



# Meta-research on reporting guidelines for artificial intelligence: are authors and reviewers encouraged enough in radiology, nuclear medicine, and medical imaging journals?

Burak Koçak

Ali Keleş

Fadime Köse

University of Health Sciences, Başakşehir Çam and Sakura City Hospital, Clinic of Radiology, Istanbul, Turkey

## PURPOSE

To determine how radiology, nuclear medicine, and medical imaging journals encourage and mandate the use of reporting guidelines for artificial intelligence (AI) in their author and reviewer instructions.

## METHODS

The primary source of journal information and associated citation data used was the Journal Citation Reports (June 2023 release for 2022 citation data; Clarivate Analytics, UK). The first- and second-quartile journals indexed in the Science Citation Index Expanded and the Emerging Sources Citation Index were included. The author and reviewer instructions were evaluated by two independent readers, followed by an additional reader for consensus, with the assistance of automatic annotation. Encouragement and submission requirements were systematically analyzed. The reporting guidelines were grouped as AI-specific, related to modeling, and unrelated to modeling.

## RESULTS

Out of 102 journals, 98 were included in this study, and all of them had author instructions. Only five journals (5%) encouraged the authors to follow AI-specific reporting guidelines. Among these, three required a filled-out checklist. Reviewer instructions were found in 16 journals (16%), among which one journal (6%) encouraged the reviewers to follow AI-specific reporting guidelines without submission requirements. The proportions of author and reviewer encouragement for AI-specific reporting guidelines were statistically significantly lower compared with those for other types of guidelines ( $P < 0.05$  for all).

## CONCLUSION

The findings indicate that AI-specific guidelines are not commonly encouraged and mandated (i.e., requiring a filled-out checklist) by these journals, compared with guidelines related to modeling and unrelated to modeling, leaving vast space for improvement. This meta-research study hopes to contribute to the awareness of the imaging community for AI reporting guidelines and ignite large-scale group efforts by all stakeholders, making AI research less wasteful.

## CLINICAL SIGNIFICANCE

This meta-research highlights the need for improved encouragement of AI-specific guidelines in radiology, nuclear medicine, and medical imaging journals. This can potentially foster greater awareness among the AI community and motivate various stakeholders to collaborate to promote more efficient and responsible AI research reporting practices.

## KEYWORDS

Artificial intelligence, machine learning, guideline, checklist, reporting

Corresponding author: Burak Koçak

E-mail: drburakkocak@gmail.com

Received 10 November 2023; revision requested 03 January 2024; accepted 10 February 2024.



Epub: 20.02.2024

Publication date: xx.xx.2024

DOI: 10.4274/dir.2024.232604

Poor or suboptimal reporting of medical research is regarded as a significant and widespread issue that contributes to the waste of scarce and valuable resources invested in research projects.<sup>1-5</sup> For such studies, readers cannot assess the validity of the method relative to existing knowledge, and thus the reliability and reproducibility of the findings.<sup>6</sup> This hinders the clinical translation of promising research findings<sup>7</sup> and their comparability with other publications for evidence synthesis or meta-analysis.<sup>8</sup> The adherence to consensus-based reporting standards (i.e., reporting guidelines) is one of the principal methods for reducing the risk of poor reporting. To promote this, vast projects, like Enhancing the QUALity and Transparency Of health Research (EQUATOR) network, were started, and several reporting guidelines were developed and published in the literature.<sup>9,10</sup> Typically, these guidelines take the form of online or offline checklists, flowcharts, or explicit texts that instruct authors on how to report their research. Several studies researched the effectiveness of adhering to reporting guidelines in various study types. They found that adherence is associated with improved manuscript quality in peer review,<sup>11</sup> favorable reviewer ratings and editorial decisions,<sup>12</sup> higher citation counts and opportunity to be published in journals with a higher impact factor,<sup>13</sup> and improved completeness and quality of the research.<sup>14-22</sup>

Similar to healthcare literature, medical artificial intelligence (AI) research faces poor or suboptimal reporting issues. With the massive growth of healthcare literature using AI, including medical imaging,<sup>23</sup> the need for complete and structured reporting of prognostic and diagnostic studies that use machine learning algorithms or models has

increased. An expanding body of research indicates that AI studies frequently fall short of expected reporting standards,<sup>24,25</sup> lacking sufficient details on modeling and its evaluation, and failing to adequately address potential sources of bias.<sup>26-32</sup> Multiple specific guidelines relevant to AI studies have been developed to address these issues.<sup>25,33-41</sup> Examples of these guidelines include Checklist for AI in Medical Imaging (CLAIM),<sup>42,43</sup> Fairness, Universality, Traceability, Usability, Robustness, and Explainability-AI (FUTURE-AI),<sup>44</sup> Minimum Information about Clinical AI Modelling (MI-CLAIM),<sup>45</sup> CheckList for Evaluation of Radiomics research (CLEAR),<sup>46</sup> and Methodological Radiomics Score (METRICS).<sup>47</sup> In addition, as a continuation of previous efforts, several guidelines are currently under development, such as Standards for the Reporting of Diagnostic Accuracy Studies-AI (STARD-AI) for AI-centered diagnostic test accuracy studies and Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis-AI (TRIPOD-AI) for those related to diagnostic models.<sup>48,49</sup> The most widely recognized AI guidelines and ones currently under development can be found in the following seminal papers.<sup>50,51</sup>

The availability of reporting guidelines and checklists has not yet resolved the problem of inadequate reporting. While editorial guidance advocating for transparent reporting is widespread and well-intentioned, authors frequently overlook or fail to adhere to these guidelines.<sup>52-57</sup> In a very recent citation analysis of an AI checklist on medical imaging and a meta-research on radiomics, claims regarding the use of checklists and quality scoring tools for self-reporting (i.e., reporting with filling checklists by study authors) have been supported.<sup>26,32</sup> Journals can significantly impact the quality of reporting by encouraging or mandating responsible reporting practices, such as the use of reporting guidelines and checklists in their author and reviewer instructions.<sup>58,59</sup> However, research on the encouragement of AI reporting guidelines by journals specialized in radiology, nuclear medicine, and medical imaging is scarce.<sup>60</sup> Investigating this issue could yield valuable insights to foster higher-quality research within these journals.

This meta-research study aims to determine how these journals encourage and mandate (i.e., requiring a filled-out checklist) the use of AI reporting guidelines in their author and reviewer instructions by comparing reporting guidelines that are specific to AI, related to modeling, and unrelated to modeling.

## Methods

Figure 1 presents the key study steps of this meta-research.

### Dataset

The primary source of journals and associated citation data used was the Journal Citation Reports (June 2023 release for 2022 citation data; Clarivate Analytics, UK). This report was based on data obtained from the Web of Science (WoS) (Clarivate Analytics, UK).

Journals indexed in the WoS category, radiology, nuclear medicine, and medical imaging, that met the following criteria were included in this study: inclusion in the Science Citation Index Expanded (SCIE) or Emerging Sources Citation Index (ESCI) and placement within the first quartile (Q1; top 25% of journals in the list) or second quartile (Q2; journals in the top 25%–50% group) based on the 2022 Journal Impact Factor. This analysis excluded journals that had a limited scope, specifically those that focused solely on review articles (i.e., not publishing original research articles), as these journals were not expected to publish articles using AI reporting guidelines.

Two readers, each in their third year of radiology residency and with prior experience conducting systematic reviews on reporting quality in AI or radiomics, accessed the author and reviewer instructions from the journals' websites and saved them as PDF files. The task was distributed evenly among the readers, and they also reviewed each other's resulting files. All author and reviewer instructions were accessed between September 4 and 7, 2023. In the case of multiple instructions, the most up-to-date version was selected.

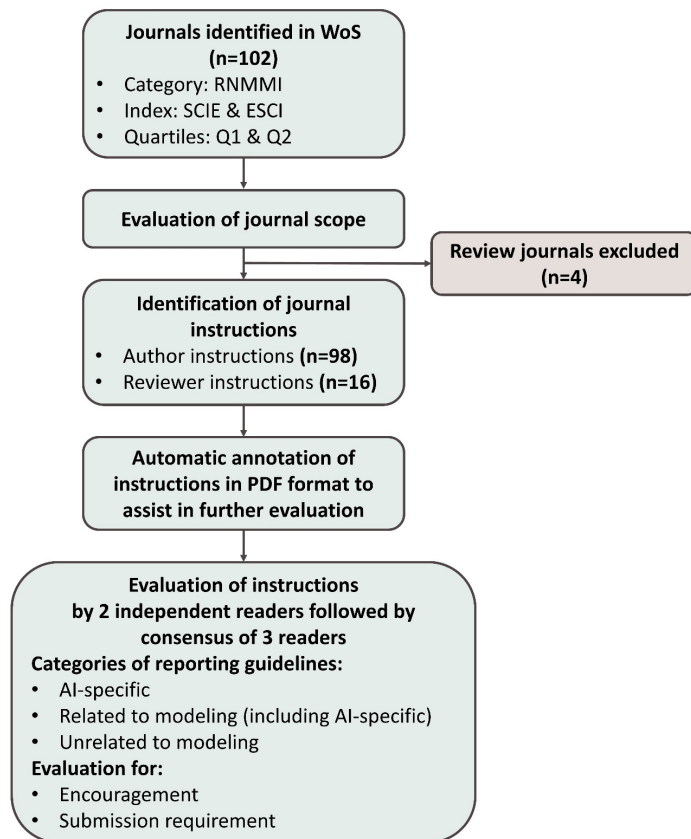
To mitigate errors during the assessment of instructions, a custom Python script based on the PyMuPDF package was used to automatically annotate certain terms within the PDF documents. The terms covered AI, machine learning, reporting, guidelines, checklists, and their specific names or acronyms. The code and exact terms can be accessed at <https://github.com/radiomic/PDFhighlighter>.

### Evaluation of author and reviewer instructions

The author and reviewer instructions that were automatically annotated by the script were evaluated by the same readers who downloaded the instructions. All evaluations underwent a meticulous review process overseen by an additional reader possessing

#### Main points

- Based on author and reviewer instructions, artificial intelligence (AI)-specific guidelines are not commonly encouraged, and they are not mandated for submission as filled-out checklists by radiology, nuclear medicine, and medical imaging journals.
- The proportions of author and reviewer encouragements for AI-specific reporting guidelines were statistically significantly lower compared with those for other types of guidelines.
- The collaboration of all stakeholders, including guideline developers, journal managers, editors, reviewers, authors, and funders, is needed to further encourage these guidelines to make AI research less wasteful.



**Figure 1.** Key study steps. WoS, Web of Science; RNMMI, Radiology, Nuclear Medicine, and Medical Imaging; SCIE, Science Citation Index Expanded; ESCI, Emerging Sources Citation Index; Q1, first quartile; Q2, second quartile; PDF, portable document format; AI, artificial intelligence.

8 years of expertise as a radiology specialist, complemented by over 5 years of research experience in machine learning, radiomics, and systematic reviews. Final decisions were reached by consensus among all readers.

The collected data primarily fell into two categories: encouragement of authors or reviewers and the presence of submission requirements for filled-out checklists in case of encouragement. When evaluating the encouragement, to elicit a positive evaluation from readers, it was imperative to explicitly state the name of the reporting guideline or make a direct reference to it. In addition, encouragement was defined as any sort of mention of specific guidelines. For instance, if authors and reviewers are recommended for adherence, referral, or usage of the guidelines, even if not explicitly intended for integration into their workflow, it was considered encouragement. The inclusion of general references to the central source or hub of guidelines or checklists, such as the EQUATOR network website, was not regarded as a specific encouragement in this work. To fulfill the submission requirement (i.e., mandating), this study sought a clear indication that the filled-out checklist would be

uploaded to the submission system as an integral part of both the manuscript and peer review processes. The submission systems were only investigated when the submission requirements were unclear in the instructions. Checklists without an associated publication in a journal (i.e., checklists as part of journal instructions without a digital object identifier) were not considered as a reporting guideline.

Three types of reporting guidelines were analyzed as follows: i) AI-specific reporting guidelines; ii) those related to modeling (e.g., diagnostic or prognostic modeling; may or may not be associated with AI or machine learning); and iii) those unrelated to modeling. AI-specific reporting guidelines and those related to modeling included those specified in two recent seminal articles.<sup>50,51</sup> For AI-specific guidelines (e.g., CLAIM, Consolidated Standards of Reporting Trials for AI (CONSORT-AI), Standard Protocol Items: Recommendations for Interventional Trials for AI (SPIRIT-AI), FUTURE-AI, MI-CLAIM), this study referred to the publication of Klontzas et al.<sup>50</sup>, which did not limit its scope to a specific data type. For guidelines related to modeling, including AI-specific ones (e.g. TRIPOD), this

study referred to the paper of Klement and El Emam<sup>51</sup>, which primarily focused on structured data. Due to the potential omission of relevant reporting guidelines in these papers, this study refrained from confining its criteria to those listed in the aforementioned articles.

## Statistical analysis

Statistical analysis was performed using Jamovi (version 2.2.5). The majority of the findings were presented through descriptive statistics, wherein percentages were rounded to the nearest whole number. The inter-reader agreement analysis of the first two readers was conducted using Cohen's kappa or percentage agreement, as appropriate. The following grading system was used to interpret Cohen's kappa: kappa  $\leq 0.00$ , no;  $0.00 < \text{kappa} \leq 0.20$ , slight;  $0.20 < \text{kappa} \leq 0.40$ , fair;  $0.40 < \text{kappa} \leq 0.60$ , moderate;  $0.60 < \text{kappa} \leq 0.80$ , substantial;  $0.80 < \text{kappa} \leq 1$ , almost perfect agreement. Comparison of the distribution of quantitative variables was conducted using either the Student's t-test or the Mann-Whitney U test, depending on the statistical normality of the data. The chi-square test or Fisher's exact test was employed to assess differences in the distribution of categorical variables across various citation variables between subjects. Furthermore, McNemar's test was used for the same purpose within subjects, and the continuity correction was also applied. A *P* value of  $< 0.05$  was considered statistically significant.

## Results

### Baseline characteristics of journals

Out of 102 Q1 and Q2 radiology, nuclear medicine, and medical imaging journals indexed in SCIE and ESCI databases, 98 were included in this study. Four journals were excluded because they published only review articles. Of the journals included, 66 were from SCIE (Q1/Q2, 32/34), with a median 2022 impact factor of 3.9 (interquartile range: 2.4). The remaining 32 journals were from ESCI (Q1/Q2, 16/16), with a median 2022 impact factor of 2.25 (interquartile range: 1.9).

For all 98 journals, instructions specific to authors were found. However, specific instructions for reviewers or referees were found for 16 journals only (16%).

### Analysis of author instructions

Table 1 summarizes the encouragement of authors to use reporting guidelines that are specific to AI, related to modeling, and

unrelated to modeling, as well as the requirement of submission for these reporting guidelines.

Considering all 98 journals, only five journals (5%) encouraged the authors to follow AI-specific reporting guidelines. Table 2 presents the AI-specific guidelines recommended in these journals: CLAIM (n = 3), Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME) (n = 1), and Checklist for AI in Medical Physics (CLAMP) (n = 1).<sup>42,61,62</sup> Of these, three (60%) required a filled-out checklist along with the submission.

In total, 30 journals (31% of 98) endorsed at least one reporting guideline related to modeling, including both general modeling guidelines and AI-specific ones: TRIPOD (n = 26), along with the three aforementioned AI reporting guidelines, namely CLAIM, PRIME, and CLAMP.<sup>42,61-63</sup> One journal encouraged two modeling-related guidelines (TRIPOD and CLAIM). Of the 30 journals, only four (13%) required a filled-out checklist along with the submission. Furthermore, only one of the journals, Ultrasound in Obstetrics and Gynecology, encouraged TRIPOD and mandated a filled-out checklist.

A total of 75 journals (77% of 98) encouraged at least one guideline unrelated to modeling. The frequency of the most well-known guidelines in these categories is as follows: CONSORT (n = 61), Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (n = 51), Animal Research: Reporting of *In Vivo* Experiments (ARRIVE) (n = 45), STARD (n = 44), and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) (n = 42).<sup>64-68</sup> Of these journals, 36 (48%) required a filled-out checklist along with the submission.

The level of encouragement for authors, both with and without submission requirements, regarding the naming of reporting guidelines, is summarized in Figure 2, alongside a comparison with that of reviewers.

Statistically significant differences were observed in the proportions of author encouragement among pairwise comparisons of AI-specific reporting guidelines, those related to modeling, and those unrelated to modeling ( $P < 0.001$  for all). Notably, the encouragement level for guidelines unrelated to modeling was consistently higher across all pairs.

There were no statistically significant differences in the distribution of author encouragement status concerning the journal index (i.e., SCIE vs. ESCI) and quartile (i.e., Q1 vs. Q2) ( $P > 0.05$  for all).

Regarding the encouragement of reporting guidelines related to modeling in general, including AI-specific ones, as well as those unrelated to modeling, the inter-rater reliability analysis yielded almost perfect agreement, with Cohen's kappa values ranging between 0.916 and 0.950.

Analysis of reviewer instructions

Table 1 summarizes the encouragement of reviewers to use reporting guidelines that are specific to AI, related to modeling, and unrelated to modeling, as well as the requirement of submission for these reporting guidelines.

Of the 16 journals that had instructions for reviewers, only one (6%), European Radiology, encouraged the reviewers to follow an AI-specific reporting guideline (CLAIM), which can also be regarded as a modeling-related guideline, without a filled-out checklist along with the submission of peer review.<sup>42</sup> The primary purpose was, however, to check whether the authors provided the checklist.

Regarding the guidelines that are not related to modeling, six journals (38% of 16) encouraged the reviewers to follow at least one of those. The journals most frequently recommended CONSORT (n = 4) and PRISMA (n = 4) without a filled-out checklist along with the submission of peer review.<sup>67,68</sup>

The summary of reviewer encouragement, both with and without submission requirements, regarding the naming of reporting guidelines, is depicted in Figure 2, alongside a comparison with that of the authors.

There was a statistically significant difference in the proportion of reviewer encouragement between AI-specific or modeling-related reporting guidelines and those unrelated to modeling ( $P < 0.025$ ), with the latter being the higher.

There were no statistically significant differences in the distribution of reviewer encouragement status against the journal index (i.e., SCIE vs. ESCI) and quartile (i.e., Q1 vs. Q2) ( $P > 0.05$  for all).

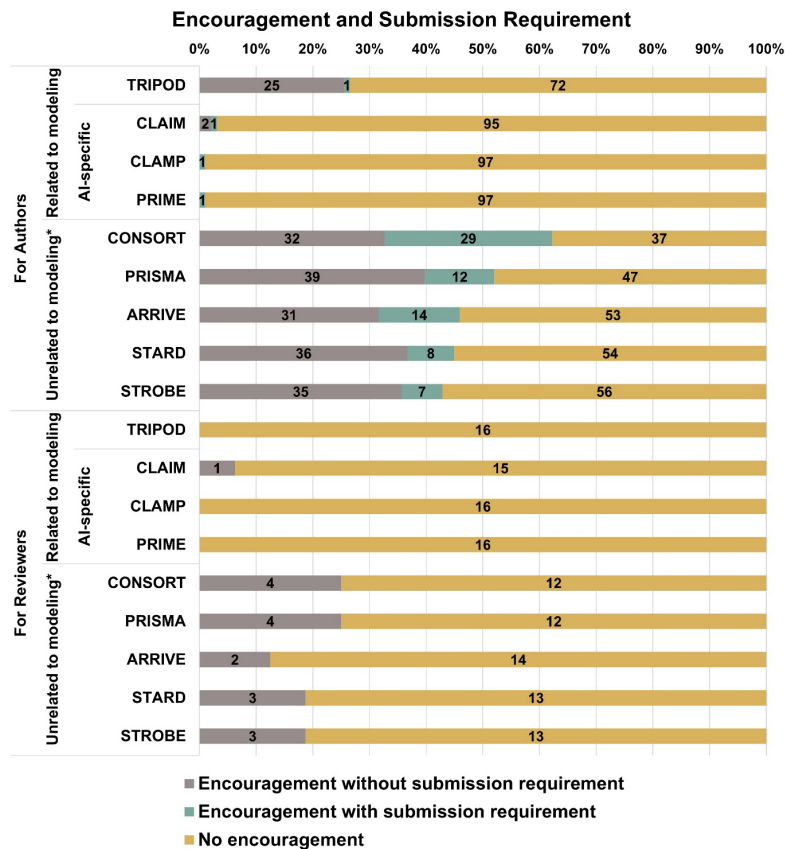
For reviewers, the encouragement of reporting guidelines related to modeling in general, including AI-specific ones, as well as those not related to modeling, resulted in high inter-rater reliability, with percentage agreement values ranging between 79% and 93%.

Table 1. Journal statistics for the encouragement and submission requirements of reporting guidelines					
Instruction	Guideline	Number of journals			
		Encouragement		Submission requirement	
		Yes	No	Yes	No
For authors	AI-specific	5 (5%)	93 (95%)	3 (60%)	2 (40%)
	Related to modeling <sup>1</sup>	30 (31%)	68 (69%)	4 (13%)	26 (87%)
	Unrelated to modeling	75 (77%)	23 (23%)	36 (48%)	39 (52%)
For reviewers	AI-specific	1 (6%)	15 (94%)	0 (0%)	16 (100%)
	Related to modeling <sup>1</sup>	1 (6%)	15 (94%)	0 (0%)	16 (100%)
	Unrelated to modeling	6 (38%)	10 (62%)	0 (0%)	16 (100%)

<sup>1</sup>Including AI-specific ones. AI, artificial intelligence.

Table 2. Encouraged AI-specific reporting guidelines for authors and their submission requirements					
Index	Journal title	2022 JIF	Q	AI-specific reporting guideline	
				Encouragement	Submission requirement
SCIE	JACC: Cardiovascular Imaging	14	Q1	PRIME	Yes
	European Radiology	5.9	Q1	CLAIM	No
	Journal of the American College of Radiology	4.5	Q1	CLAIM	No
	Medical Physics	3.8	Q2	CLAMP	Yes
ESCI	Radiology: Artificial Intelligence	9.8	Q1	CLAIM	Yes

SCIE, Science Citation Index Expanded; ESCI, Emerging Sources Citation Index; JIF, Journal Impact Factor; Q, quartile (according to the category radiology, nuclear medicine, and medical imaging); AI, artificial intelligence; Q1, first quartile; Q2, second quartile; PRIME, Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation; CLAIM, Checklist for AI in Medical Imaging; CLAMP, Checklist for AI in Medical Physics.



**Figure 2.** Encouragement of authors and reviewers for the reporting guidelines with and without submission requirements. AI, artificial intelligence. \*Only the top five most frequently encouraged reporting guidelines were included.

## Discussion

### Overview

This meta-research investigated how radiology, nuclear medicine, and medical imaging journals encourage and mandate (i.e., requiring a filled-out checklist) the use of AI reporting guidelines in their author and reviewer instructions. The results were presented by comparing reporting guidelines that are specific to AI, related to modeling, and unrelated to modeling. It was found that only a very small number of journals encouraged (5%, 5/98) and mandated (3%, 3/98) the use of AI reporting guidelines (i.e., CLAIM, PRIME, and CLAMP) for authors. In addition, only one journal (6% of 16 available reviewer instructions) encouraged the reviewers to follow AI reporting guidelines (i.e., CLAIM), without any requirement of submission. Encouragement and the mandated use of AI-specific guidelines and those related to modeling in the journals were generally lower compared with those unrelated to modeling.

### Previous related works

There is only one recent closely related study to this research in which the endorse-

ment of AI reporting guidelines in radiology journals has been systematically analyzed.<sup>60</sup> In their seminal study, Zhong et al.<sup>60</sup> investigated the endorsement of 15 general reporting guidelines and 10 AI reporting guidelines. Of the 117 SCIE journals included, the authors found that CLAIM (1.7%, 2/117) was the only and the most implemented AI reporting guideline, while the other nine AI reporting guidelines were not mentioned. This study found that five (5%) out of the 98 journals encouraged AI-specific guidelines. The disparity in rates can be attributed to two methodological issues. First, the journals differed in their index sources. Second, our study encompassed half of the journals indexed in SCIE and ESCI (i.e., Q1–Q2). In contrast, Zhong et al.<sup>60</sup> exclusively included all SCIE journals. Furthermore, the AI reporting guidelines considered for these works were different. This study referenced two prior works and imposed no additional restrictions on their use, provided that they were in the form of a publication (i.e., not a custom checklist that appears on the instructions of journals).<sup>50,51</sup> The authors of the prior investigation restricted their assessment to 10 AI reporting guidelines. Furthermore, both studies reached the same conclusion that the

endorsement or encouragement to follow AI reporting guidelines in these journals was remarkably low. Their main findings were complementary and mutually reinforcing.

Given the scarcity of literature on the encouragement of AI reporting guidelines in radiology, nuclear medicine, and medical imaging journals, it would be beneficial to discuss studies that are not specifically pertinent to AI but are nevertheless extremely relevant to the encouragement of and mandating reporting guidelines. In a cross-sectional study, Malički et al.<sup>69</sup> analyzed a representative sample of journal instructions for authors across multiple scientific fields, including health sciences. The instructions of 13% of journals suggested the use of reporting guidelines, while only 2% mandated its use. In addition, the authors discovered that journals in the health or life sciences, as well as those published by prominent publishers, were more likely to include reporting guidelines or standards in their author instructions. In a different study, Agha et al.<sup>70</sup> investigated the impact of the mandatory implementation of reporting guidelines on the quality of reporting in a surgical journal. Compliance with STROBE, CONSORT, and PRISMA dramatically improved after the policy implementation. The authors observed that implementing a policy demanding the submission of a completed reporting checklist for observational research, randomized controlled trials, and systematic reviews can increase compliance. In addition, they recommended similar approaches for various journals and study types. In another seminal study, Hirst and Altman focused on the encouragement of reviewers to utilize reporting guidelines for 116 health research publications.<sup>59</sup> They discovered that 41 (35%) of the journals offered reviewers with online instructions. In addition, they revealed that nearly half of the online instructions referred to these tools without providing clear instructions on how to use them.

### Potential reasons for low rates

Considering the relevant works above and the present study, it is evident that journals do not encourage and mandate AI reporting guidelines frequently. The potential causes can only be speculated because their analysis falls outside the scope of this study. The editorial team of the journals may wrongly presume that researchers are aware of these fundamental aspects of rigorous and transparent reporting and that authors are entirely responsible for implementing them, not the journals. The journals may also be hesitant to incorporate appropriate

reporting practices through reporting guidelines, and they may be unwilling to address scientific misconduct and correct publication errors.<sup>71-73</sup> The editors may also not want to unintentionally overburden the authors with too many instructions. Even if journals encourage good reporting practices, researchers may be resistant to fundamental change. Furthermore, despite the validity of these tools, journals may not agree on the importance of reporting guidelines and may be hesitant to recommend their usage in the absence of convincing proof of their effectiveness.

### What are the following steps?

In light of the outstanding and exponential growth of AI research on medical imaging over the past decade,<sup>23</sup> it is necessary to promote the highest-quality research. It would be advantageous to conduct additional research to define the effectiveness of AI reporting guidelines. Such research will help persuade journals to encourage and mandate them. Hence, there is a need for further assessment of AI reporting guidelines to determine their optimal utilization. This assessment should consider whether they should be incorporated into the study design, applied during ongoing research, utilized solely for reporting purposes post-study completion, or implemented at the request of journals, among other potential considerations. Enhancing our understanding of the factors that influence the dissemination and implementation of these tools and strategies is crucial for improving their efficacy and promoting their broader adoption. Future research should investigate the obstacles journals might experience when adopting such policy changes in their journals, as well as how automated tools could minimize their workload while guaranteeing adherence to these reporting guidelines. Furthermore, radiology, nuclear medicine, and medical imaging journals may collaborate to improve reporting standards for research. These group initiatives should also be supported by scientific organizations, universities, institutions, societies, and funding agencies. This would make it more difficult for authors receiving negative reviews due to inadequate reporting to choose journals with more flexible reporting policies. This could enhance the overall reporting quality of the scientific literature. In certain areas of medical research, such as rehabilitation and disability, the journals established such collaborations.<sup>74</sup> As of 2014, 28 prominent rehabilitation and disability journals have joined

a group to require the adherence to reporting guidelines to increase the quality of research reporting, not just inside their journal but also within their field of medicine and research. They jointly published an editorial, announcing their agreement and urging authors to adhere to appropriate EQUATOR reporting guidelines when preparing articles for submission. They also requested reviewers to utilize reporting guidelines when evaluating submissions.<sup>74</sup> A similar group effort is crucial to improve the overall reporting quality of AI research in radiology, nuclear medicine, and medical imaging journals.

### Limitations

This study has a few limitations. First, it assumed that instructions are the sole location where reporting guidelines that are encouraged or mandated can be found. However, some of the requirements editors put on authors and reviewers may not be necessarily outlined in the instructions. For instance, the submission systems of all the journals were not thoroughly analyzed to check whether they encouraged or requested the use of guidelines during the submission and/or review processes. It was presumed that this was not common practice. Nonetheless, their submission systems were only investigated when the submission requirements were not clear in the instructions. Second, only Q1 and Q2 SCIE and ESCI journals indexed in the WoS were included due to their well-known high standards for indexing. Therefore, it is unlikely to represent the editorial standards of all journals. To diversify the journal characteristics, Q1 and Q2 ESCI journals were included instead of Q3 and Q4 SCIE journals. However, achieving a perfect representation of journals in terms of diversity should not be a major concern in an exploratory study focusing on a new area of reporting guidelines. Third, while downloading the journal instructions, they were double-checked for accuracy. Due to the complex and multi-layered design of certain journal websites, some parts of the instructions may have been omitted. Additionally, this study aimed to evaluate the automatically annotated content of the instructions through independent readings by two readers, with consensus reached through consultation with a third reader. This study may have missed any reporting guidelines that were recommended or deemed necessary in the submission. However, the impact of missing instructions and their content analysis will likely be minor. Finally, the instructions were downloaded over a brief time frame (between September 4 and 7,

2023). If journals had improved their instructions after this period, these changes would not have been reflected in the results.

In conclusion, this meta-research study provides an overview of instructions for authors and peer reviewers across radiology, nuclear medicine, and medical imaging journals. It specifically examines the encouragement of AI-specific reporting guidelines and their submission requirements, comparing them with guidelines related to modeling and those unrelated to modeling. However, the findings indicate that these AI-specific guidelines are not commonly encouraged and mandated (i.e., requiring a filled-out checklist) by these journals, compared with other guidelines. To further encourage the use of these tools, all stakeholders, including developers, journal managers, editors, reviewers, authors, and funders, are required to collaborate. Given their position at the forefront of AI, if more of these journals enforce or encourage responsible reporting through guidelines, the value of articles and AI research may increase and become less wasteful.

### Conflict of interest disclosure

Burak Koçak, MD, is Section Editor in Diagnostic and Interventional Radiology. He had no involvement in the peer-review of this article and had no access to information regarding its peer-review. Other authors have nothing to disclose.

### References

1. Jin Y, Sanger N, Shams I, et al. Does the medical literature remain inadequately described despite having reporting guidelines for 21 years? - A systematic review of reviews: an update. *J Multidiscip Healthc*. 2018;11:495-510. [\[CrossRef\]](#)
2. Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet Lond Engl*. 2014;383(9913):267-276. [\[CrossRef\]](#)
3. Macleod MR, Michie S, Roberts I, et al. Biomedical research: increasing value, reducing waste. *Lancet*. 2014;383(9912):101-104. [\[CrossRef\]](#)
4. Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ*. 2005;330(7494):753. [\[CrossRef\]](#)
5. Altman DG, Simera I. Responsible reporting of health research studies: transparent, complete, accurate and timely. *J Antimicrob Chemother*. 2010;65(1):1-3. [\[CrossRef\]](#)
6. Goodman SN, Fanelli D, Ioannidis JPA. What does research reproducibility mean? *Sci Transl Med*. 2016;8(341):341ps12. [\[CrossRef\]](#)

7. Casas JP, Kwong J, Ebrahim S. Telemonitoring for chronic heart failure: not ready for prime time. *Cochrane Database Syst Rev*. 2010;2011:ED000008. [\[CrossRef\]](#)
8. Fuller T, Pearson M, Peters J, Anderson R. What affects authors' and editors' use of reporting guidelines? Findings from an online survey and qualitative interviews. *PLoS One*. 2015;10(4):e0121585. [\[CrossRef\]](#)
9. Moher D, Weeks L, Ocampo M, et al. Describing reporting guidelines for health research: a systematic review. *J Clin Epidemiol*. 2011;64(7):718-742. [\[CrossRef\]](#)
10. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med*. 2010;8(1):24. [\[CrossRef\]](#)
11. Cobo E, Cortés J, Ribera JM, et al. Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: masked randomised trial. *BMJ*. 2011;343:d6783. [\[CrossRef\]](#)
12. Botos J. Reported use of reporting guidelines among JNCI: Journal of the National Cancer Institute authors, editorial outcomes, and reviewer ratings related to adherence to guidelines and clarity of presentation. *Res Integr Peer Rev*. 2018;3:7. [\[CrossRef\]](#)
13. Stevanovic A, Schmitz S, Rossaint R, Schürholz T, Coburn M. CONSORT item reporting quality in the top ten ranked journals of critical care medicine in 2011: a retrospective analysis. *PLoS One*. 2015;10(5):e0128061. [\[CrossRef\]](#)
14. Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust*. 2006;185(5):263-267. [\[CrossRef\]](#)
15. Moher D, Jones A, Lepage L; CONSORT Group (Consolidated Standards for Reporting of Trials). Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA*. 2001;285(15):1992-1995. [\[CrossRef\]](#)
16. Prady SL, Richmond SJ, Morton VM, Macpherson H. A systematic evaluation of the impact of STRICTA and CONSORT recommendations on quality of reporting for acupuncture trials. *PLoS One*. 2008;3(2):e1577. [\[CrossRef\]](#)
17. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ*. 2010;340:c723. [\[CrossRef\]](#)
18. Wynne KE, Simpson BJ, Berman L, Rangel SJ, Grosfeld JL, Moss RL. Results of a longitudinal study of rigorous manuscript submission guidelines designed to improve the quality of clinical research reporting in a peer-reviewed surgical journal. *J Pediatr Surg*. 2011;46(1):131-137. [\[CrossRef\]](#)
19. Smidt N, Rutjes AW, van der Windt DA, et al. The quality of diagnostic accuracy studies since the STARD statement: has it improved? *Neurology*. 2006;67(5):792-797. [\[CrossRef\]](#)
20. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst Rev*. 2012;1:60. [\[CrossRef\]](#)
21. Tunis AS, McInnes MD, Hanna R, Esmail K. Association of study quality with completeness of reporting: have completeness of reporting and quality of systematic reviews and meta-analyses in major radiology journals changed since publication of the PRISMA statement? *Radiology*. 2013;269(2):413-426. [\[CrossRef\]](#)
22. Hong PJ, Korevaar DA, McGrath TA, et al. Reporting of imaging diagnostic accuracy studies with focus on MRI subgroup: Adherence to STARD 2015. *J Magn Reson Imaging*. 2018;47(2):523-544. [\[CrossRef\]](#)
23. Kocak B, Baessler B, Cuocolo R, Mercaldo N, Pinto Dos Santos D. Trends and statistics of artificial intelligence and radiomics research in radiology, nuclear medicine, and medical imaging: bibliometric analysis. *Eur Radiol*. 2023;33(11):7542-7555. [\[CrossRef\]](#)
24. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. 2019;1(6):271-297. [\[CrossRef\]](#)
25. Cabitza F, Campagner A. The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical AI studies. *Int J Med Inform*. 2021;153:104510. [\[CrossRef\]](#)
26. Kocak B, Keles A, Akinci D'Antonoli T. Self-reporting with checklists in artificial intelligence research on medical imaging: a systematic review based on citations of CLAIM. *Eur Radiol*. 2023. [\[CrossRef\]](#)
27. Marwaha JS, Chen HW, Habashy K, et al. Appraising the quality of development and reporting in surgical prediction models. *JAMA Surg*. 2023;158(2):214-216. [\[CrossRef\]](#)
28. Emam KE, Klement W, Malin B. Reporting and methodological observations on prognostic and diagnostic machine learning studies. *JMIR AI*. 2023;2(1):e47995. [\[CrossRef\]](#)
29. Ibrahim H, Liu X, Denniston AK. Reporting guidelines for artificial intelligence in healthcare research. *Clin Experiment Ophthalmol*. 2021;49(5):470-476. [\[CrossRef\]](#)
30. Shelmerdine SC, Arthurs OJ, Denniston A, Sebire NJ. Review of study reporting guidelines for clinical studies using artificial intelligence in healthcare. *BMJ Health Care Inform*. 2021;28(1):e100385. [\[CrossRef\]](#)
31. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. *BMJ*. 2020;370:m3210. [\[CrossRef\]](#)
32. Kocak B, Akinci D'Antonoli T, Ates Kus E, et al. Self-reported checklists and quality scoring tools in radiomics: a meta-research. *Eur Radiol*. 2024. [\[CrossRef\]](#)
33. Olczak J, Pavlopoulos J, Prijs J, et al. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. *Acta Orthop*. 2021;92(5):513-525. [\[CrossRef\]](#)
34. Padula WV, Kreif N, Vanness DJ, et al. Machine learning methods in health economics and outcomes research-the PALISADE checklist: a good practices report of an ISPOR task force. *Value Health*. 2022;25(7):1063-1080. [\[CrossRef\]](#)
35. Loftus TJ, Tighe PJ, Ozrazgat-Baslanti T, et al. Ideal algorithms in healthcare: explainable, dynamic, precise, autonomous, fair, and reproducible. *PLOS Digit Health*. 2022;1(1):e0000006. [\[CrossRef\]](#)
36. Weaver CGW, Basmadjian RB, Williamson T, et al. Reporting of model performance and statistical methods in studies that use machine learning to develop clinical prediction models: protocol for a systematic review. *JMIR Res Protoc*. 2022;11(3):e30956. [\[CrossRef\]](#)
37. Kotecha D, Asselbergs FW, Achenbach S, et al. CODE-EHR best-practice framework for the use of structured electronic health-care records in clinical research. *Lancet Digit Health*. 2022;4(10):757-764. [\[CrossRef\]](#)
38. Seedat N, Imrie F, van der Schaar M. DC-Check: a data-centric AI checklist to guide the development of reliable machine learning systems. Published online November 9, 2022. [\[CrossRef\]](#)
39. 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. *BMJ*. 2022;377:e070904. [\[CrossRef\]](#)
40. Lu JH, Callahan A, Patel BS, et al. Assessment of adherence to reporting guidelines by commonly used clinical prediction models from a single vendor: a systematic review. *JAMA Netw Open*. 2022;5(8):e2227779. [\[CrossRef\]](#)
41. Crossnohere NL, Elsaid M, Paskett J, Bose-Brill S, Bridges JFP. Guidelines for artificial intelligence in medicine: literature review and content analysis of frameworks. *J Med Internet Res*. 2022;24(8):e36823. [\[CrossRef\]](#)
42. Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging

- (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell.* 2020;2(2):e200029. [\[CrossRef\]](#)
43. Tejani AS, Klontzas ME, Gatti AA, et al. Updating the checklist for artificial intelligence in medical imaging (CLAIM) for reporting AI research. *Nat Mach Intell.* 2023;5(9):950-951. [\[CrossRef\]](#)
  44. Lekadir K, Osuala R, Gallin C, et al. FUTURE-AI: Guiding principles and consensus recommendations for trustworthy artificial intelligence in medical imaging. Published online October 1, 2023. [\[CrossRef\]](#)
  45. Norgeot B, Quer G, Beaulieu-Jones BK, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat Med.* 2020;26(9):1320-1324. [\[CrossRef\]](#)
  46. Kocak B, Baessler B, Bakas S, et al. Checklist for evaluation of radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMI. *Insights Imaging.* 2023;14:75. [\[CrossRef\]](#)
  47. Kocak B, Akinci D'Antonoli T, Mercaldo N, et al. Methodological radiomics score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMI. *Insights Imaging.* 2024;15(1):8. [\[CrossRef\]](#)
  48. Sounderajah V, Ashrafian H, Golub RM, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. *BMJ Open.* 2021;11(6):e047709. [\[CrossRef\]](#)
  49. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open.* 2021;11(7):e048008. [\[CrossRef\]](#)
  50. Klontzas ME, Gatti AA, Tejani AS, Kahn CE Jr. AI reporting guidelines: how to select the best one for your research. *Radiol Artif Intell.* 2023;5(3):e230055. [\[CrossRef\]](#)
  51. Klement W, El Emam K. Consolidated reporting guidelines for prognostic and diagnostic machine learning modeling studies: development and validation. *J Med Internet Res.* 2023;25:e48763. [\[CrossRef\]](#)
  52. Leung V, Rousseau-Blass F, Beauchamp G, Pang DSJ. ARRIVE has not ARRIVED: support for the ARRIVE (animal research: reporting of in vivo experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. *PLoS One.* 2018;13(5):e0197882. [\[CrossRef\]](#)
  53. Reveiz L, Villanueva E, Iko C, Simera I. Compliance with clinical trial registration and reporting guidelines by Latin American and Caribbean journals. *Cad Saude Publica.* 2013;29(6):1095-1100. [\[CrossRef\]](#)
  54. Pussegoda K, Turner L, Garritty C, et al. Systematic review adherence to methodological or reporting quality. *Syst Rev.* 2017;6(1):131. [\[CrossRef\]](#)
  55. Page MJ, Shamseer L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med.* 2016;13(5):e1002028. [\[CrossRef\]](#)
  56. Diong J, Butler AA, Gandeveia SC, Héroux ME. Poor statistical reporting, inadequate data presentation and spin persist despite editorial advice. *PLoS One.* 2018;13(8):e0202121. [\[CrossRef\]](#)
  57. Innocenti T, Giaggio S, Salvioi S, et al. Completeness of reporting is suboptimal in randomized controlled trials published in rehabilitation journals, with trials with low risk of bias displaying better reporting: a meta-research study. *Arch Phys Med Rehabil.* 2022;103(9):1839-1847. [\[CrossRef\]](#)
  58. Diong J, Bye E, Djajadikarta Z, Butler AA, Gandeveia SC, Héroux ME. Encouraging responsible reporting practices in the instructions to authors of neuroscience and physiology journals: there is room to improve. *PLoS One.* 2023;18(3):e0283753. [\[CrossRef\]](#)
  59. Hirst A, Altman DG. Are peer reviewers encouraged to use reporting guidelines? A survey of 116 health research journals. *PLoS One.* 2012;7(4):e35621. [\[CrossRef\]](#)
  60. Zhong J, Xing Y, Lu J, et al. The endorsement of general and artificial intelligence reporting guidelines in radiological journals: a meta-research study. *BMC Med Res Methodol.* 2023;23(1):292. [\[CrossRef\]](#)
  61. Sengupta PP, Shrestha S, Berthon B, et al. Proposed requirements for cardiovascular imaging-related machine learning evaluation (PRIME): a checklist: reviewed by the American College of Cardiology Healthcare Innovation Council. *JACC Cardiovasc Imaging.* 2020;13(9):2017-2035. [\[CrossRef\]](#)
  62. El Naqa I, Boone JM, Benedict SH, et al. AI in medical physics: guidelines for publication. *Med Phys.* 2021;48(9):4711-4714. [\[CrossRef\]](#)
  63. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ. BMC Med.* 2015;13(1):1. [\[CrossRef\]](#)
  64. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med.* 2007;147(8):573-577. [\[CrossRef\]](#)
  65. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015;351:h5527. [\[CrossRef\]](#)
  66. Percie du Sert N, Hurst V, Ahluwalia A, et al. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *PLoS Biol.* 2020;18(7):e3000410. [\[CrossRef\]](#)
  67. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. [\[CrossRef\]](#)
  68. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;8(1):18. [\[CrossRef\]](#)
  69. Malički M, Aalbersberg IJ, Bouter L, Ter Riet G. Journals' instructions to authors: a cross-sectional study across scientific disciplines. *PLoS One.* 2019;14(9):e0222157. [\[CrossRef\]](#)
  70. Agha RA, Fowler AJ, Limb C, et al. Impact of the mandatory implementation of reporting guidelines on reporting quality in a surgical journal: A before and after study. *Int J Surg.* 2016;30:169-172. [\[CrossRef\]](#)
  71. Bosch X, Hernández C, Pericas JM, Doti P, Marušić A. Misconduct policies in high-impact biomedical journals. *PLoS One.* 2012;7(12):e51928. [\[CrossRef\]](#)
  72. Williams P, Wager E. Exploring why and how journal editors retract articles: findings from a qualitative study. *Sci Eng Ethics.* 2013;19(1):1-11. [\[CrossRef\]](#)
  73. Wager E. Coping with scientific misconduct. *BMJ.* 2011;343:d6586. [\[CrossRef\]](#)
  74. Chan L, Heinemann AW, Roberts J. Elevating the quality of disability and rehabilitation research: mandatory use of the reporting guidelines. *Am J Phys Med Rehabil.* 2014;93(4):279-281. [\[CrossRef\]](#)