

Michael X. Jin, MD¹, Austin Young², Katherine Wang², Kian Avilla², Kevin Gilotra², Pamela Nguyen MD³, Pablo Ros MD¹

1. Department of Radiology, Stony Brook University Hospital, Stony Brook, NY

2. Renaissance School of Medicine at Stony Brook University, Stony Brook, NY

3. Department of Radiology, Columbia University Irving Medical Center, New York, NY

INTRODUCTION

- In radiology, the generation of reports is critical for patient management, though can be time-consuming and labor-intensive.
- Studies have demonstrated the potential of AI in generating comprehensible and clinically relevant radiology reports.
- This study aims investigate the feasibility of a hands-free radiology reporting system leveraging voice-to-text technology and generative AI to streamline the reporting process (i.e., improve time efficiency, aid report standardization, and reduce radiologist workload).

METHODS

- This study utilized a prospective design:
 - 100 chest X-ray and MSK reports were generated using traditional methods and an automated process using generative AI.
 - A radiologist dictated their findings into a voice-to-text software (Dragon) creating a raw text transcript.
 - GPT-4 was used to develop an AI-generated report based on our institution's template.
 - With the radiologist's traditional reports as comparisons, two additional radiologists evaluated all reports for quality on a 5-point Likert scale.

5/5: The report is comprehensive, offering an exceptional impression that matches the quality and substance of the radiologists' assessment. No modifications are deemed necessary.

4/5: The report is detailed and conveys a good impression, closely approximating the radiologists' impression. Minimal modifications required.

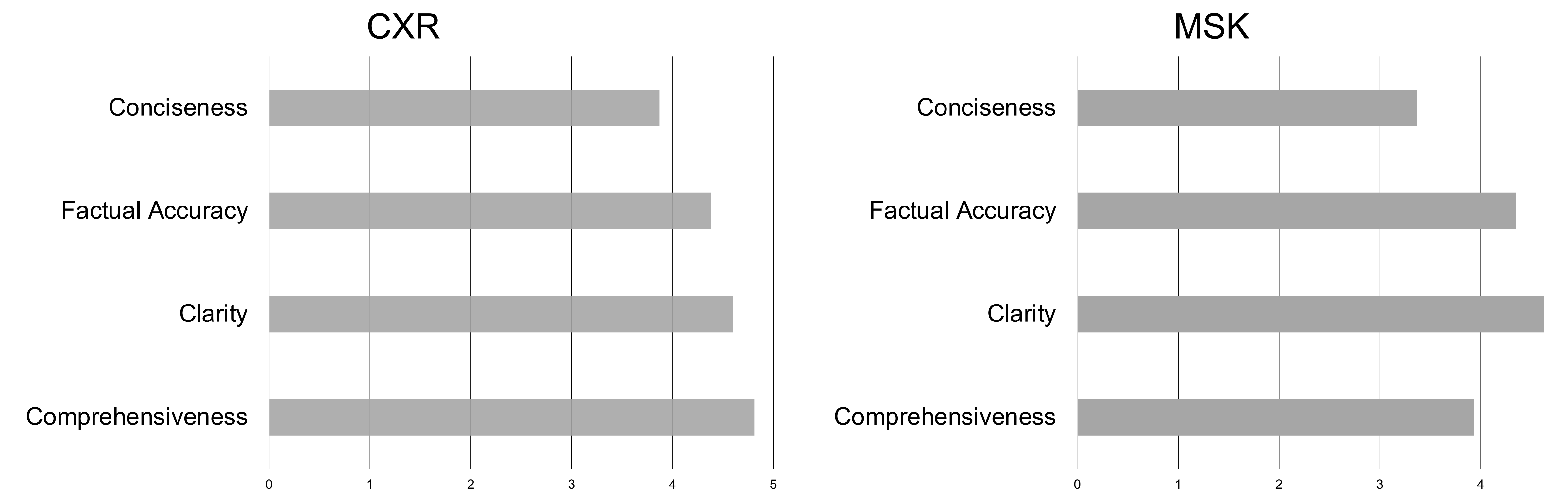
3/5: The report is satisfactory and conveys an adequate impression but omits minor details. Moderate modifications required.

2/5: The report is missing critical information and fails to provide an impression that meets basic professional standards. Substantial modifications required.

1/5: The report is fundamentally inadequate, providing a poor impression lacking in key areas of information. Extensive revisions required.

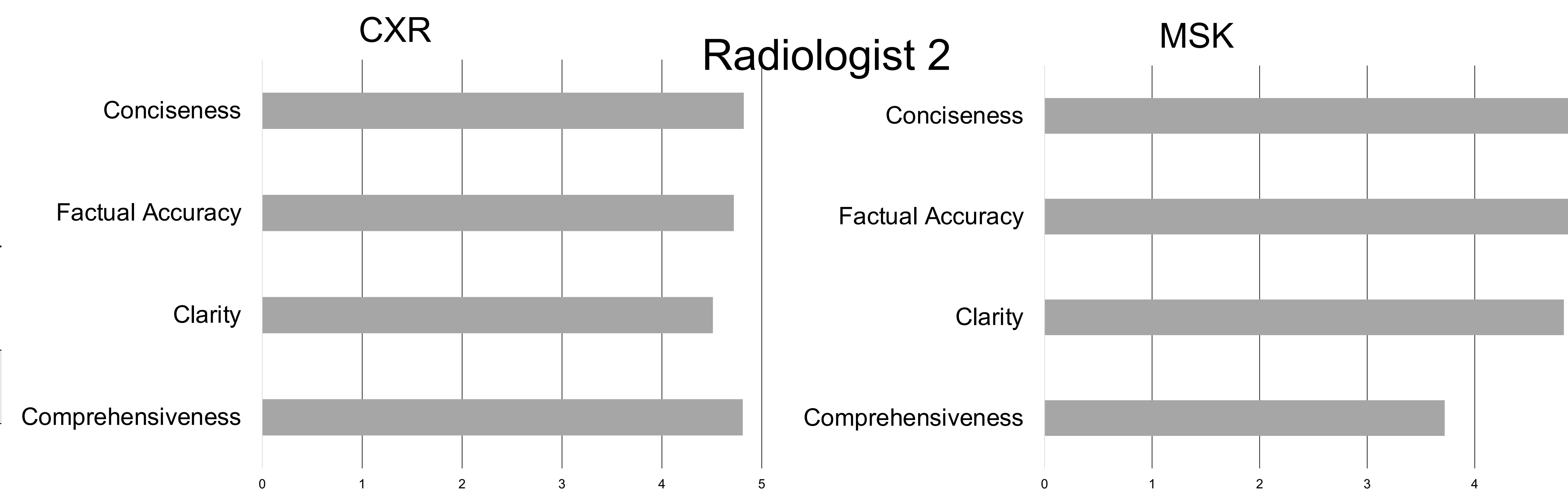
Figure 1: Likert Scale utilized to evaluate conciseness, factual accuracy, clarity, and comprehensiveness of radiology reports.

Radiologist 1



Figures 2a, 2b: Average values of conciseness, factual accuracy, clarity, and comprehensiveness of chest X-ray (CXR) and musculoskeletal (MSK) generative AI reports, as evaluated by radiologist 1.

Radiologist 2



Figures 3a, 3b: Average values of conciseness, factual accuracy, clarity, and comprehensiveness of chest X-ray (CXR) and musculoskeletal (MSK) generative AI reports, as evaluated by radiologist 2.

RESULTS

- This study evaluated the performance of generative AI in producing radiology reports for chest X-ray (CXR) and musculoskeletal (MSK) studies.
- Two independent radiologists assessed the AI-generated reports using a 5-point Likert scale across four metrics: comprehensiveness, clarity, factual accuracy, and conciseness. Likert scores (combined CXR and MSK):
 - Average among the two radiologists: 4.35 (comprehensiveness), 4.64 (clarity), 4.59 (factual accuracy), and 4.26 (conciseness).
 - Individual from radiologist 1: 4.39 (comprehensiveness), 4.61 (clarity), 4.35 (factual accuracy), and 3.63 (conciseness).
 - Individual from radiologist 2: 4.31 (comprehensiveness), 4.66 (clarity), 4.82 (factual accuracy), and 4.89 (consciousness).
- No significant differences were found between the two radiologists' scores for comprehensiveness ($p=0.56$) and clarity ($p=0.63$).
- A significant difference was observed in the scores for factual accuracy ($p=0.0004$) and conciseness ($p=0.0001$).

CONCLUSIONS

- This study highlights the potential of voice-to-text and generative AI integration in radiology reporting.
- Increased time efficiency can allow radiologists to focus more on complex cases and patient care.
- Further development and refinement are required to address comprehensiveness and occasional inaccuracies.
- Ongoing advancements in AI and speech recognition technologies suggest a positive outlook for future implementations in medical reporting.

REFERENCES

1. Amin KS, Davis MA, Doshi R, Haims AH, Khosla P, Forman HP. Accuracy of ChatGPT, Google Bard, and Microsoft Bing for Simplifying Radiology Reports. *Radiology*. 2023;309(2):e232561. doi:10.1148/radiol.232561
2. Bajaj S, Gandhi D, Nayar D. Potential Applications and Impact of ChatGPT in Radiology. *Acad Radiol*. Published online October 5, 2023. doi:10.1016/j.acra.2023.08.039
3. Adams LC, Truhn D, Busch F, et al. Leveraging GPT-4 for Post Hoc Transformation of Free-text Radiology Reports into Structured Reporting: A Multilingual Feasibility Study. *Radiology*. 2023;307(4):e230725. doi:10.1148/radiol.230725