijnenvoer 35 PB7001, 3000 Leuven, Belgium, Institute of Health and Care Sciences, University of Gothenburg, Arvid Wallgrens backe 141346 Gothenburg, Sweden, Department of Paediatrics and Child Health, University of Cape Town, Klipfontein Rd, Rondebosch, 7700 Cape Town, South Africa), Alejandro Quiroga-Garza (Human Anatomy Department, School of Medicine, Universidad Autonoma de Nuevo Leon, Monterrey 64460, Mexico), Riaz Qureshi (Department of Ophthalmology, Department of Epidemiology, University of Colorado Anschutz Medical Campus, USA), Ximing Xu (Big Data Center for Children's Medical Care, Children's Hospital of Chongqing Medical University, National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing, China), Stephen R Ali (Reconstructive Surgery and Regenerative Medicine Research Centre, Institute of Life Sciences, Swansea University Medical School, Swansea SA2 8PP, UK), Nash Anderson (Tuggeranong Chiropractic Centre, Canberra, Australian Capital Territory, Australia, Canberra City Chiro, Canberra, Australian Capital Territory, Australia), Hiroj Bagde (Department of Periodontology, Chhattisgarh Dental College and Research Institute, Uttar Pradesh, India), Charlotte Blease (Participatory eHealth and Health Data Research Group, Department of Women's and Children's Health, Uppsala Universitet, Uppsala, Sweden, Digital Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, 25 Shattuck Street 3, Boston, Massachusetts, USA), Randy D'Amico (Department of Neurological Surgery, Lenox Hill Hospital/Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, New York, New York, USA),

Hannah Decker (University of California, San Francisco Department of Surgery, USA), Adrian Egli (Institute of Medical Microbiology, University of Zurich, Switzerland), Shijian Feng (Department of Urology and Institute of Urology (Laboratory of Reconstructive Urology), State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, Chengdu, China), Sheng Li (Department of Anorectal Surgery, Ningbo No.2 Hospital, Ningbo, Zhejiang, China), Nav Persaud (Department of Family and Community Medicine, University of Toronto, Toronto, Ontario, Canada), Murali Ramanathan (Department of Neuroscience, School of Translational Medicine, Monash University, Australia), Gemma Sharp (Department of Neuroscience, School of Translational Medicine, Monash University, Australia), Ye Wang (School of Public Health, Lanzhou University, Lanzhou, China), Wah Yang (Department of Metabolic and Bariatric Surgery, The First Affiliated Hospital of Jinan University, Guangzhou, China), Qing-xin Yu (Department of Pathology, Ningbo Clinical Pathology Diagnosis Center, Ningbo City, Zhejiang Province, China).

ORCID iDs
Mauro Giuffrè http://orcid.org/0000-0002-9910-3514
Kyle Lam http://orcid.org/0000-0001-6407-4912
Che-Wei Hsu http://orcid.org/0000-0002-3297-3961
Akihiko Ozaki http://orcid.org/0000-0003-4415-9657
Yaolong Chen http://orcid.org/0000-0002-9841-5233

References

1 **Perez-Lopez R**, Ghaffari Laleh N, Mahmood F, *et al*. A guide to artificial intelligence for cancer researchers. *Nat Rev Cancer* 2024;24:427–41.

2 **Allen JW**, Earp BD, Koplin J, *et al*. Consent-GPT: is it ethical to delegate procedural consent to conversational AI? *J Med Ethics* 2024;50:77–83.

3 **Meng X**, Yan X, Zhang K, *et al*. The application of large language models in medicine: A scoping review. *iScience* 2024;27:109713.

4 **Lin Z**. Techniques for supercharging academic writing with generative AI. *Nat Biomed Eng* 2024;9:426–31.

5 **Nordling L**. How ChatGPT is transforming the postdoc experience. *Nature New Biol* 2023;622:655–7.

6 **Ordak M**. Using ChatGPT in Statistical Analysis: Recommendations for JACC Journals. *JACC Adv* 2024;3:100776.

7 Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature New Biol* 2023;613:612.

8 **Ganjavi C**, Eppler MB, Pekcan A, *et al*. Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: bibliometric analysis. *BMJ* 2024;384:e077192.

9 **Zhao X**, Cox A, Cai L. ChatGPT and the digitisation of writing. *Humanit Soc Sci Commun* 2024;11:482.

10 **Naddaf M**. ChatGPT generates fake data set to support scientific hypothesis. *Nature New Biol* 2023;623:895–6.

11 **Anderson N**, Belavy DL, Perle SM, *et al*. AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of ChatGPT and AI in Sports & Exercise Medicine manuscript generation. *BMJ Open Sport Exerc Med* 2023;9:e001568.

12 **Wu X**, Duan R, Ni J. Unveiling security, privacy, and ethical concerns of ChatGPT. *Journal of Information and Intelligence* 2024;2:102–15.

13 **Zhou J**, Müller H, Holzinger A, *et al*. Ethical ChatGPT: Concerns, Challenges, and Commandments. *Electronics (Basel)* 2024;13:3417.

14 **Heyman T**, Heyman G. The impact of ChatGPT on human data collection: A case study involving typicality norming data. *Behav Res Methods* 2024;56:4974–81.

15 Moher D, Collins G, Hoffmann T, *et al*. Reporting on data sharing: executive position of the EQUATOR Network. *BMJ* 2024;386:e079694.

16 Collins GS, Moons KGM, Dhiman P, *et al*. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378.

17 Sounderajah V, Ashrafian H, Aggarwal R, *et al*. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: The STARD-AI Steering Group. *Nat Med* 2020;26:807–8.

18 Cruz Rivera S, Liu X, Chan A-W, *et al*. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med* 2020;26:1351–63.

19 Liu X, Cruz Rivera S, Moher D, *et al*. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020;26:1364–74.

20 Vasey B, Nagendran M, Campbell B, *et al*. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med* 2022;28:924–33.

21 Cacciamani GE, Collins GS, Gill IS. ChatGPT: standard reporting guidelines for responsible use. *Nature New Biol* 2023;618:238.

22 Tejani AS, Klontzas ME, Gatti AA, *et al*. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update. *Radiol Artif Intell* 2024;6:e240300.

23 Gallifant J, Afshar M, Ameen S, *et al*. The TRIPOD-LLM reporting guideline for studies using large language models. *Nat Med* 2025;31:60–9.

24 Macdonald H, Abbasi K. Riding the whirlwind: BMJ's policy on artificial intelligence in scientific publishing. *BMJ* 2023;382:1923.

25 Flanagin A, Pirracchio R, Khera R, *et al*. Reporting Use of AI in Research and Scholarly Publication-JAMA Network Guidance. *JAMA* 2024;331:1096–8.

26 Moy L. Guidelines for Use of Large Language Models by Authors, Reviewers, and Editors: Considerations for Imaging Journals. *Radiology* 2023;309:e239024.

27 Luo X, Estill J, Chen Y. The use of ChatGPT in medical research: do we need a reporting guideline? *Int J Surg* 2023;109:3750–1.

28 International Committee of Medical Journal Editors (ICMJE). Recommendations for the conduct, reporting, editing and publication of scholarly work in medical journals. 2024. Available: https://www.icmje.org/recommendations

29 COPE position statement: Authorship and AI tools, Available: https://publicationethics.org/cope-position-statements/ai-author

30 WAME Recommendations on Chatbots and Generative Artificial Intelligence in Relation to Scholarly Publications, Available: https://wame.org/page3.php?id=106

31 SAGE. The distinction between Assistive AI tools and Generative AI tools, Available: https://group.sagepub.com/assistive-and-generative-ai-guidelines-for-authors

32 Islam N, van der Schaar M. Use of generative artificial intelligence in medical research. *BMJ* 2024;384:q119.

33 Moher D, Schulz KF, Simera I, *et al*. Guidance for developers of health research reporting guidelines. *PLoS Med* 2010;7:e1000217.

34 Luo X, Tham YC, Daher M, *et al*. Protocol for developing the reporting guideline for the use of chatbots and other Generative Artificial intelligence tools in MEdical Research (GAMER). medRxiv [Preprint].

35 Yin S, Huang S, Xue P, *et al*. Generative artificial intelligence (GAI) usage guidelines for scholarly publishing: a cross-sectional study of medical journals. *BMC Med* 2025;23:77.

36 Assessment of the efficacy of alkaline water in conjunction with conventional medication for the treatment of chronic gouty arthritis: A randomized controlled study: Retraction. *Medicine (Baltimore)* 2024;103:e38913.

37 Zhang M, Wu L, Yang T, *et al*. RETRACTED: The three-dimensional porous mesh structure of Cu-based metal-organic-framework - Aramid cellulose separator enhances the electrochemical performance of lithium metal anode batteries. *Surfaces and Interfaces* 2024;46:104081.

38 Sreekumar S, Kiran MS. Localized trans-browning and pro-angiogenesis inductive self-assembled collagen resveratrol bio-matrix for tissue repair and regeneration in obese conditions. *Int J Biol Macromol* 2024;263:130322.

39 Luo X, Chen F, Zhu D, *et al*. Potential Roles of Large Language Models in the Production of Systematic Reviews and Meta-Analyses. *J Med Internet Res* 2024;26:e56780.

40 Jazi AHD, Mahjoubi M, Shahabi S, *et al*. Bariatric Evaluation Through AI: a Survey of Expert Opinions Versus ChatGPT-4 (BETA-SEOV). *Obes Surg* 2023;33:3971–80.

41 Kim HJ, Yang JH, Chang D-G, *et al*. Assessing the Reproducibility of the Structured Abstracts Generated by ChatGPT and Bard Compared to Human-Written Abstracts in the Field of Spine Surgery: Comparative Analysis. *J Med Internet Res* 2024;26:e52001.

42 Huang AS, Hirabayashi K, Barna L, *et al*. Assessment of a Large Language Model's Responses to Questions and Cases About Glaucoma and Retina Management. *JAMA Ophthalmol* 2024;142:371–5.

43 Perlis RH, Goldberg JF, Ostacher MJ, *et al*. Clinical decision support for bipolar depression using large language models. *Neuropsychopharmacology* 2024;49:1412–6.

44 Lim ZW, Pushpanathan K, Yew SME, *et al*. Benchmarking large language models' performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard. *EBioMedicine* 2023;95:104770.

45 Eriksen AV, Möller S, Ryg J. Use of GPT-4 to Diagnose Complex Clinical Cases. *NEJM AI* 2024;1.

46 Williams CYK, Zack T, Miao BY, *et al*. Use of a Large Language Model to Assess Clinical Acuity of Adults in the Emergency Department. *JAMA Netw Open* 2024;7:e248895.

47 Ayers JW, Poliak A, Dredze M, *et al*. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med* 2023;183:589–96.

48 Giray L. Prompt Engineering with ChatGPT: A Guide for Academic Writers. *Ann Biomed Eng* 2023;51:2629–33.

49 Meskó B. Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial. *J Med Internet Res* 2023;25:e50638.

50 Gartlehner G, Kahwati L, Hilscher R, *et al*. Data extraction for evidence synthesis using a large language model: A proof-of-concept study. *Res Synth Methods* 2024;15:576–89.

51 Huang J, Yang DM, Rong R, *et al*. A critical assessment of using ChatGPT for extracting structured data from clinical notes. *NPJ Digit Med* 2024;7:106.

52 Peng D, Zheng L, Liu D, *et al*. Large-language models facilitate discovery of the molecular signatures regulating sleep and activity. *Nat Commun* 2024;15:3685.

53 Liu S, McCoy AB, Wright AP, *et al*. Leveraging large language models for generating responses to patient messages—a subjective analysis. *J Am Med Inform Assoc* 2024;31:1367–79.

54 Salem AC, Gale RC, Fleegle M, *et al*. Automating Intended Target Identification for Paraphasias in Discourse Using a Large Language Model. *J Speech Lang Hear Res* 2023;66:4949–66.

55 Chen CC, Wei CJ, Tseng TY, *et al*. Applying Object Detection and Large Language Model to Establish a Smart Telemedicine Diagnosis System with Chatbot: A Case Study of Pressure Injuries Diagnosis System. *Telemed J E Health* 2024;30:e1705–12.

56 Jamrat S, Sukasem C, Sratthaphut L, *et al*. A precision medicine approach to personalized prescribing using genetic and nongenetic factors for clinical decision-making. *Comput Biol Med* 2023;165:107329.

57 Tayebi Arasteh S, Han T, Lotfinia M, *et al*. Large language models streamline automated machine learning for clinical studies. *Nat Commun* 2024;15:1603.

58 Castellanos-Gomez A. Good Practices for Scientific Article Writing with ChatGPT and Other Artificial Intelligence Language Models. *Nanomanufacturing* 2023;3:135–8.

59 Magalhães Araujo S, Cruz-Correia R. Incorporating ChatGPT in Medical Informatics Education: Mixed Methods Study on Student Perceptions and Experiential Integration Proposals. *JMIR Med Educ* 2024;10:e51151.

60 Perkins M, Roe J. Academic publisher guidelines on AI usage: A ChatGPT supported thematic analysis. *F1000Res* 2023;12:1398.

61 Blanco-González A, Cabezón A, Seco-González A, *et al*. The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. *Pharmaceuticals (Basel)* 2023;16:891.

62 **Pu Z**, Shi C-L, Jeon CO, *et al*. ChatGPT and generative AI are revolutionizing the scientific community: A Janus-faced conundrum. *Imeta* 2024;3:e178.

63 **Song M**, Wang J, Yu Z, *et al*. PneumoLLM: Harnessing the power of large language model for pneumoconiosis diagnosis. *Med Image Anal* 2024;97:103248.

64 **Kacena MA**, Plotkin LI, Fehrenbacher JC. The Use of Artificial Intelligence in Writing Scientific Review Articles. *Curr Osteoporos Rep* 2024;22:115–21.

65 **Suwała S**, Szulc P, Guzowski C, *et al*. ChatGPT-3.5 passes Poland's medical final examination-Is it possible for ChatGPT to become a doctor in Poland? *SAGE Open Med* 2024;12:20503121241257777.

66 **Chen J**, Liu Z, Huang X, *et al*. When large language models meet personalization: perspectives of challenges and opportunities. *World Wide Web* 2024;27:42.

67 **Hoppe JM**, Auer MK, Strüven A, *et al*. ChatGPT With GPT-4 Outperforms Emergency Department Physicians in Diagnostic Accuracy: Retrospective Analysis. *J Med Internet Res* 2024;26:e56110.

68 **Au K**, Yang W. Auxiliary use of ChatGPT in surgical diagnosis and treatment. *Int J Surg* 2023;109:3940–3.

69 **Masanneck L**, Schmidt L, Seifert A, *et al*. Triage Performance Across Large Language Models, ChatGPT, and Untrained Doctors in Emergency Medicine: Comparative Study. *J Med Internet Res* 2024;26:e53297.

70 **Elsevier**. The use of generative AI and AI-assisted technologies in the review process for Elsevier: Policy for book and commissioned content reviewers, 2024. Available: https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-the-review-process [Accessed 20 Aug 2024].

71 **Ueda D**, Walston SL, Matsumoto T, *et al*. Evaluating GPT-4-based ChatGPT's clinical potential on the NEJM quiz. *BMC Digit Health* 2024;2.

72 **Boulogne LH**, Lorenz J, Kienzle D, *et al*. The STOIC2021 COVID-19 AI challenge: Applying reusable training methodologies to private data. *Med Image Anal* 2024;97:103230.

73 **Lim SS**, Semnani-Azad Z, Morieri ML, *et al*. Reporting guidelines for precision medicine research of clinical relevance: the BePRECISE checklist. *Nat Med* 2024;30:1874–81.

74 **Kolbinger FR**, Veldhuizen GP, Zhu J, *et al*. Reporting guidelines in medical artificial intelligence: a systematic review and meta-analysis. *Commun Med (Lond)* 2024;4:71.

75 **Birhane A**, Kasirzadeh A, Leslie D, *et al*. Science in the age of large language models. *Nat Rev Phys* 2023;5:277–80.

76 **Collaborative C**. Protocol for the development of the Chatbot Assessment Reporting Tool (CHART) for clinical advice. *BMJ Open* 2024;14:e081155.