



OPEN

Use of large language models as artificial intelligence tools in academic research and publishing among global clinical researchers

Tanisha Mishra¹, Edward Sutanto^{2,3}, Rini Rossanti⁴, Nayana Pant⁵, Anum Ashraf⁶, Akshay Raut⁷, Germaine Uwabareze⁸, Ajayi Oluwatomiwa⁹ & Bushra Zeeshan^{10,11}✉

With breakthroughs in Natural Language Processing and Artificial Intelligence (AI), the usage of Large Language Models (LLMs) in academic research has increased tremendously. Models such as Generative Pre-trained Transformer (GPT) are used by researchers in literature review, abstract screening, and manuscript drafting. However, these models also present the attendant challenge of providing ethically questionable scientific information. Our study provides a snapshot of global researchers' perception of current trends and future impacts of LLMs in research. Using a cross-sectional design, we surveyed 226 medical and paramedical researchers from 59 countries across 65 specialties, trained in the Global Clinical Scholars' Research Training certificate program of Harvard Medical School between 2020 and 2024. Majority (57.5%) of these participants practiced in an academic setting with a median of 7 (2,18) PubMed Indexed published articles. 198 respondents (87.6%) were aware of LLMs and those who were aware had higher number of publications ($p < 0.001$). 18.7% of the respondents who were aware ($n = 37$) had previously used LLMs in publications especially for grammatical errors and formatting (64.9%); however, most (40.5%) did not acknowledge its use in their papers. 50.8% of aware respondents ($n = 95$) predicted an overall positive future impact of LLMs while 32.6% were unsure of its scope. 52% of aware respondents ($n = 102$) believed that LLMs would have a major impact in areas such as grammatical errors and formatting (66.3%), revision and editing (57.2%), writing (57.2%) and literature review (54.2%). 58.1% of aware respondents were opined that journals should allow for use of AI in research and 78.3% believed that regulations should be put in place to avoid its abuse. Seeing the perception of researchers towards LLMs and the significant association between awareness of LLMs and number of published works, we emphasize the importance of developing comprehensive guidelines and ethical framework to govern the use of AI in academic research and address the current challenges.

Keywords Large language models, Artificial intelligence, Academic writing, Biomedical research

Large Language Models (LLMs) represent a significant breakthrough in Natural Language Processing (NLP) and Artificial Intelligence (AI)¹. Prior to 2017, while NLP models could perform several language processing tasks, they were not easily accessible to non-domain experts. The introduction of the Transformer architecture in 2017 revolutionized the field, enabling NLP models to efficiently synthesize and analyze datasets using simple prompts. This allowed large-scale use by people worldwide, significantly broadening access to advanced language processing tools². The Transformer technology led to the development of two game changers: Bidirectional

¹Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India.

²Nuffield Department of Medicine, Centre for Tropical Medicine and Global Health, University of Oxford, Oxford OX3 7LG, UK. ³Faculty of Medicine, Oxford University Clinical Research Unit Indonesia, Universitas Indonesia, Jakarta 10430, Indonesia. ⁴Department of Child Health, Dr. Hasan Sadikin General Hospital/Faculty of Medicine, Universitas Padjadjaran, Bandung, Indonesia. ⁵Royal Free NHS Foundation Trust Hospital, Pond Street, London NW32QG, UK. ⁶Department of Pharmacology & Therapeutics, Allama Iqbal Medical College, Jinnah Hospital, Lahore, Pakistan. ⁷Department of Internal Medicine, Guthrie Robert Packer Hospital, Sayre, PA 18840, USA. ⁸World Health Organization, Comoros Country Office, Moroni, Comoros. ⁹Federal Neuropsychiatric Hospital Yaba, Lagos, Nigeria. ¹⁰Department of Dermatology, Niazi Hospital, Lahore, Pakistan. ¹¹Allama Iqbal Medical College, Jinnah Hospital, Lahore, Pakistan. ✉email: dr_bushrahassan@hotmail.com

Encoder Representations from Transformers (BERT) and Generative Pretrained Transformer (GPT), that used semi-supervised approach and acquired exceptional generalization capabilities with the ability to interpret and generate human-like text³. The launch of ChatGPT in 2022 gained public attention in almost every field of life owing to its accessibility and user-friendly interface. LLMs offer AI driven support particularly in literature review, summarizing articles, abstract screening, extracting data and drafting manuscript. Due to workload reduction and the ease offered, there has been an increasing interest in the incorporation of LLMs like ChatGPT, Perplexity, Llama by Meta (formerly Facebook), Google Bard and Claude, in academic research, as indicated by a rapid increase in the number of articles after ChatGPT's release^{3,4}.

Although there are numerous efficiency gains in utilizing LLMs in research, they however cannot replace humans particularly in contexts where meticulous understanding and original thought, along with accountability are crucial^{5,6}. With deeper understanding of LLMs, it was found that LLMs are also capable of generating fake citations, rapidly generating large volumes of questionable information, and also amplifying biases^{3,7}. This has led to negative ethical implications like authorship integrity and a surge in predatory practices and as a consequence, an "AI-driven infodemic" has emerged⁵. There is also a risk of public health threat resulting from ghost-written scientific articles, fake news and misinforming content³. In addressing these issues, as a first step, it is pertinent to understand attitudes of researchers towards LLMs in research by assessing researcher's awareness and practices of the use of LLMs.

Our study provides a unique analysis of a targeted group of medical and paramedical researchers enrolled in a one-year-certification course- Global Clinical Scholars Research Training (GCSRT) Program, at Harvard Medical School (HMS). We aim to provide insights into the current trends in AI usage in research and publication along with a peek into the future scope and impact of LLMs. We strongly believe that the results of our study can aid journals to formulate future policies regarding the use of AI tools in the process of publication, thus ensuring credibility and maintaining integrity of medical publications.

Methods

Study design and population

This global survey was carried out using a cross-sectional design. It was conducted between April and June 2024, amongst a diverse group of medical and paramedical researchers who received training at the GCSRT program at Harvard. This program consists of researchers from over 50 countries and 6 continents spanning various specialties, career stages, age groups and genders. At the program, all participants receive advanced training in every stage of research including statistical analysis, publishing and grant writing⁸. They are therefore an ideal group to assess AI tools usage in research.

Study objectives

We had three primary objectives for this study. First, to assess the level of awareness of LLMs amongst global researchers. Second, to identify how LLMs are currently used in academic research and publishing amongst our survey respondents. And third, to analyze the potential future impact and ethical implications of AI tools in medical research and publishing.

Eligibility criteria

- (a) Inclusion Criteria: Medical and paramedical researchers who have been participants of the GCSRT program at HMS belonging to any cohort between 2020 and 2024, irrespective of their country of origin, research interests, active years in research, age or gender. Researchers who were members of the unofficial class WhatsApp groups and were proficient in reading and writing in English language were included specifically.
- (b) Exclusion Criteria: Researchers from cohorts outside of the above specified years, those who were not accessible through class WhatsApp groups, or were not proficient in reading and writing in English language were excluded from the study. Medical and paramedical researchers who have not undergone training at this program as well as non-medical researchers, were not invited for this study.

Questionnaire development and survey dissemination strategy

The survey was drafted using Google Forms, in English Language. It consisted of a total of 4 sections to cover our primary objectives- (1) Background, (2) Awareness of LLMs, (3) Impact of LLMs and (4) Future Policy. Each question was carefully reviewed for its relevance, validity, and unbiasedness. Data collectors for the study were voluntarily chosen from amongst the participants of the GCSRT Program. The data collectors from each of the targeted cohorts were made primary in-charge of reaching out to our target population in their cohort via personal messaging on WhatsApp and LinkedIn. The contact information of the survey respondents was obtained from the unofficial class WhatsApp groups and personal networks of the data collectors. A total of 3 personal messages including 2 reminders, spaced 7 days apart each, were sent to each prospective participant. Informed consent was obtained, and Google survey forms were filled out by a total of 226 researchers from over 59 countries.

Sample size and statistical methods

The link to the Google survey form was distributed to 5 cohorts of the GCSRT program consisting of a total of 550 medical and paramedical researchers. A total sample size of 220 was calculated by considering a margin of error of 5%, a confidence level of 95% and power of 0.8. Descriptive statistics of the survey respondents were presented as mean \pm standard deviation for normally distributed continuous data, median (interquartile range) for non-normally distributed continuous data, and frequencies & percentages for categorical data. Continuous

data were tested for normality using the Shapiro–Wilk test. Normally distributed data were analyzed using one way ANOVA while non-normally distributed data were analyzed using the Kruskal–Wallis test. Categorical data were analyzed with Chi-squared test or Fisher’s exact test. Qualitative data from open-ended questions were studied via thematic analysis. All statistical analyses were performed in Stata MP version 17.0 (StataCorp, College Station, TX, USA). All tests were two-tailed and considered significant at $P < 0.05$.

Ethical consideration

In accordance with the declaration of Helsinki⁸, this study was approved by the ethical review board at Allama Iqbal Medical College/ Jinnah Hospital Lahore, Pakistan (Reference no: ERB 163/9/30-04-2024/S1 ERB). It is not supported or endorsed by HMS. However, timely notification about the study was provided to the administration of the GCSRT program. Consent to participate was collected from every respondent as the first, mandatory response to the questionnaire. All personal information like email ID, nationality, and age was carefully de-identified and handled confidentially. The respondents were provided necessary information on the voluntariness of the study as well as contact information of the principal investigator.

Results

We analyzed the responses of 226 global researchers from over 59 countries and practicing across 65 different medical and paramedical specialties. Across the various countries of origin (Supplementary Table S1), the two most common regions of origin were the region of Americas (23.5%) and the South-East Asian region (23.5%).

Table 1 Represents academic and demographic characteristics of our survey respondents and compares them amongst respondents who were aware of LLMs to those who were not. The median of PubMed indexed

Characteristics	Overall (N = 226)	Awareness of LLMs		p-value
		Yes (n = 198)	No (n = 28)	
Age (in years)				
21–40	133 (58.8)	118 (59.6)	15 (53.6)	0.688
41–60	86 (38.0)	74 (37.4)	12 (42.9)	
60+	7 (3.2)	6 (3.0)	1 (3.5)	
Region of Origin				
African Region	30 (13.3)	27 (13.6)	3 (10.7)	0.728
Region of the Americas	53 (23.5)	45 (22.7)	8 (28.6)	
South-East Asian Region	53 (23.5)	49 (24.8)	4 (14.3)	
European Region	29 (12.8)	26 (13.1)	3 (10.7)	
Eastern Mediterranean Region	41 (18.1)	34 (17.2)	7 (25.0)	
Western Pacific Region	20 (8.8)	17 (8.6)	3 (10.7)	
Area of Practice*				
Medical Specialty	146 (64.6)	127 (64.1)	19 (67.8)	.**
Surgical Specialty	57 (25.2)	52 (26.2)	5 (17.8)	
Pharmacy	8 (3.5)	7 (3.5)	1 (3.5)	
Clinical Research	62 (27.4)	53 (26.8)	9 (32.1)	
Other	14 (6.2)	14 (100.0)	0 (0.0)	
Type of Practice*				
Academic (such as college or university)	129 (57.1)	117 (59.1)	12 (42.9)	.**
Public/Community hospital, Health Center, or Clinic	68 (30.1)	57 (28.8)	11 (39.3)	
Private Hospital, Health Center, or Clinic	56 (24.8)	49 (24.7)	7 (25.0)	
Non-profit Organization	24 (10.6)	23 (11.6)	1 (3.6)	
Industry / Pharmaceutical Company	26 (11.5)	20 (10.1)	6 (21.4)	
Government, such as Ministry of Health	20 (8.8)	18 (9.1)	2 (7.1)	
Other	6 (2.7)	5 (2.5)	1 (3.6)	
Number of PubMed Indexed Publications	7 (2–18)	8 (3–20)	3 (0–7)	<0.001
Active Years in Medical Research				
0–5 years	106 (46.9)	89 (45.0)	17 (60.7)	0.101
6–10 years	71 (31.4)	67 (33.8)	4 (14.3)	
10 + years	49 (21.7)	42 (21.2)	7 (25.0)	

Table 1. Academic and Demographic Characteristics of Survey Respondents Abbreviations: LLMs, large language models. Data are presented in n (%) or median (interquartile range). *May add up to more than 100% as respondents were able to select multiple options. **No statistical test was conducted as individuals may select multiple options.

publications among survey respondents were 7 (interquartile range: 2–18). 198 (87.6%) survey respondents were previously aware of LLMs. None of the characteristics were significantly associated with awareness of LLMs, except the number of PubMed indexed publications. Those who were aware of the use of LLMs have a higher number of publications compared to those who were not aware ($p < 0.001$).

Table 2 represents aware respondents' ($n = 198$) knowledge, attitudes and practices with respect to LLMs. Most were somewhat and moderately familiar with LLMs (33.3% and 30.8%, respectively). Of these aware respondents, the ones that have personally used LLMs previously (18.7%), mostly used it for grammatical error and formatting (64.9%), followed by writing (45.9%) and finally revision and editing (45.9%). When stratified with the number of active years in medical research, none of these variables were significantly associated.

Figure 1 displays the level of perceived future impact of LLMs in various stages of publication amongst aware respondents. Majority believe that LLMs will have a major overall impact (52.0%). Areas that will be majorly impacted are grammatical errors and formatting (66.3%), revision and editing (57.2%), and writing (57.2%). Areas that will not be impacted to moderately impacted were methodology (74.3%), journal selection (73.3%), and study ideas (71.1%).

Table 3 represents aware respondent perceptions on future scope of LLMs. Majority perceived that it will bring a positive impact (50.8%), yet a significant proportion were unsure (32.6%). While most respondents believe that journal should allow usage of AI tools in publishing (58.1%), the majority (78.3%) also believe that some regulations (i.e. modified journal policies, AI review boards, tools to detect LLM usage) should be put in

Variables	Overall ($n = 198$)	Active Years In Academic Research			p-value
		0–5 years ($n = 89$)	6–10 years old ($n = 67$)	10 + years old ($n = 42$)	
Level of Awareness					
Slightly familiar	57 (28.8)	22 (24.7)	21 (31.3)	14 (33.3)	0.586
Somewhat familiar	66 (33.3)	36 (40.5)	20 (29.9)	10 (23.8)	
Moderately familiar	61 (30.8)	26 (29.2)	21 (31.3)	14 (33.3)	
Extremely familiar	14 (7.1)	5 (5.6)	5 (7.5)	4 (9.5)	
Awareness of LLMs Prior to 2022					
Yes	27 (13.6)	10 (11.2)	12 (17.9)	5 (11.9)	0.453
No	171 (86.4)	79 (88.8)	55 (82.1)	37 (88.1)	
Usage of LLMs in Previous Research/Publication					
Yes	37 (18.7)	16 (18.0)	9 (13.4)	12 (28.6)	0.139
No	161 (81.3)	73 (82.0)	58 (86.6)	30 (71.4)	
Level of Use					
1 (Rarely)	6 (16.2)	2 (12.5)	1 (11.1)	3 (25.0)	0.891
2	4 (10.8)	1 (6.3)	1 (11.1)	2 (16.7)	
3 (Moderately)	11 (29.7)	5 (31.3)	3 (33.3)	3 (25.0)	
4	8 (21.6)	4 (25.0)	3 (33.3)	1 (8.3)	
5 (Frequently)	8 (21.6)	4 (25.0)	1 (11.1)	3 (25.0)	
Stage of Publication LLM Used Previously*					
Study ideas	14 (37.8)	6 (37.5)	3 (33.3)	5 (41.7)	.**
Literature Review	15 (40.5)	10 (62.5)	2 (22.2)	3 (25.0)	
Methodology	9 (24.3)	4 (25.0)	3 (33.3)	2 (16.7)	
Data Analysis	10 (27.0)	3 (18.8)	3 (33.3)	4 (33.3)	
Writing	17 (45.9)	9 (56.3)	2 (22.2)	6 (50.0)	
Journal Selection	2 (5.4)	1 (6.3)	0 (0.0)	1 (8.3)	
Revision and Editing	17 (45.9)	8 (50.0)	5 (55.6)	4 (33.3)	
Grammatical error and formatting	24 (64.9)	10 (62.5)	6 (66.7)	8 (66.7)	
References	5 (13.5)	2 (12.5)	1 (11.1)	2 (16.7)	
Acknowledgement of Use of LLM in Previous Publication*					
Did not acknowledge the use of LLM	15 (40.5)	5 (31.3)	4 (44.4)	6 (50.0)	.**
Cited the specific LLM model used (e.g., GPT-3) in the acknowledgments	6 (16.2)	4 (25.0)	1 (11.1)	1 (8.3)	
Included a statement in the Methods section describing the use of LLMs	7 (18.9)	4 (25.0)	1 (11.1)	2 (16.7)	
Disclosed LLM usage in the funding or conflict of interest statement	3 (8.1)	1 (6.3)	0 (0.0)	2 (16.7)	
Provided a footnote or disclaimer regarding LLM-generated content	5 (13.5)	4 (25.0)	0 (0.0)	1 (8.3)	
Others	9 (24.3)	2 (12.5)	3 (33.3)	4 (33.3)	

Table 2. Knowledge, attitudes and practices of Aware respondents. Abbreviations: LLMs, large language models. Data are presented in n (%). *May add up to more than 100% as respondents were able to select multiple options. **No statistical test was conducted as respondents may select multiple options.

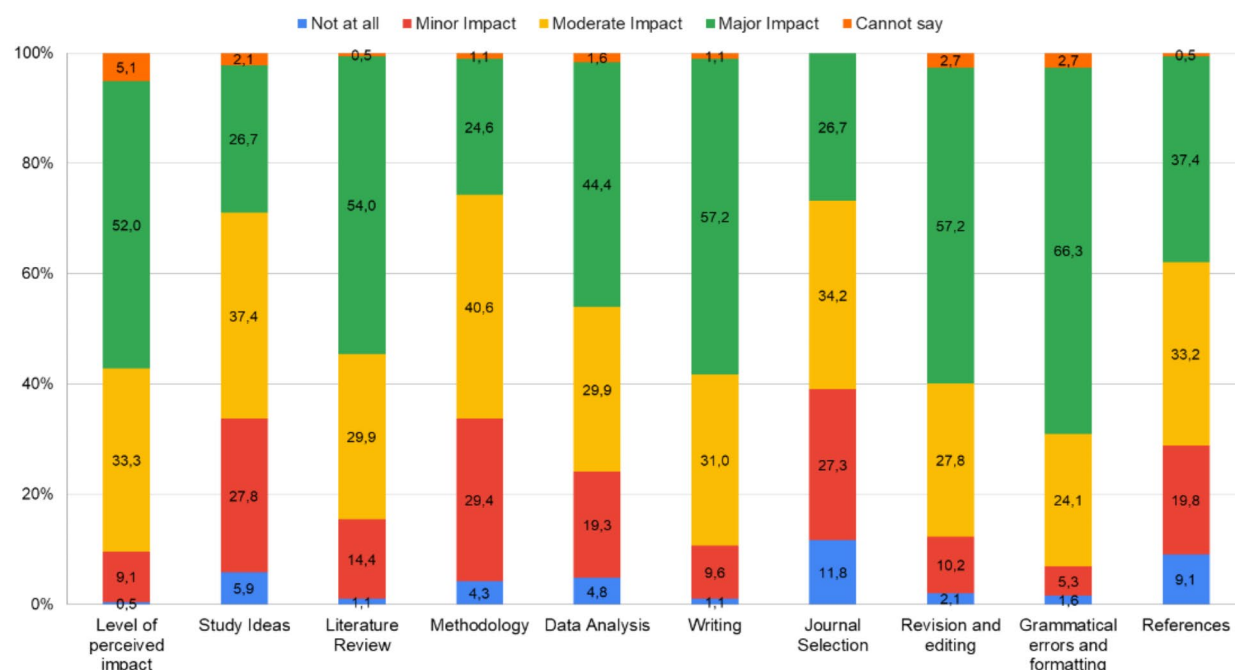


Fig. 1. Future of LLMs in various stages of the publication process.

place to make AI tools in publishing ethical. When stratified with the number of active years in medical research, none of these variables were significantly associated.

In our survey, 79% ($n=179$) of the respondents were willing to share their overall opinion into future scope and challenges of LLMs. Their views fall into one or more categories in Fig. 2. 28% ($n=64$) respondents expressed that LLMs are helpful tools in the publication process, particularly in organizing and writing large topics in a systematic way. Additionally, around a quarter of respondents ($n=55$) stated that with the use of LLMs, researchers are able to spend less time on different sections of their research projects such as literature review, data analysis, and manuscript preparation. However, the survey respondents also revealed several concerns and challenges associated with using LLMs in academic research. 14% ($n=33$) of the respondents expressed uncertainty or their lack of experience with LLMs. Ethical apprehensions about the use of LLMs in academic research and publication, including potential biases, privacy issues, and plagiarism, were noted in 8% of the participants ($n=18$).

Discussion

AI has generated seismic waves around the world; the field of research is no exception. Our study assessed the awareness, trends of usage and future scope of LLMs to better analyze this impact in the field of academia. It captured the perception of researchers from all walks of medical and paramedical research representing 59 countries and spanning 65 specialties. Our respondents mainly belonged to Medical subspecialties (64.6%) rather than Surgical or Paramedical subspecialties, similar (68%) to respondent characteristics seen in a study by Abdelhafiz et al.⁹ Our respondents were mostly working in academic settings (57.1%) followed by public and private healthcare settings, similar to the study by Abdelhafiz et al. where 75% of participants were from universities or research centers⁹. The respondents with 10+, 6–10 and 0–5 years of research experience were 21.7%, 31.4% and 46.9% respectively, suggesting that our target population well represented academicians at varying stages of their careers.

A significant majority of our respondents (87.6%) were aware of LLMs, which was similar (85%) to a survey conducted among medical students in Jordan¹⁰, and higher in comparison to a study done in Pakistan where only 21.3% of the respondents had familiarity with AI¹¹. A plausible explanation for the high level of awareness amongst GCSRT participants could be that they had already completed an advanced training in research, and might have come across the applications of LLMs in contemporary research and publication during this training period. Also, their keen interest in research might have rendered them to explore the latest advancements in this field, amongst which usage of LLMs and AI tools probably tops the list^{12,13}. Interestingly, the participants who were aware of LLMs had a higher number of publications compared to those who were not (p -value < 0.001). This finding coincides with previous studies where it has been reported that greater familiarity and access to LLMs is associated with a greater pre-print and publication turn-out ratio amongst academic authors, probably due to the fast-paced nature of LLMs research and the use of LLMs for writing assistance¹⁴. None of the other variables like age, country of respondent, or field of practice were significantly associated with awareness of the use of LLMs. An overwhelmingly large proportion of the respondents who were aware of LLMs, reported that they

	Overall (n = 198)	Active Years in Academic Research			p-value
		0–5 years (n = 89)	6–10 years old (n = 67)	10 + years old (n = 42)	
Anticipated Future Impact of LLMs					
Strongly positive	18 (9.6)	7 (8.3)	5 (7.7)	6 (15.8)	0.143
Positive	95 (50.8)	44 (52.4)	27 (41.5)	24 (63.2)	
Cannot say	61 (32.6)	27 (32.1)	27 (41.5)	7 (18.4)	
Negative	9 (4.8)	5 (6.0)	4 (6.1)	0 (0.0)	
Strongly negative	4 (2.1)	1 (1.2)	2 (3.1)	1 (2.6)	
Should Journals Allow Usage of AI Tools in Publications?					
Yes	115 (58.1)	48 (53.9)	39 (58.2)	28 (66.7)	0.741
No	31 (15.7)	15 (16.9)	11 (16.4)	5 (11.9)	
Cannot say	52 (26.3)	26 (29.2)	17 (25.4)	9 (21.4)	
Potential Problems with Using AI in Research Publications*					
Impaired creativity	119 (60.1)	52 (58.4)	43 (64.2)	24 (57.1)	**
Data security	100 (50.5)	43 (48.3)	33 (49.3)	24 (57.1)	
Misinformation	132 (66.7)	60 (67.4)	41 (61.2)	31 (73.8)	
Unintended bias	130 (65.7)	57 (64.0)	45 (67.2)	28 (66.7)	
Lack of accountability	115 (58.1)	59 (66.3)	36 (53.7)	20 (47.6)	
Copyright infringement	125 (63.1)	58 (65.2)	45 (67.2)	22 (52.4)	
Impersonation	92 (46.5)	38 (42.7)	33 (49.3)	21 (50.0)	
Others	16 (8.1)	7 (7.9)	5 (7.4)	4 (9.5)	
Is The Use of AI Tools in Publishing Ethical?					
Yes, in current form	21 (10.6)	9 (10.1)	8 (11.9)	4 (9.5)	0.072
Only if some regulations like modified journal policies, AI review boards or tools to detect LLM usage are put in place	155 (78.3)	65 (73.0)	54 (80.6)	36 (85.7)	
Never	9 (4.5)	9 (10.1)	0 (0.0)	0 (0.0)	
Others	13 (6.6)	6 (6.7)	5 (7.4)	2 (4.8)	

Table 3. Insights into the future scope of LLMs from aware respondents. LLMs, large language models; AI: Artificial intelligence. Data are presented in n (%). *May add up to more than 100% as respondents were able to select multiple options. **No statistical test was conducted as respondents may select multiple options.



Fig. 2. Overall Opinion on Future Scopes & Challenges (Thematic Categories).

were not aware of AI tools prior to 2022 (86.4%). This corresponds to a compelling trajectory of publications pertaining to LLM in medical research, from May 2021 to July 2023¹⁵.

81.3% of our aware respondents never previously used LLMs in their research projects or publications. This is in contrast to Eppler et al.'s earlier study¹⁶, where nearly half of the respondents reported having used LLMs in their academic practice. Amongst those who previously used LLMs in their publications, most rated their usage to be of moderate to frequent intensity for tasks such as grammatical error corrections, editing and manuscript writing. These results were in concordance with the study by Eppler et al.¹⁶, which showed that the most common use of LLMs in scientific publishing was for writing (36.6%) followed by checking grammar (30.6%). With the help of LLMs based on NLP, it is possible to conveniently rectify grammatical errors using categorization models and algorithm-based sentence construction^{17,18}. Despite the frequent use of LLMs for various components of academic writing, a considerable proportion of these respondents (~40%) did not acknowledge its usage in their publications. There are multiple reasons as to why a researcher may not reveal the inclusion of AI tools in their research papers. First, is the lack of information or comprehension on part of the researchers regarding the technologies they are using, due to which they remain oblivious to the degree to which AI has been integrated into their research^{19,20}. And second, is the skepticism or negative perceptions associated with the use of AI – like the notion that a machine was deployed to generate proposals or scientific discussion of their study²¹. Thus, the question of whether or not to acknowledge the use of AI in research studies remains an ethical imbroglio. Publishers may ask authors to submit or include a declaration about whether or not they have used AI systems in their writing^{22–24}.

Figures 1 and 2; Table 3 provide insights into future scopes and challenges of LLMs among global researchers. Figure 1 reveals a substantial belief in the transformative potential of LLMs, with slightly more than half of respondents anticipating a major overall impact. Specific areas identified as being most significantly influenced by LLMs in the future included grammatical errors and formatting, revision and editing, writing, and literature review. These results align with current literature suggesting that LLMs can greatly enhance the efficiency and accuracy of these tasks, thus facilitating quicker and higher-quality academic outputs^{25–27}. Conversely, the areas perceived to be less impacted, such as methodology, journal selection, and study ideas, reflect apprehension about AI to critically assess research design and journal suitability.

As shown in Table 3, slightly more than half of the participants view the impact of LLMs positively, yet around one-third of them remain uncertain. This uncertainty underscores a significant concern regarding the ethical implications and potential misuse of AI technologies. Ethical concerns are well-documented in existing studies that highlight issues such as data privacy, misinformation, and unintended biases that can arise from AI-generated content^{28,29}. In addition, our study reveals that while the majority of respondents support the use of AI tools in publishing, there is a strong consensus on the necessity of implementing regulatory measures, such as modified journal policies, AI review boards, and tools to detect LLM usage. This finding is consistent with broader ethical guidelines proposed in the literature, which advocate for robust oversight and ethical frameworks to mitigate the risks associated with AI deployment in sensitive fields, such as medical research^{27,30}. Interestingly, the perception of AI's ethical use varies with experience levels. In our study, participants with more than 10 years of research experience were more likely to view AI tools as positive and support their use under regulated conditions compared to those with lesser years of research experience; however, this result was not statistically significant.

Conclusions

The discipline of academic writing has seen a noticeable transformation following the advent of LLMs, with an increasing number of researchers incorporating these tools at varying stages of their research publications. However, as the applications of LLMs rise, there is a corresponding rise in the concerns regarding their validity, accountability, potential exploitation and ethical implications.

While there is a broad recognition of the beneficial impact of LLMs on certain aspects of academic research and publishing, addressing associated ethical risks and apprehensions is of paramount importance. Our study emphasizes the need for developing comprehensive guidelines and ethical frameworks to govern the use of AI in medical and paramedical research. The growing utility of LLMs necessitates the implementation of such regulatory policies promptly to ensure their safe, responsible and effective usage.

Limitations

Our study has certain methodological limitations which need to be addressed. First, since this is a cross-sectional study, causal inferences cannot be drawn from the findings as well as the temporal relevance of our study findings are subjected to change over time. Second, despite our extensive attempts to maintain the anonymity of survey responses, the study findings are prone to social desirability bias. Third, since our study population was exclusively limited to the participants of the GCSRT program with extensive knowledge surrounding academic research, a selection bias could have been introduced which limits the generalizability of the study findings. Fourth, our study did not collect several respondent's characteristics that may be associated with awareness of the use of LLMs, such as sex, level of education, and level of income. And finally, our study is susceptible to sampling bias due to the use of WhatsApp and LinkedIn for data collection. Participants using LinkedIn might be concurrently using other platforms for their research, while participants using WhatsApp might be younger and more well-versed in technology and AI. Hence, concerns about the overrepresentation of certain demographics could affect the external validity of the findings.

Data availability

The data of this study will be shared upon reasonable request to the corresponding author.

Received: 25 June 2024; Accepted: 26 November 2024

Published online: 30 December 2024

References

1. Naveed, H. et al. A comprehensive overview of large language models. (2023). <https://arxiv.org/abs/2307.06435>.
2. Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. Improving language understanding by generative pre-training (2018). https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.
3. De Angelis, L. et al. ChatGPT and the rise of large language models: The new AI-driven infodemic threat in public health. *Front. Public Health* **11**, (2023).
4. Bornmann, L., Haunschild, R. & Mutz, R. Growth rates of Modern Science: A latent piecewise growth curve Approach to Model publication numbers from established and New Literature databases. *Humanit. Soc. Sci. Commun.* **8**, (2021).
5. Kendall, G. & Teixeira Da Silva, J. A. Risks of abuse of large language models, like ChatGPT, in scientific publishing: Authorship, predatory publishing, and paper mills. *Learn. Pub* **37**, (2024).
6. Rane, N., Choudhary, S. P., Tawde, A. & Rane, J. ChatGPT is not capable of serving as an author: Ethical concerns and challenges of large language models in education. *IRJMETs* **5**, (2023).
7. Hosseini, M. & Horbach, S. P. J. M. Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Res. Integr. Peer Rev.* **8**, (2023).
8. <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>
9. Abdelhafiz, A. S. et al. Knowledge, perceptions and attitude of researchers towards using ChatGPT in Research. *J. Med. Syst.* **48**, (2024).
10. Al Saad, M. M. et al. Medical students' knowledge and attitude towards Artificial Intelligence: An online survey. *Open. Public. Health J.* **15**, (2022).
11. Umer, M. et al. Investigating awareness of artificial intelligence in healthcare among medical students and professionals in Pakistan: A cross-sectional study. *Ann. Med. Surg.* **86**, (2024).
12. Watkins, R. Guidance for researchers and peer-reviewers on the ethical use of large language models (LLMs) in scientific research workflows. *AI Ethics* (2023).
13. Li, H. et al. Ethics of large language models in medicine and medical research. *Lancet Digit. Health* **5**, (2023).
14. Liang, W. et al. Mapping the Increasing Use of LLMs in Scientific Papers. (2024). <https://arxiv.org/abs/2404.01268v1>.
15. Meng, X. et al. The application of large language models in medicine: A scoping review. *iScience* **27**, (2024).
16. Eppler, M. et al. Awareness and use of ChatGPT and large Language models: A prospective cross-sectional global survey in Urology. *Eur. Urol.* **85**, (2024).
17. Cano, P. M. A. et al. Natural language processing of grammar checker tools for academic writing: A systematic literature review. *J. Electr. Syst.* **20**, (2024).
18. Wu, L. & Pan, M. English grammar detection based on LSTM-CRF machine learning model. *Comput. Intell. Neurosci.* (2021).
19. Christou, P. A. Critical perspective over whether and how to acknowledge the use of Artificial Intelligence (AI) in qualitative studies. *Qual. Rep.* **28**, (2023).
20. Bender, E. M. & Koller, A. Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (2020).
21. Barocas, S. & Selbst, A. D. Big Data's disparate impact. *Calif. L. Rev.* **104**, (2016).
22. Tools such. As ChatGPT threaten transparent science; here are our ground rules for their use. *Nature* **613**, (2023).
23. Thorp, H. H. ChatGPT is fun, but not an author. *Science* **379**, (2023).
24. Flanagan, A., Bibbins-Domingo, K., Berkwitz, M. & Christiansen, S. L. Nonhuman Authors and Implications for the Integrity of Scientific Publication and Medical Knowledge. *JAMA* **329**, (2023).
25. Wagner, G., Lukyanenko, R. & Paré, G. Artificial intelligence and the conduct of literature reviews. *J. Inf. Technol.* **37**, (2022).
26. Patil, S. & Tovani-Palone, M. R. The rise of intelligent research: How should artificial intelligence be assisting researchers in conducting medical literature searches? *Clinics* **78**, (2023).
27. Dave, T., Athaluri, S. A. & Singh, S. ChatGPT in medicine: An overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front. Artif. Intell.* **6**, (2023).
28. Biswas, S. ChatGPT and the future of Medical writing. *Radiology* **307**, (2023).
29. Preiksaitis, C. & Rose, C. Opportunities challenges, and future directions of generative artificial intelligence in medical education: Scoping review. *JMIR Med. Educ.* **20**, (2023).
30. Alkaissi, H. & McFarlane, S. I. Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus* **19**, (2023).

Acknowledgements

We would like to sincerely acknowledge Dr. Le Huu Nhat Minh for his invaluable contribution in collecting 27 responses for this study. Additionally, we express our heartfelt gratitude to all research scholars of the GCSRT program at Harvard for participating in our survey and helping us complete this study.

Author contributions

T.M. conceptualized the study, defined the context and purpose of the study and drafted the survey questionnaire; B.Z., T.M., A.R., G.U., R.R. and N.P. collected data for the study. E.S., G.U. and T.M. performed data cleaning, data analysis and drafted tables. T.M., B.Z., E.S., R.R., N.P., A.A. A.R. and A.O. provided expertise for the study, conducted literature review and drafted the first draft of the manuscript. All authors were involved in the critical review of the manuscript. B.Z. and T.M. provided supervision, organization of the manuscript and project administration. All authors have made significant contributions and reviewed the final version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-81370-6>

[0.1038/s41598-024-81370-6](https://doi.org/10.1038/s41598-024-81370-6).

Correspondence and requests for materials should be addressed to B.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024