

Large language models in simplifying radiological reports: systematic review

Yaara Artsi, BMedSC¹; Vera Sorin, MD²⁻⁴; Eli Konen, MD²⁻³; Benjamin S. Glicksberg, PhD⁵; Girish Nadkarni, MD⁵⁻⁶; Eyal Klang, MD²⁻⁶

¹ Azrieli Faculty of Medicine, Bar-Ilan University, Zefat, Israel.

² Department of Diagnostic Imaging, Chaim Sheba Medical Center, Israel

³ Tel-Aviv University School of Medicine, Israel

⁴ DeepVision Lab, Chaim Sheba Medical Center, Israel

⁵ Division of Data-Driven and Digital Medicine (D3M), Icahn School of Medicine at Mount

Sinai, New York, New York, USA

⁶ The Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at

Mount Sinai, New York, New York, USA.

Abstract

Objectives

Simplifying medical information to make it understandable for patients, specifically in the case of radiology reports, is challenging. It requires time and effort from medical personnel. This systematic review focuses on the application of large language models (LLMs) in generating simplified radiological imaging reports, as well as answering patient inquiries regarding radiological procedures.

Materials and Methods

The authors searched for studies published up to January 2024. Search terms focused on LLMs generated simplified radiological reports and answers to patient inquiries regarding radiological procedures. MEDLINE was used as a search database.

Results

Overall, eight studies published between May 2023 and November 2023 were included. All studies showed that LLMs can produce simplified medical information for patients. Four studies (50%) used GPT-3.5, Two studies (25%) conducted a comparative analysis between GPT-3.5 and GPT-4. One study (12.5%) examined Microsoft Bing. One study (12.5%) utilized GPT-4. Four studies (50%) used LLMs to simplify radiological reports. Four studies (50%) used LLMs to answer patient questions regarding radiological procedures. Only two studies (25%) used patients to evaluate the LLMs output. One study (12.5%) compared their initial prompt with optimized prompt. Five studies (62.5%) showed missing, inaccurate and potentially harmful AI outputs.

Conclusion

LLMs can be used to simplify medical imaging reports and procedures, for improved patient comprehension. However, their limitations cannot be ignored. Further study in this field is essential and more conclusive evidence is needed.

Keywords

Large language models, Generative Pre-trained Transformer, Patient Education, Artificial intelligence, Radiology

Introduction

Reading, understanding, and interpreting radiographic images, reports and procedures is challenging, even for non-radiologist physicians [1, 2]. The task is many more times difficult for patients who have no medical training [3, 4]. Nowadays, medical imaging is an integral part of the clinical decision making process [5, 6, 7].

According to the patient-centered care approach, patients should be active participants in their care [4, 8]. Although patients have access to their imaging reports, these reports are frequently incomprehensible to the average patient [3, 4]. Using complex medical terminology can create patient anxiety and a perception of exacerbated severity of their condition [9]. The lack of access to simplified information for patients regarding their healthcare highlights a shortcoming of modern healthcare practice.

Large language models (LLMs), such as ChatGPT, can be used to analyze free-text and generate human-like responses to various inquiries [10, 11]. It is possible that this technology might hold the key to bridge the gap between patients and complicated medical imaging reports jargon. However, the possibility of patients seeking clarification from AI poses risks. The accuracy and credibility of such models is still up for debate and can potentially misinform patients' diagnoses and outcomes [12].

The aim of this study is to systematically review the literature on applications of LLMs for patient education and simplification of radiological reports and procedures.

Methods

Literature search

For this retrospective review, we conducted a search to identify studies describing LLMs' applications for patient education. We searched PubMed/MEDLINE for papers published not earlier than 2023 up to January 2024. The following keywords were used with Boolean operators AND/OR: *large language models, ChatGPT, openAI, patient, education*.

We checked the references list of selected publications for more relevant papers. Sections as 'Similar Articles' (e.g., PubMed) were also inspected for possible additional articles.

Our study followed the Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) guidelines. The study is registered with PROSPERO

Inclusion and exclusion process

Publications resulting from the search were initially assessed by one author (YA) for relevant titles and abstracts. Next, full-text papers underwent an independent evaluation by two authors (EK and VS).

We included full length studies describing LLMs application for patient education focusing on radiology and imaging reports. We excluded papers published before 2023, non-English papers and non-original studies (**Figure 1**).

Discrepancies were discussed and resolved to achieve a consensus. Risk of bias and applicability were evaluated using the tailored QUADAS-2 tool (**Figure 2**).

Results

Study selection and characteristics

The initial literature search resulted in 729 articles. Eight studies met our inclusion criteria (**Figure 1**). Six studies were retrospective. One study is cross-sectional (descriptive). Majority of the studies used ChatGPT (versions 3.5 or 4) as an AI model of choice, one study used Microsoft's Bing. The prompts were phrased differently in each study. One study conducted a comparison between initial and optimized prompt. Two studies involved patients' evaluation of the simplified reports (**Figure 3**).

Descriptive summary of results

Lyu et al. [13] collected a total of 138 imaging reports, 76 chest CT and 62 brain MRI (**Table 2**). All reports were anonymized. GPT-3.5 was provided with three initial prompts requesting simplification of the reports (**Table 7**). The evaluation focused on three aspects: overall score, completeness, and correctness. The number of places

with missing information and with incorrect information was recorded as well. Two radiologists evaluated the results using a 5-point system and word count. Patients did not evaluate the simplified reports. For the chest CT reports, 85.5% of the translated results (53 of 62) were shorter than the corresponding original reports. The overall length reduction was 26.7%. In brain MRI radiology reports, 72.4% of the translated results (55 of 76) contained fewer words than the corresponding original reports. The overall length reduction was 21.1% (**Table 3**).

Negative performance of ChatGPT was documented as well, showing instances of inaccurate, missing or incorrect information (**Table 5**). They've also compared GPT-3.5 to GPT-4, with GPT-4 outperforming in all aspects. Lastly, they tested the difference between the original prompt and an optimized prompt. The overall quality of translation increased from 55.2% to 77.2%, and the measures on information that were completely omitted, partially translated, and misinterpreted were reduced to 9.2%, 13.6%, and 0%, respectively (**Table 4**).

Li et al. [14] randomly sampled 100 radiographs (XR), 100 ultrasound (US), 100 CT, and 100 MRI radiology reports (**Table 2**). They prompted GPT-3.5 for simplified reports. Mean report length, Flesch reading ease score (FRES), and Flesch-Kincaid reading level (FKRL) were calculated for each original report and GPT-3.5 simplified output (**Table 3**). Patients did not evaluate the simplified reports. Negative GPT-3.5 performance was not detailed. Following simplification by GPT-3.5, all reports had a FKRL <8.5 and 77/100 (77%) of XR, 76/100 (76%) of US, 65/100 (65%) of CT, and 58/100 (58%) of MRI reports had a FKRL <6.5 (**Table 3**)

Grewal et al. [15] tested GPT-4 application in radiology across several fields, including patient education. GPT-4 generated patient-oriented explanations of

radiological findings, and assisted in patient inquiries. Patients did not evaluate the simplified reports. Negative GPT-4 performance was not detailed (**Table 3**).

Kuckelman et al. [16] selected three common radiologic examinations and procedures: CT, MRI, and bone biopsy. Ten patient questions for each type of examination or procedure were compiled (**Table 2**). This is the only study that utilized Microsoft Bing. The questions were asked on three different chatbot settings in two trials, for a total of 360 reviews. Attending radiologist and a fourth-year medical student rated the responses independently for accuracy and completeness on a 1–3 scale. They used radiologyinfo.org, an accepted online resource for comparison [17]. The Fleisch-Kincaid level of readability was also examined. Overall, 336 (93%) ratings were “entirely correct”, and 235 (65%) ratings were “complete”. No responses were rated as “inaccurate/incomplete” by either reviewer. The Fleisch-Kincaid level of readability was an eighth-grade level (**Table 3**). Patients did not evaluate the simplified reports. Negative Microsoft Bing performance included missing details about bone biopsy procedure (**Table 5**).

Jeblick et al. [18] wrote three fictitious radiology reports, for knee MRI, head MRI and whole-body CT. The reports were simplified by prompting GPT-3.5. They generated 15 different simplified reports per original report, 45 in total (**Table 2**). Many different prompt designs were tried. Radiologists evaluated the quality of the simplified reports using a 5-point Likert scale in three categories: factual correctness, completeness and potential harm. All simplified reports were factually correct and complete. For quality criteria 75% rated “Agree” (**Table 3**). Negative GPT-3.5 performance included incorrect text passages in 23 simplified reports (51%), missing relevant

information for 10 simplified reports (22%) and potentially harmful conclusions for 16 simplified reports (36%) (**Table 5**). Patients did not evaluate the simplified reports.

Scheschenja et al.[19] compiled 133 questions related to three specific interventional radiology procedures (Port implantation, percutaneous transluminal angioplasty and transcatheter arterial chemoembolization). They assessed both GPT-3.5 and GPT-4 responses. The chatbot was primed to respond to specific inquiries (**Table 2**).

Grading was performed using a 5-point Likert scale. For “completely correct” GPT-3.5 scored 30.8% while GPT-4 scored 35.3%. GPT-4 was found to give significantly more accurate responses than GPT-3 ($p = 0.043$) (**Table 3**). Negative performance included “mostly incorrect” responses in 5.3% of instances for GPT-3. For GPT-4 just 2.3%. No response was identified as potentially harmful (**Table 5**).

Gordon et al. [20] assessed GPT-3.5 for accuracy, relevance and readability in answering patient imaging-related questions. They compiled 22 imaging-related questions (**Table 2**). The categories for the questions included: safety, the radiology report, the procedure, preparation before imaging, meaning of terms and medical staff. Questions were posed to ChatGPT with and without a prompt. Four board-certified radiologists evaluated the answers for accuracy, consistency and relevance. Two patients also reviewed the responses. Readability was assessed by Flesch Kincaid Grade Level (FKGL). For accuracy GPT-3.5 scored 87% (229/264). Consistency of the responses was 86% (76/88). Nearly all responses 99% (261/264) were partially relevant for both prompt and non-prompt questions. The average FKGL was high at 13.6. When provided with a prompt, GPT-3.5 performed better in all parameters (**Table 3**). Negative GPT-4 performance was not detailed.

Schmidt et al. [21] evaluated the ability of GPT-3.5 for simplifying radiological MRI findings. They created five versions of a simplified radiological report using ChatGPT 3.5 (**Table 2**). They created different prompts until one prompt was selected for the best outcomes (**Table 4**). They asked GPT-3.5 for varying levels of complexity: simple, moderate and complex. Two orthopedic surgeons and two radiologists evaluated the reports for quality, completeness and comprehensibility using a questionnaire. All simplified reports were evaluated by 20 patients. They used a patient-specific questionnaire for comprehension and simplification. The simplified radiology reports were factually correct regardless of complexity. The majority of participants indicated “Agree” with respect to the simplicity and comprehensibility (**Table 3**). Negative performance included missing 53.8% (7/13) or incorrect 23% (3/13) information across all simplified findings. For potentially harmful conclusions the simplified reports misinterpreted crucial information 6 times. An incorrect need for therapy was indicated two times, and degeneration was interpreted as injury once (**Table 5**). In addition, patient evaluation showed that while they knew what the text was about, the majority responded that the simplified text did not inform them as well as a doctor.

Some studies presented examples for different prompts they used, as well as examples for the simplification process done by ChatGPT. We included examples from each study of simplification and prompt generation, presented in **Table 6** and **Table 7**, respectively.

Discussion

In this review we examined LLMs capability to simplify radiological reports and procedures, enhancing patient comprehension and education. All studies reviewed demonstrated LLMs capability in generating simplified, understandable radiological reports.

In the past, radiology was considered a paraclinical field [22]. Radiology reports were written for referring physicians and healthcare providers. Nowadays, radiology emerges to be more clinical and patient centered [23]. Patients can access their imaging reports, but the reports readability is still complex and incomprehensible [24].

Making medicine approachable to patients is a formidable challenge to the medical community. Doctors often use complicated medical terms that patients have trouble understanding [25]. Patients' misperception of medical jargon can lead to confusion, stress and overtreatment [26, 27]. The application of advanced AI to explain and simplify imaging reports and procedures could be a step toward accessible and approachable medicine.

LLMs benefits

One of the most limited resources for a medical doctor is time [28, 29]. Working long hours, with many tasks and responsibilities, leaves little time to address patients' inquiries and concerns [30]. AI continues to evolve, becoming more integrated in various medical applications [31]. AI performance is fast and efficient [32]. Utilizing LLMs chatbots for patient education can save time for both the overworked physician and the patient waiting for answers [33].

Another important advantage is the chatbot's ability to simplify complicated medical texts into plain language [34]. Gotlieb et al. showed that several common medical phrases are often misunderstood. The interpreted meaning is frequently the exact opposite of what is intended [35]. LLMs ability to make medical terms understandable to patients can alleviate patient concern and anxiety [36].

Lastly, ChatGPT reached over 100 million users in only 2 months [37]. This rapid adoption highlights its potential role in improving accessibility of medical information to patients seeking answers. In every study we reviewed, LLMs significantly improved clarity and simplicity of radiological reports. These models may provide the solution to the knowledge gap between patients and their medical information.

LLMs drawbacks

When patients rely on LLMs for simplifying their medical imaging reports, they also need to be certain of the medical accuracy. A known limitation for LLMs is called "hallucination". This occurs when generative AI misinterprets the given prompt, resulting in outputs that lack logical consistency. When relying on AI for accurate medical information, this phenomenon is unacceptable [38]. Also, LLMs can often misinterpret clinical findings [39]. For example, some of the studies we reviewed presented AI output that mistook benign findings as malignant. This can lead to needless patient anxiety and interfere with the physician ability to reach the correct diagnosis [40]

Another concern is the patient's medical information safety. Medical imaging reports hold private medical information, and can often be the target of cyber-attacks [41]. Finally, another consideration is the disparity between different age groups in their ability to use technology. When applying LLMs for simplifying patients' imaging reports, a certain level of technological abilities is needed. Older individuals might find it harder to apply LLMs to obtain readability for their medical reports [42, 43].

It is imperative to take into consideration those significant shortcomings and challenges. LLMs should be used with caution while utilized to simplify important medical information.

Prompt engineering

For each study we examined the process of crafting the prompts. We noticed a wide range of approaches to writing and designing prompts. Only one study [13] conducted prompt optimization, which significantly improved the LLM's outputs. This emphasizes the importance and sensitivity of prompts. Prompt-engineering may be a task that requires specific training, so that the prompt is phrased correctly and the quality of the simplified medical report is not impaired.

Limitations

Our review has several limitations. Due to heterogeneity in study design and data, a meta-analysis was not performed. Only two studies tested the simplified result with patients. Several studies used word count as representation of simplification.

However, a shorter text is not always a guarantee for simplification. This was not examined. Only one study conducted prompt optimization and evaluated its

outcomes. Two studies were at high risk of bias. One study did not present clear parameters for evaluation of bias. Additional studies will be needed to further solidify the usefulness of LLMs in simplifying radiological reports and clarifying radiologic procedures. Lastly, we limited our search to PubMed/MEDLINE. We did so due to its relevance in biomedical research. We recognize this choice narrows our review's scope. This might exclude studies from other databases, possibly limiting diverse insights.

Conclusion

Utilizing LLMs for simplifying medical imaging reports and procedures is feasible. In the majority of the studies we reviewed, LLMs demonstrated promise in their capability to generate accessible medical imaging reports. However, their use warrants cautious, critical evaluation. Awareness of LLMs limitations is needed in order to avoid misuse and harming patients' diagnosis and treatment. Currently, further research in this field is warranted. Until further advancements are achieved, AI should be used with caution when applying it to simplify medical information for patients.

Disclosure statement

The authors report there are no competing interests to declare

Additional information

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Acknowledgments

None

Disclaimer

None

References

1. Atsina, Kofi-Buaku et al. "Advanced Imaging Interpretation by Radiologists and Nonradiologist Physicians: A Training Issue." *AJR. American journal of roentgenology* vol. 214,1 (2020): W55-W61. doi:10.2214/AJR.19.21802
2. Madrigano, Renata Rodrigues et al. "Evaluation of non-radiologist physicians' knowledge on aspects related to ionizing radiation in imaging." *Radiologia brasileira* vol. 47,4 (2014): 210-6. doi:10.1590/0100-3984.2013.1840
3. Rosenkrantz, Andrew B, and Eric R Flagg. "Survey-Based Assessment of Patients' Understanding of Their Own Imaging Examinations." *Journal of the American College of Radiology : JACR* vol. 12,6 (2015): 549-55. doi:10.1016/j.jacr.2015.02.006
4. Martin-Carreras, Teresa et al. "Readability of radiology reports: implications for patient-centered care." *Clinical imaging* vol. 54 (2019): 116-120. doi:10.1016/j.clinimag.2018.12.006
5. Sneider, Michael B, and Corey D Kershaw. "The Importance of Imaging in the Assessment of Interstitial Lung Diseases." *Journal of thoracic imaging* vol. 38,Suppl 1 (2023): S2-S6. doi:10.1097/RTI.0000000000000708

6. Sotoudeh, Housman, and Masoumeh Gity. "The Role of Medical Imaging in COVID-19." *Advances in experimental medicine and biology* vol. 1318 (2021): 413-434. doi:10.1007/978-3-030-63761-3_24

7. Ballard, David H et al. "The Role of Imaging in Health Screening: Overview, Rationale of Screening, and Screening Economics." *Academic radiology* vol. 28,4 (2021): 540-547. doi:10.1016/j.acra.2020.03.038

8. Reynolds, April. "Patient-centered Care." *Radiologic technology* vol. 81,2 (2009): 133-47.

9. Nickel, Brooke et al. "Words do matter: a systematic review on how different terminology for the same condition influences management preferences." *BMJ open* vol. 7,7 e014129. 10 Jul. 2017, doi:10.1136/bmjopen-2016-014129

10. Clusmann, Jan et al. "The future landscape of large language models in medicine." *Communications medicine* vol. 3,1 141. 10 Oct. 2023, doi:10.1038/s43856-023-00370-1

11. De Angelis, Luigi et al. "ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health." *Frontiers in public health* vol. 11 1166120. 25 Apr. 2023, doi:10.3389/fpubh.2023.1166120

12. Hatherley, Joshua James. "Limits of trust in medical AI." *Journal of medical ethics* vol. 46,7 (2020): 478-481. doi:10.1136/medethics-2019-105935

13. Lyu, Qing et al. "Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential." *Visual computing for industry, biomedicine, and art* vol. 6,1 9. 18 May. 2023, doi:10.1186/s42492-023-00136-5

14. Li, Hanzhou et al. "Decoding radiology reports: Potential application of OpenAI ChatGPT to enhance patient understanding of diagnostic reports." *Clinical imaging* vol. 101 (2023): 137-141. doi:10.1016/j.clinimag.2023.06.008

15. Grewal, Harpreet et al. "Radiology Gets Chatty: The ChatGPT Saga Unfolds." *Cureus* vol. 15,6 e40135. 8 Jun. 2023, doi:10.7759/cureus.40135

16. Kuckelman, Ian J., et al. "Assessing AI-Powered Patient Education: A case study in radiology." *Academic Radiology*, 2023, <https://doi.org/10.1016/j.acra.2023.08.020>.

17. Radiology Info for Patients. Computed tomography of the abdomen and pelvis; MRI spine; bone biopsy. Retrieved from <<https://www.radiologyinfo.org/>

18. Jeblick, Katharina et al. "ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports." *European radiology*, 10.1007/s00330-023-10213-1. 5 Oct. 2023, doi:10.1007/s00330-023-10213-1
19. Scheschenja, Michael et al. "Feasibility of GPT-3 and GPT-4 for in-Depth Patient Education Prior to Interventional Radiological Procedures: A Comparative Analysis." *Cardiovascular and interventional radiology*, 10.1007/s00270-023-03563-2. 23 Oct. 2023, doi:10.1007/s00270-023-03563-2
20. Gordon, Emile B et al. "Enhancing patient communication with Chat-GPT in radiology: evaluating the efficacy and readability of answers to common imaging-related questions." *Journal of the American College of Radiology : JACR*, S1546-1440(23)00775-5. 18 Oct. 2023, doi:10.1016/j.jacr.2023.09.011
21. Schmidt, Sebastian et al. "Simplifying radiologic reports with natural language processing: a novel approach using ChatGPT in enhancing patient understanding of MRI results." *Archives of orthopaedic and trauma surgery*, 10.1007/s00402-023-05113-4. 11 Nov. 2023, doi:10.1007/s00402-023-05113-4
22. Boey, Hong Khim. "The evolution of radiology from paraclinical to clinical." *Annals of the Academy of Medicine, Singapore* vol. 38,7 (2009): 653-7.

23. Itri, Jason N. "Patient-centered Radiology." *Radiographics : a review publication of the Radiological Society of North America, Inc* vol. 35,6 (2015): 1835-46. doi:10.1148/rg.2015150110
24. Patil, Siya et al. "Radiology Reporting in the Era of Patient-Centered Care: How Can We Improve Readability?." *Journal of digital imaging* vol. 34,2 (2021): 367-373. doi:10.1007/s10278-021-00439-0
25. Gotlieb, Rachael et al. "Accuracy in Patient Understanding of Common Medical Phrases." *JAMA network open* vol. 5,11 e2242972. 1 Nov. 2022, doi:10.1001/jamanetworkopen.2022.42972
26. Chapple, A et al. "Clinical terminology: anxiety and confusion amongst families undergoing genetic counseling." *Patient education and counseling* vol. 32,1-2 (1997): 81-91. doi:10.1016/s0738-3991(97)00065-7
27. Nickel, Brooke et al. "Words do matter: a systematic review on how different terminology for the same condition influences management preferences." *BMJ open* vol. 7,7 e014129. 10 Jul. 2017, doi:10.1136/bmjopen-2016-014129
28. Prasad, Kriti et al. "Time Pressure During Primary Care Office Visits: a Prospective Evaluation of Data from the Healthy Work Place Study." *Journal of general internal medicine* vol. 35,2 (2020): 465-472. doi:10.1007/s11606-019-05343-6

29. Moura, Felipe Scipião et al. "Physicians' working time restriction and its impact on patient safety: an integrative review." *Revista brasileira de medicina do trabalho : publicacao oficial da Associacao Nacional de Medicina do Trabalho-ANAMT* vol. 16,4 482-491. 24 Apr. 2020, doi:10.5327/Z1679443520180294
30. Dugdale, D C et al. "Time and the patient-physician relationship." *Journal of general internal medicine* vol. 14 Suppl 1,Suppl 1 (1999): S34-40. doi:10.1046/j.1525-1497.1999.00263.x
31. Rajpurkar, Pranav et al. "AI in health and medicine." *Nature medicine* vol. 28,1 (2022): 31-38. doi:10.1038/s41591-021-01614-0
32. Yin, Jiamin et al. "Role of Artificial Intelligence Applications in Real-Life Clinical Practice: Systematic Review." *Journal of medical Internet research* vol. 23,4 e25759. 22 Apr. 2021, doi:10.2196/25759
33. Priolo, Manuela, and Marco Tartaglia. "The Right to Ask, the Need to Answer- When Patients Meet Research: How to Cope with Time." *International journal of environmental research and public health* vol. 20,5 4573. 4 Mar. 2023, doi:10.3390/ijerph20054573

34. Kirchner, Gregory J et al. "Can Artificial Intelligence Improve the Readability of Patient Education Materials?." *Clinical orthopaedics and related research* vol. 481,11 (2023): 2260-2267. doi:10.1097/CORR.0000000000002668
35. Gotlieb, Rachael et al. "Accuracy in Patient Understanding of Common Medical Phrases." *JAMA network open* vol. 5,11 e2242972. 1 Nov. 2022, doi:10.1001/jamanetworkopen.2022.42972
36. Mueller, P R et al. "Interventional radiologic procedures: patient anxiety, perception of pain, understanding of procedure, and satisfaction with medication--a prospective study." *Radiology* vol. 215,3 (2000): 684-8. doi:10.1148/radiology.215.3.r00jn33684
37. Mesko, Bertalan. "The ChatGPT (Generative Artificial Intelligence) Revolution Has Made Artificial Intelligence Approachable for Medical Professionals." *Journal of medical Internet research* vol. 25 e48392. 22 Jun. 2023, doi:10.2196/48392
38. Sharun, Khan et al. "ChatGPT and artificial hallucinations in stem cell research: assessing the accuracy of generated references - a preliminary study." *Annals of medicine and surgery (2012)* vol. 85,10 5275-5278. 1 Sep. 2023, doi:10.1097/MS9.0000000000001228

39. Seyyed-Kalantari, Laleh et al. "Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations." *Nature medicine* vol. 27,12 (2021): 2176-2182. doi:10.1038/s41591-021-01595-0
40. Bernstein, Michael H et al. "Can incorrect artificial intelligence (AI) results impact radiologists, and if so, what can we do about it? A multi-reader pilot study of lung cancer detection with chest radiography." *European radiology* vol. 33,11 (2023): 8263-8269. doi:10.1007/s00330-023-09747-1
41. Sorin, Vera et al. "Adversarial attacks in radiology - A systematic review." *European journal of radiology* vol. 167 (2023): 111085. doi:10.1016/j.ejrad.2023.111085
42. Köttl, Hanna et al. ""But at the age of 85? Forget it!": Internalized ageism, a barrier to technology use." *Journal of aging studies* vol. 59 (2021): 100971. doi:10.1016/j.jaging.2021.100971
43. Kunonga, Tafadzwa Patience et al. "Effects of Digital Technologies on Older People's Access to Health and Social Care: Umbrella Review." *Journal of medical Internet research* vol. 23,11 e25887. 24 Nov. 2021, doi:10.2196/25887

