Tracing the Footprints of AI in Radiology Literature: A Detailed Analysis of Journal Abstracts

Die Spuren der KI in der radiologischen Literatur verfolgen: eine detaillierte Analyse von Journal-Abstracts

Authors
Ismail Mese

Affiliations
1 Department of Radiology, Istanbul Erenkoy Mental and Nervous Diseases Training and Research Hospital, Istanbul, Turkey

Keywords
abstract, artificial intelligence, quantitative analysis, radiology, research

received 4.10.2023
accepted 24.11.2023
published online 16.1.2024

Bibliography
Fortschr Röntgenstr 2024; 196: 843–849
DOI 10.1055/a-2224-9230
ISSN 1438-9029

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Correspondence
Dr. Ismail Mese
Department of Radiology, Istanbul Erenkoy Mental and Nervous Diseases Training and Research Hospital, Kadikoy, 34736 Istanbul, Turkey
Tel.: +90 50 68 95 16 53
ismail_mese@yahoo.com

ABSTRACT

Purpose To assess and compare the probabilities of AI-generated content within scientific abstracts from selected Q1 journals in the fields of radiology, nuclear medicine, and imaging, published between May and August 2022 and May and August 2023.

Materials and Methods An extensive list of Q1 journals was acquired from Scopus in the fields of radiology, nuclear medicine, and imaging. All articles in these journals were acquired from the Medline databases, focusing on articles published between May and August in 2022 and 2023. The study specifically compared abstracts for limitations of the AI detection tool in terms of word constraints. Extracted abstracts from the two different periods were categorized into two groups, and each abstract was analyzed using the AI detection tool, a system capable of distinguishing between human and AI-generated content with a validated accuracy of 97.06 %. This tool assessed the probability of each abstract being AI-generated, enabling an in-depth comparison between the two groups in terms of the prevalence of AI-generated content probability.

Results Group 1 and Group 2 exhibit significant variations in the characteristics of AI-generated content probability. Group 1, consisting of 4,727 abstracts, has a median AI-generated content probability of 3.8 % (IQR1.9–9.9 %) and peaks at 49.9 %, with the computation times contained within a range of 2 to 10 seconds (IQR 3–8 s). In contrast, Group 2, which is composed of 3,917 abstracts, displays a significantly higher median AI-generated content probability at 5.7 % (IQR2.8–12.9 %) surging to a maximum of 69.9 %, with computation times spanning from 2 to 14 seconds (IQR 4–11 s). This comparison yields a statistically significant difference in median AI-generated content probability between the two groups ($p = 0.005$). No significant correlation was observed between word count and AI probability, as well as between article type, primarily original articles and reviews, and AI probability, indicating that AI probability is independent of these factors.

Conclusion The comprehensive analysis reveals significant differences and variations in AI-generated content probabilities between 2022 and 2023, indicating a growing presence of AI-generated content. However, it also illustrates that abstract length or article type does not impact the likelihood of content being AI-generated.

Key Points

- The study examines AI-generated content probability in scientific abstracts from Q1 journals between 2022 to 2023.
- The AI detector tool indicates an increased median AI content probability from 3.8 % to 5.7 %.
- No correlation was found between abstract length or article type and AI probability.

Citation Format

- Mese I. Tracing the Footprints of AI in Radiology Literature: A Detailed Analysis of Journal Abstracts. Fortschr Röntgenstr 2024; 196: 843–849

ZUSAMMENFASSUNG

Zweck Ziel ist es, die Wahrscheinlichkeiten von KI-generierten Inhalten in wissenschaftlichen Abstracts aus ausgewähl-

ten Q1-Zeitschriften in den Bereichen Radiologie, Nuklearmedizin und Bildgebung zu bewerten und zu vergleichen, die zwischen Mai und August 2022 sowie Mai und August 2023 veröffentlicht wurden.

Materialien und Methoden Eine umfangreiche Liste von Q1-Zeitschriften wurde von Scopus in den Bereichen Radiologie, Nuklearmedizin und Bildgebung erworben. Alle Artikel aus diesen Zeitschriften wurden aus den Medline-Datenbanken bezogen, wobei der Fokus auf Artikeln lag, die zwischen Mai und August 2022 und 2023 veröffentlicht wurden. Die Studie verglich speziell Abstracts hinsichtlich der Grenzen des verwendeten KI-Erkennungstools in Bezug auf Wortbeschränkungen. Die aus den beiden Zeiträumen extrahierten Abstracts wurden in zwei Gruppen eingeteilt, und jedes Abstract wurde mit dem KI-Erkennungstool analysiert, einem System, das zwischen von Menschen und KI generierten Inhalten mit einer validierten Genauigkeit von 97,06 % unterscheiden kann. Dieses Tool bewertete die Wahrscheinlichkeit, dass jedes Abstract KI-generiert war, und ermöglichte so einen detaillierten Vergleich zwischen den beiden Gruppen in Bezug auf die Verbreitung der Wahrscheinlichkeit von KI-generierten Inhalten.

Ergebnisse Gruppe 1 und Gruppe 2 zeigen signifikante Unterschiede in den Merkmalen der Wahrscheinlichkeit von KI-generierten Inhalten. Gruppe 1, bestehend aus 4.727 Abstracts, hat eine mediane KI-generierte Inhalts-Wahrscheinlichkeit von 3,8 % (IQR 1,9 %–9,9 %) und erreicht Spitzenwerte von 49,9 %, mit Berechnungszeiten von 2 bis 10 Sekunden (IQR 3s–8 s). Im Gegensatz dazu zeigt Gruppe 2, die aus 3.917 Abstracts besteht, eine deutlich höhere mediane KI-generierte Inhalts-Wahrscheinlichkeit von 5,7 % (IQR 2,8 %–12,9 %) und erreicht ein Maximum von 69,9 %, mit Berechnungszeiten von 2 bis 14 Sekunden (IQR 4s–11 s). Dieser Vergleich ergibt einen statistisch signifikanten Unterschied in der medianen Wahrscheinlichkeit von KI-generierten Inhalten zwischen den beiden Gruppen ($p = 0{,}005$). Es wurde keine signifikante Korrelation zwischen der Wortanzahl und der KI-Wahrscheinlichkeit beobachtet, ebenso wie zwischen dem Artikeltyp, hauptsächlich Originalartikel und Übersichten, und der KI-Wahrscheinlichkeit, was darauf hinweist, dass die KI-Wahrscheinlichkeit unabhängig von diesen Faktoren ist.

Schlussfolgerung Die umfassende Analyse zeigt signifikante Unterschiede und Variationen in den Wahrscheinlichkeiten von KI-generierten Inhalten zwischen 2022 und 2023, was auf eine zunehmende Präsenz von KI-generierten Inhalten hinweist. Es zeigt jedoch auch, dass weder die Länge des Abstracts noch der Artikeltyp die Wahrscheinlichkeit beeinflussen, dass der Inhalt KI-generiert ist.

Kernaussagen

- Die Studie untersucht die Wahrscheinlichkeit von KI-generierten Inhalten in wissenschaftlichen Abstracts aus Q1-Zeitschriften zwischen 2022 und 2023.
- Das KI-Erkennungstool zeigt eine erhöhte mediane Wahrscheinlichkeit für KI-Inhalte von 3,8 % auf 5,7 %.
- Es wurde keine Korrelation zwischen Abstract-Länge oder Artikeltypen und KI-Wahrscheinlichkeit gefunden.

Introduction

Artificial intelligence (AI), a distinguished subfield of computer science, is dedicated to creating intelligent machines capable of performing tasks that typically require human intelligence [1]. AI utilizes its capability to learn and make decisions based on the environment and acquired information, taking forms such as machine learning, which learns from data to make predictions, and Natural Language Processing (NLP), which employs algorithms to understand and simulate human-like interactions [1, 2]. In recent years, AI has diversified into fields like healthcare and the Internet of Things, collectively known as the Artificial Intelligence of Things [3–5].

OpenAI introduced ChatGPT, a state-of-the-art natural language processing (NLP) system, in November 2022 with the launch of its version 3.5 designed to generate human-like dialogue [6]. It understands the context of conversations and produces appropriate responses, showcasing versatility in understanding different conversational contexts and generating responses in varied styles, thus representing a notable advancement in AI applications [6]. However, the integration of AI systems like ChatGPT, especially in medical research, has sparked debates and raised concerns, mainly related to privacy, security, misuse, and over-reliance [6]. If not secured, the extensive data AI systems access, like medical records, risk unauthorized access and misuse [6]. Additionally, there is growing concern regarding the potential over-dependence and unwarranted trust placed in AI systems by medical professionals, often without fully comprehending the limitations and the possibilities of inaccuracies inherent in such systems [6]. Given these issues, examining the role and impact of AI systems like ChatGPT in medical research is crucial. Our hypothesis suggests there has been an increase in the integration of AI in medical research. This increase should be objectively demonstrated. Therefore, this study aims to assess and compare the probabilities of AI-generated content within scientific abstracts from selected Q1 journals in the fields of radiology, nuclear medicine, and imaging, published between May and August 2022 and May and August 2023. This study employs an advanced plagiarism detection tool to distinguish between human and AI-generated content and conducts a detailed statistical analysis to elucidate any significant disparities in the probabilities of AI-generated content between the two periods.

Materials and Methods

Our study, which focuses on the examination of scientific publications, necessitates neither direct involvement with human subjects nor ethical committee approval, in compliance with the prin-

ciples of the Declaration of Helsinki. We took rigorous measures to ensure the privacy and confidentiality of extracted data.

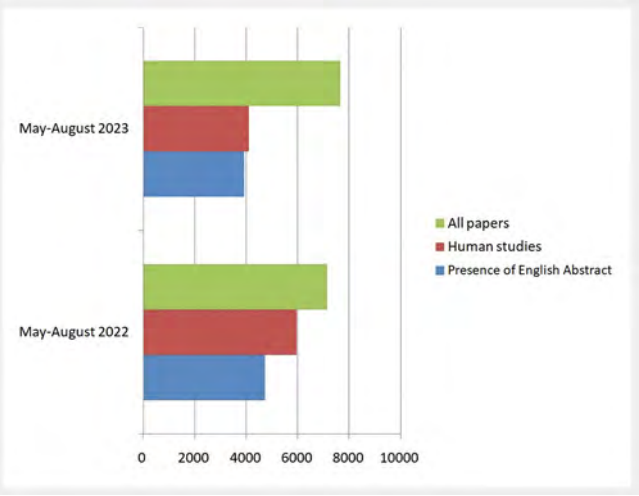
Selection criteria

Initially, an exhaustive list of Q1 journals within the fields of radiology, nuclear medicine, and imaging was obtained by consulting the Scopus database [7]. Subsequently, the Medline database was employed to acquire a list of all articles related to that journal using journal filters [8]. This precise application of filters facilitated the acquisition of all articles published between May and August 2022, and May and August 2023, within the identified journals. Time frames were selected to consistently compare AI-generated content probabilities across consecutive years. These specific intervals ensure comparability by controlling for potential seasonal variations in publication volume and content generation, thereby providing more reliable insight into trends and patterns. Every article identified within these periods underwent a rigorous assessment to determine the presence of an abstract. Articles without abstracts were excluded as the study focuses on comparing abstracts. Articles involving non-human subjects were deliberately excluded from our study to maintain focus and for simplicity. Including non-human subjects could introduce additional variability and nuances, such as those pertinent to veterinary or botanical studies, which could skew the results and obscure the specific insights we seek regarding AI's role in generating content in human medical research. Articles were also excluded if their abstracts were less than 30 words or exceeded 1000 words, due to the limitations of the AI detector tool utilized in this study. ▶ **Fig. 1** shows the quantity of eligible papers for each time frame, showcasing the number of publications that met the defined selection criteria.

Categorization and Analysis with AI Detector Tool

Titles, word counts, article types, and abstract texts were recorded for abstracts meeting the criteria. Subsequently, abstracts from May to August 2022 were categorized as Group 1, and those from May to August 2023 as Group 2.

The preserved text of each abstract from each group was subjected to analysis using the CopyLeaks detection tool, a system renowned for its precision in distinguishing between human and AI-generated content. The preserved text of each abstract was analyzed using the CopyLeaks detection tool. The tool's developer claims an accuracy rate of 99.1 % [9], but an independent study reported a slightly lower accuracy rate of 97.06 % [10]. CopyLeaks identifies AI-generated content by analyzing distinctive writing patterns, word choices, and syntax. In this regard, the abstracts, along with their full titles, were uploaded to the CopyLeaks website [9]. The website then provided a report section detailing the probability of each abstract being AI-generated after conducting an analysis (▶ **Fig. 2**). Identified AI-generated content was quantified in percentages and recorded systematically to facilitate a more nuanced understanding. We also recorded analysis computation times. The analysis focused on comparing the means or medians of AI-generated content probabilities between the groups, treating AI-generated probabilities as continuous variables due to the absence of a definitive threshold to conclusively determine AI authorship. This methodological approach enabled detailed exploration and evaluation of the emergence and prevalence of AI-generated content probability within time frames. All assessments of manuscripts and data extractions were performed solely by the author of this study.



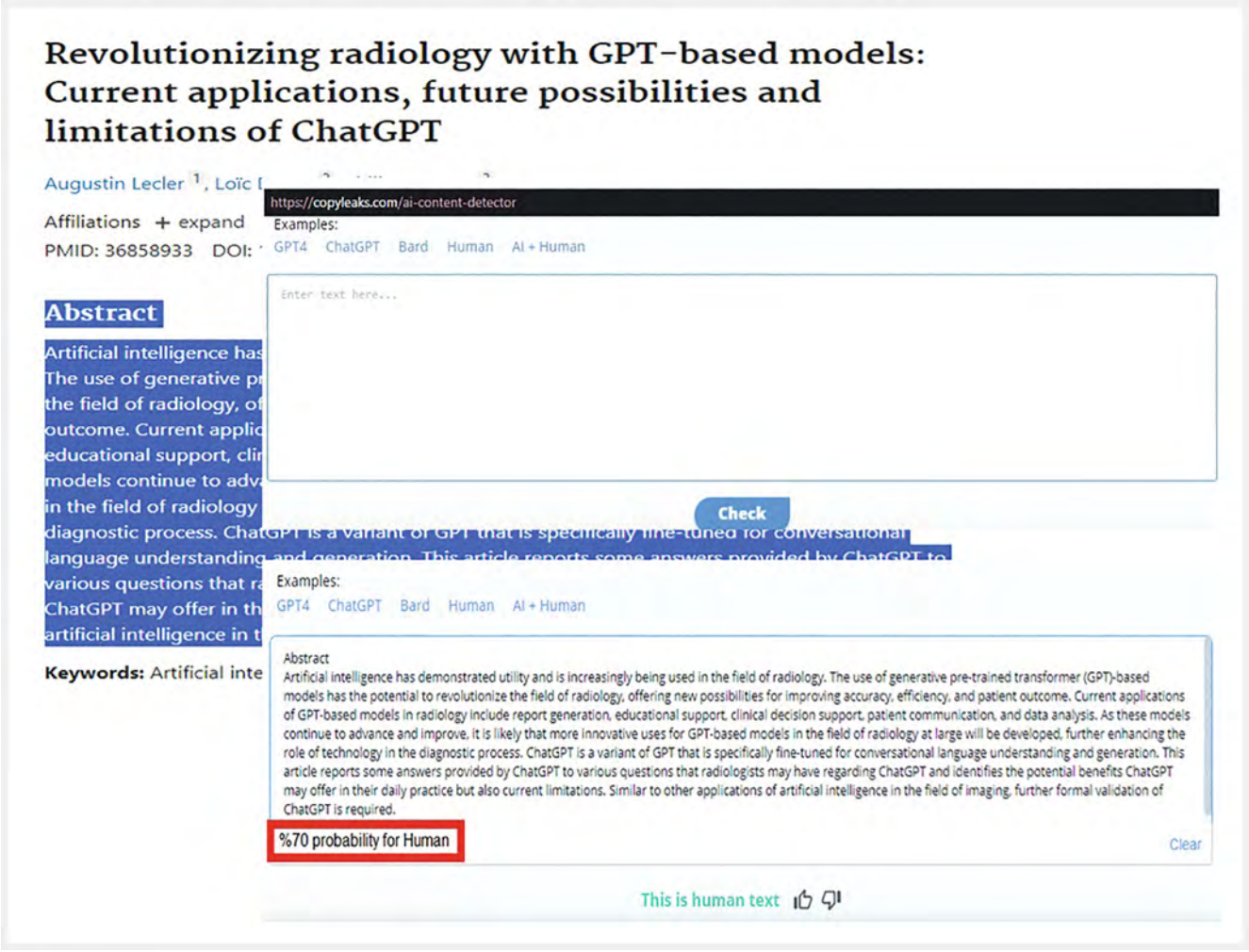
▶ **Fig. 1** Number of eligible papers for each time frame based on selection criteria.

Statistical Methods

The distribution of AI-generated probabilities will be determined using the Shapiro-Wilk test for normality. Non-normally distributed data will be compared using the Mann-Whitney U-Test for continuous or ordinal variables and the chi-square test for categorical variables, assessing the discrepancies between observed and expected frequencies. Conversely, normally distributed data will undergo appropriate parametric tests. Descriptive statistics will be applied to characterize our data more profoundly. These will encompass measures of central tendency, such as mean or median, and measures of variability, like standard deviation or interquartile range, offering an in-depth insight into data distribution. A p-value less than 0.05 will be considered indicative of statistical significance in all tests conducted. Statistical analysis was performed using IBM SPSS version 21.

Results

Group 1, encompassing 4727 abstracts, disclosed a median AI-generated content probability of 3.8 % (IQR: 1.9–9.9 %), with a peak value of 49.9 %. The computation times for this group ranged from 2 to 10 seconds, with an IQR of 3 to 8 seconds, illustrating data variability around the median. In contrast, Group 2, consisting of 3917 abstracts, had a median AI-generated content probability of 5.7 % (IQR: 2.8–12.9 %), with probabilities escalating to a maximum of 69.9 %. The computation times for this group were more varied, ranging from 2 to 14 seconds, with an IQR of 4 to 11 seconds. It is noteworthy that our dataset was complete, as there were no missing values for any of the variables of interest in either group. In terms of data distribution, the Shapiro-Wilk test was employed ($p = < 0.001$). Based on these find-



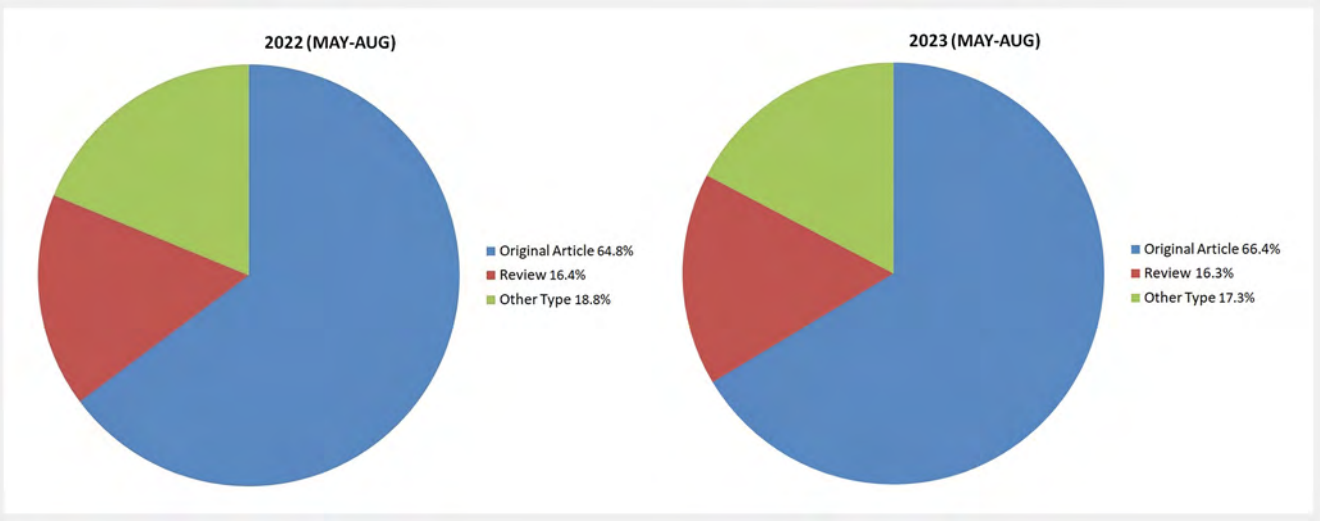
▶ **Fig. 2** Abstract extraction and AI content analysis using Copyleaks: We extracted all relevant abstracts that met our criteria, categorizing them into two groups based on the year, 2022 or 2023. Each abstract was then copied and pasted into the tool section on the Copyleaks website to calculate the probability of being human-generated or AI-generated. In this instance, the tool calculated a 30 % probability of the content being AI-generated.

ings, and given the non-normal nature of our data, we opted for the Mann-Whitney U-Test for comparative analysis between the two groups (▶ **Table 1**).

We conducted a comprehensive analysis of the word counts in abstracts from two distinct groups. For Group 1, the abstracts exhibited word counts that spanned between 50 and 488 words, with the median word count being 252 words, while the abstracts in Group 2 displayed a spectrum of word counts ranging from 42 to 459 words, with a median of 249 words. To assess the differences in word counts between the two groups, a Mann-Whitney U-test was conducted, yielding a p-value of 0.453, indicating no significant statistical difference between the two groups. Subsequently, we explored the potential correlation between word count and AI probability within these abstracts using Spearman's rank correlation coefficient due to the non-normal distribution of our data. The Spearman correlation coefficient (ρ) was calculated to be 0.12 with a p-value of 0.31, suggesting no significant correlation between word count and AI probability in our studied con-

▶ **Table 1** Mann-Whitney U-Test Results for Comparative Analysis Between Group 1 and Group 2.

Attribute	Group 1	Group 2	p-value
Number of abstracts	4727	3917	
Computation time range (25th–75th IQR)	2 s to 10 s (3–8 s)	2 s to 14 s (4–11 s)	
Peak AI-generated content probability	49.9 %	69.9 %	
Median AI-generated content probability (25th–75th IQR)	3.8 % (1.9–9.9 %)	5.7 % (2.8–12.9 %)	**0.005**



▶ **Fig. 3** Distribution of surveyed article types.

text. These findings suggest that the length of the abstract does not impact the likelihood of AI probability.

Further breaking down Group 1 in terms of article type, there were 3,065 original articles, making up roughly 64.8 % of the articles. 772 articles were reviews, comprising about 16.4 % of the total. The remaining 890 articles, amounting to 18.8 %, fell into different categories. Similarly, in Group 2, there were 2,599 original articles, which constituted approximately 66.4 % of the articles. Reviews were the next most common with 642 articles, accounting for about 16.3 % of the total. The remaining 676 articles, or 17.3 % (676 out of 3,917), were of various other types. We assessed the correlation between AI probability and article type within both groups using Chi-square tests of independence, to understand if there were any dependencies between them. We conclude that there is no significant correlation between AI probability and article type for both Group 1 and Group 2, indicating that AI probability is independent of article type in our study (p-value 0.72, and 0.75 respectively). ▶ **Fig. 3** displays the distribution of articles.

Discussion

The integration of AI into medical research has emerged as a transformative force [11]. Our meticulous examination of scientific article abstracts from 2022 to 2023 illuminates this evolving trend, showcasing a significant increase in the probability of AI-generated content. The comparative analysis of abstracts from two distinct time periods reveals fluctuations in AI content probability and computation times. This suggests a potential expansion in the acceptance and application of AI technologies in scientific research. Our study focused on the fields of radiology, nuclear medicine, and imaging, which naturally align with AI due to their heavy reliance on technology. However, similar trends may be observed in other medical fields.

In our study, we exclusively utilized abstracts instead of full texts for analysis. The rationale behind this methodological choice stems from the essence of abstracts as compact summations of the key components of scholarly texts, encompassing objectives, methodology, results, and conclusions [12]. By analyzing abstracts, we were able to condense our investigation into the pivotal elements of the manuscripts, which is particularly imperative given the limitations of the AI detection tool utilized in terms of word constraints. This approach allowed for a more efficient and streamlined analysis, facilitating the handling of a vast amount of documents swiftly and effectively.

Utilizing AI for writing scientific material, like generating or refining content, and charting its probability is crucial for maintaining transparency, accuracy, and reliability in scholarly publications. When AI is used to aid in the creation of scientific content, disclosing the probability scores or confidence intervals assigned by the AI models ensures that readers, peer reviewers, and other researchers can critically assess the validity and reliability of the presented information. It provides insight into the likelihood of the AI-generated content being accurate and allows researchers to weigh the information appropriately. This transparency is vital to uphold the rigor and trustworthiness of scientific discourse, promoting an environment where knowledge is not only generated and disseminated efficiently but also scrutinized and validated rigorously, ensuring that the advancement of science is predicated on robust and reliable foundations.

It's crucial to highlight the lack of significant correlation between the length of abstracts and AI probability. This lack of correlation implies that the variations in abstract length do not impact the likelihood of AI-related content within the studied parameters, suggesting a non-dependency between content length and AI relevance. Additionally, the investigation into the types of articles within both groups did not reveal any meaningful association between article type and AI probability. This indicates that the variety of articles, whether they are original articles, reviews, or other forms, does not influence the probability of the content being related to AI.

Recognizing the distinct advantages and capabilities that AI introduces to medical research elucidates why its incorporation in

scientific abstracts is intensifying [13]. The emerging trend may be credited to the unparalleled efficacy and acceleration that AI introduces. Generative AI models, due to their ability to process extensive data swiftly, not only expedite the research procedure but also aid in uncovering patterns and insights that could be overwhelming or intricate for the human intellect to comprehend rapidly [13]. These models possess the ability to transform vast and complex datasets into concise and lucid summaries, potentially making the research more comprehensible and reachable for a diversified audience. Babl et al. investigated the capability of the openly accessible version of ChatGPT to formulate a quality conference abstract. The study used a hypothetical but accurately derived data table, resulting in an abstract that was coherent, error-free, and complied with the established guidelines [14].

Generative AI provides invaluable tools for academic writing, aiding in literature reviews, suggesting relevant topics, streamlining citations, and enhancing the structuring and clarity of manuscripts [13]. ChatGPT, in particular, is adept at organizing references and citations, thereby facilitating the academic writing process [15]. However, studies have raised concerns over its accuracy. Ariyaratne et al. found significant inaccuracies in articles generated by ChatGPT in the field of radiology [16]. Similarly, a study by Wu et al. revealed that only 10 % of references provided by ChatGPT were entirely accurate in the field of head and neck surgery [17]. Alkaissi et al. discussed the implications of using ChatGPT for scientific writing in medicine, highlighting that while the model can generate coherent and scholarly text, it can also produce inaccurate, unverified, or incorrect information, sometimes referring to non-existent or irrelevant academic citations [15]. These inaccuracies or "artificial hallucinations" pose serious ethical and practical challenges in fields requiring stringent factual accuracy [18, 19]. The authors of the study recommended updated editorial policies, such as AI output detectors and full disclosure practices, to uphold the integrity of academic writing [18].

The rising prevalence of AI in academic research is raising significant questions and concerns, primarily related to research integrity and originality. Large language models (LLMs) like ChatGPT might produce content that is too similar or identical to existing works, potentially causing issues related to plagiarism and copyright infringement. There are also inquiries about authorship recognition, especially in cases where substantial parts of academic content are AI-generated, leading to debates on the rightful attribution of authorship [5, 6, 20]. A study by Ali et al. highlights the issues surrounding authorship ethics, suggesting that AI tools like ChatGPT should not be recognized as authors as they do not fulfill standard authorship guidelines, which require entities to agree to be listed and take responsibility for their contributions [21]. These systems also can't manage copyright and license agreements, and thus should not be granted authorship status.

Recognizing the surge in both the quantity and sophistication of AI-driven content, the adoption of advanced plagiarism detection tools such as CopyLeaks has become indispensable. CopyLeaks is recognized for its exceptional detection capabilities but is also lauded for its user-friendly interface, enabling users with varying degrees of technical expertise to navigate through its features with ease [9, 10]. Its reputation for reliability and accuracy stems from its proven track record, having been identified as one of the best in a recent article for its AI detection rate [10]. Another notable feature of CopyLeaks is its multilingual detection capability, allowing users to scan content in various languages, thus expanding its utility to a wider, more diverse user base.

It's imperative to understand that the AI detector tool employed in this research was primarily designed to identify and analyze the presence of AI-generated content within scientific abstracts, focusing on distinctive writing patterns, word choices, and syntax, and not to verify the factual accuracy or assess the need for language refinement in the content analyzed. Therefore, our investigation did not explicitly reveal instances of fabrication of facts or highlight specific use cases for language correction. This limitation underscores the necessity for comprehensive and diversified research methodologies to fully explore and understand the varied dimensions of AI's influence and interaction within the medical research field, including assessing the credibility and the intricate language modifications made by AI tools, which were beyond the scope of our current study's capability to detect.

To manage the complexity and volume of data, our study was primarily focused on Q1 journals, which streamlined the research process but also somewhat limited the scope and applicability of our findings. Incorporating journals from Q1 to Q4 would have allowed for a more encompassing and varied understanding of the prevalence of AI-generated content in scientific publications. However, due to constraints in data handling and analysis, a decision was made to concentrate on top-tier journals. Additionally, the accuracy of the CopyLeaks AI detector, while reported to be high, does have a margin of error, potentially leading to misclassification of AI-generated content.

Another significant limitation of our study is the exclusive analysis of abstracts without delving into the full texts of the articles. This approach, while efficient for handling a large volume of documents, inherently restricts our insight into the depth and complexity of the content. This limitation is particularly relevant in the context of AI-generated content detection, where subtle nuances and complex argumentation in full texts could offer a more comprehensive understanding of AI's influence on scientific writing. By focusing solely on abstracts, we may overlook key aspects of AI integration and its implications on the quality and integrity of scientific discourse. Therefore, this methodological constraint should be acknowledged and considered when interpreting the findings of our study.

The study's temporal scope, particularly in relation to the evolving landscape of AI-generated content in scientific literature, is limited. While our analysis focuses on abstracts from May to August in 2022 and 2023, it is crucial to acknowledge the potential presence of AI-generated content prior to this period. The launch of ChatGPT version 3.5 by OpenAI in November 2022 marked a significant milestone in the development of Large Language Models (LLMs), attracting substantial attention [6]. However, earlier versions of GPT, notably GPT-2 released in 2019, were already accessible to the public [22]. This availability suggests that the use of AI-generated content in scientific literature, including in our fields of interest may have commenced before our study period.

The analysis would have greatly benefited from a trend analysis over a more extended period, ideally covering at least three years. Such an analysis could provide deeper insight into the progression and adoption rate of AI-generated content in scientific literature. It would enable us to trace the evolution of AI's role in content creation more accurately, thereby offering a more comprehensive understanding of its impact on scientific discourse.

This study elucidates a discernible increase in AI-generated content in medical research abstracts between 2022 and 2023, emphasizing a growing reliance on and integration of AI in scientific documentation and exploration. While AI offers unparalleled efficiency and insight, it also raises substantial concerns about accuracy, integrity, and ethical conduct within scientific discourse. Therefore, a balanced and conscientious approach is imperative to leverage AI's benefits while mitigating its potential risks and maintaining the rigor and authenticity of scientific endeavors.

Statement: It is important to clarify that the use of LLMs in our manuscript was confined strictly to enhancing grammar, punctuation, and similar language-related aspects. We utilized these technologies solely to improve the readability and language quality of our document, while ensuring that the core content and research findings remained the product of human expertise and intellectual rigor. The author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

CLINICAL RELEVANCE OF THE STUDY

- This study highlights a significant increase in the probability of AI-generated content within medical research abstracts between 2022 and 2023, reflecting the amplifying role and integration of AI tools, such as ChatGPT, in scholarly medical publications.
- The findings underscore the necessity for transparency, reliability, and ethical considerations when utilizing AI in scientific writings, particularly given the potential inaccuracies and "artificial hallucinations" produced by such tools.
- The lack of correlation between AI-generated content probability and abstract length or article type suggests the application of AI is widespread across different forms and lengths of medical articles, reinforcing the need for thorough scrutiny and validation irrespective of article characteristics.
- The prevalent use of advanced plagiarism tools like CopyLeaks, notable for its precision and user-friendly interface, emphasizes the crucial role such tools play in maintaining the integrity and originality of scientific discourse in the face of increasing AI integration.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Deng J, Lin Y. The benefits and challenges of ChatGPT: an overview. Front Comput Intelligent Syst 2022; 2 (2): 81–83

[2] Lecun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521: 436–444

[3] Kamdar J, Jeba Praba J, Georrge JJ. Artificial intelligence in medical diagnosis: methods, algorithms and applications. In: Machine Learning with Health Care Perspective. Springer; 2020: 27–37

[4] Rose K, Eldridge S, Chapin L. The internet of things: an overview. Internet Soc 2015; 80: 1–50

[5] Yu K, Guo Z, Shen Y et al. Secure artificial intelligence of things for implicit group recommendations. IEEE Internet Things J 2021; 9 (4): 2698–2707

[6] King MR. The future of ai in medicine: a perspective from a chatbot. Ann Biomed Eng 2022; 51 (2): 291–295

[7] Scopus. Content Sources. [accessed on September 2, 2023]. Available at: https://www.scopus.com/sources.uri

[8] MEDLINE. Database Homepage. [accessed on September 2, 2023]. Available at: https://pubmed.ncbi.nlm.nih.gov/

[9] CopyLeaks. CopyLeaks AI Content Detector beta. [accessed on Aug 23, 2023]. Available at: 2023. https://copyleaks.com/ai-content-detector

[10] Orenstrakh MS, Karnalim O, Suárez CA et al. Detecting LLM-Generated Text in Computing Education: A Comparative Study for ChatGPT Cases. 2023. doi:10.1145/nnnnnnn.nnnnnnn

[11] Varghese J, Chapiro J. ChatGPT: The transformative influence of generative AI on science and healthcare. J Hepatol 2023. doi:10.1016/j.jhep.2023.07.028

[12] Andrade C. How to write a good abstract for a scientific paper or conference presentation. Indian J Psychiatry 2011; 53 (2): 172–175

[13] Huang J, Tan M. The role of ChatGPT in scientific communication: writing better scientific review articles. Am J Cancer Res 2023; 13: 1148–1154

[14] Babl FE, Babl MP. Generative artificial intelligence: Can ChatGPT write a quality abstract? Emerg Med Australas 2023. doi:10.1111/1742-6723.14233

[15] Alkaissi H, McFarlane SI. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. Cureus 2023; 15: e35179

[16] Ariyaratne S, Iyengar KP, Nischal N et al. A comparison of ChatGPT-generated articles with human-written articles. Skeletal Radiol 2023; 52: 1755–1758

[17] Wu RT, Dang RR. ChatGPT in head and neck scientific writing: A precautionary anecdote. Am J Otolaryngol 2023; 44: 103980

[18] Ji Z, Lee N, Frieske R et al. Survey of hallucination in natural language generation. ACM Comput Surv 2022; 1: 1

[19] Else H. Abstracts written by ChatGPT fool scientists. Nature 2023; 613: 423

[20] Zielinski C, Winker MA, Aggarwal R et al. Chatbots, Generative AI, and scholarly manuscripts: WAME recommendations on Chatbots and Generative artificial intelligence in relation to scholarly publications. [accessed on August 23, 2023]. https://wame.org/page3.php?id=106

[21] Ali MJ, Djalilian A. Readership awareness series – Paper 4: Chatbots and ChatGPT – Ethical considerations in scientific publications. Ocul Surf 2023; 28: 153–154

[22] Nazir A, Wang Z. A comprehensive survey of ChatGPT: Advancements, applications, prospects, and challenges. Meta Radiol 2023; 1: 100022