



Leveraging GPT-4 enables patient comprehension of radiology reports[☆]

M.H.Elise van Driel^a, Noa Blok^a, Jan A.J.G. van den Brand^{b,c}, Davy van de Sande^{b,c}, Marianne de Vries^d, Bram Eijlers^d, Fokko Smits^d, Jacob J. Visser^d, Diederik Gommers^{b,c}, Cornelis Verhoef^a, Michel E. van Genderen^{b,c}, Dirk J. Grünhagen^{a,1}, Denise E. Hilling^{a,b,1,*}

^a Department of Surgical Oncology and Gastrointestinal Surgery, Erasmus MC Cancer Centre, University Medical Centre Rotterdam, the Netherlands

^b Erasmus MC Datahub, Erasmus University Medical Centre, Rotterdam, the Netherlands

^c Department of Adult Intensive Care, Erasmus University Medical Centre, Rotterdam, the Netherlands

^d Department of Radiology & Nuclear Medicine, Erasmus University Medical Centre, Rotterdam, the Netherlands

ARTICLE INFO

Keywords:

Artificial Intelligence
Large Language Models
Health Literacy
Patient-Centred Care

ABSTRACT

Objective: To assess the feasibility of using GPT-4 to simplify radiology reports into B1-level Dutch for enhanced patient comprehension.

Methods: This study utilised GPT-4, optimised through prompt engineering in Microsoft Azure. The researchers iteratively refined prompts to ensure accurate and comprehensive translations of radiology reports. Two radiologists assessed the simplified outputs for accuracy, completeness, and patient suitability. A third radiologist independently validated the final versions. Twelve colorectal cancer patients were recruited from two hospitals in the Netherlands. Semi-structured interviews were conducted to evaluate patients' comprehension and satisfaction with AI-generated reports.

Results: The optimised GPT-4 tool produced simplified reports with high accuracy (mean score 3.33/4). Patient comprehension improved significantly from 2.00 (original reports) to 3.28 (simplified reports) and 3.50 (summaries). Correct classification of report outcomes increased from 63.9% to 83.3%. Patient satisfaction was high (mean 8.30/10), with most preferring the long simplified report.

Conclusion: RADiANT successfully enhances patient understanding and satisfaction through automated AI-driven report simplification, offering a scalable solution for patient-centred communication in clinical practice. This tool reduces clinician workload and supports informed patient decision-making, demonstrating the potential of LLMs beyond English-based healthcare contexts.

1. Introduction

Since July 2020, Dutch hospitals have been legally required to provide patients with access to their electronic health records (EHR) [1]. Approximately 75 % of medical specialists in the Netherlands report that patients can now access an online portal to view their medical results [2]. Many patients, particularly those with limited health literacy, struggle to understand the EHR reports which are often written in complex medical language [3–5]. The improvement of health literacy is crucial for achieving favourable and equitable health outcomes [6].

Previous studies have demonstrated that patient comprehension and engagement can be enhanced by integrating interactive diagrams and

plain language explanations into radiology reports [7,8]. However, these efforts typically require radiologists to manually create simplified reports, limiting scalability. Advances in artificial intelligence (AI), such as ChatGPT's (Chat Generative Pre-trained Transformer, OpenAI) Large Language Model (LLM) GPT-4, offer a potential solution by automating this process. LLMs are a subset of generative AI designed to process natural language, allowing for the automation of tasks that require understanding and generating text. Research shows that LLM's can improve administrative burden, outperform clinicians and enhance patient comprehension of medical notes, particularly among individuals with low health literacy [9–11].

In this brief communication, we describe RADiANT (Radiology

[☆] This article is part of a special issue entitled: 'Large Language Models in Radiology' published in European Journal of Radiology.

* Corresponding author at: Erasmus MC Cancer Centre, University Medical Centre Rotterdam, Department of Surgical Oncology and Gastrointestinal Surgery Doctor Molewaterplein 40, 3015 GD Rotterdam, the Netherlands.

E-mail address: d.hilling@erasmusmc.nl (D.E. Hilling).

¹ Joint senior authorship.

Artificial Intelligence Navigation and Translation), a tool that seeks to automatically simplify radiology reports for patients by using GPT-4 to improve their understanding. This could be a crucial addition to the EHR's open patient portal because it is well known that patients tend to overestimate their true understanding of health information, resulting in incorrect understanding of reports [12]. To our knowledge this is the first study that demonstrates that EHR information can be translated by GPT-4 in another language than English, underscoring the unique contribution of the RADiANT tool in bridging this gap globally [13].

2. Methods and patients

The study was reviewed and approved by a Medical Ethics Committee (METC) [MEC-2024-0057]. All patients provided written

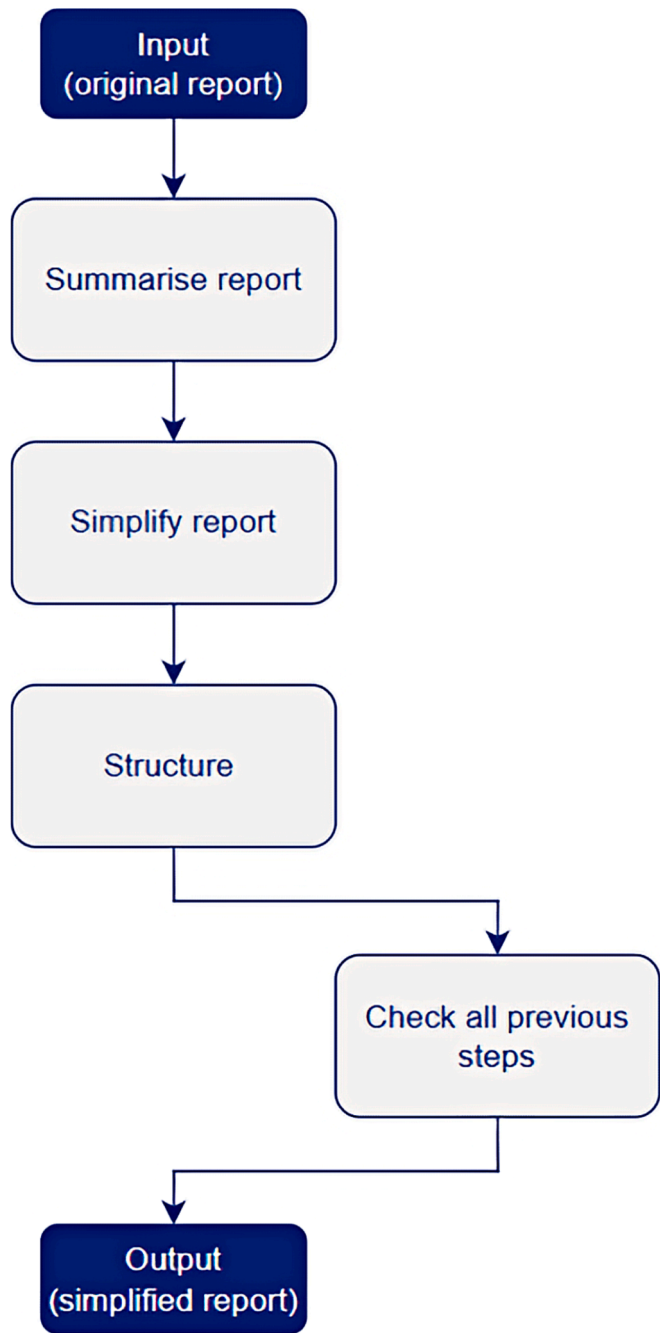


Fig. 1. Visual representation of the prompt flow.

approval for study participation.

2.1. Tool optimisation

We used the LLM GPT-4o (version 2024-05-13), developed by OpenAI which is publicly accessible, and iteratively optimised its performance in translating original radiology reports into B1 level Dutch, through structured prompt engineering in Microsoft Azure. This was guided by two main principles: providing clear and specific instructions and allowing the tool time to process information [14]. To implement the latter principle, a prompt flow was created, consisting of multiple sequential steps to ensure the tool processed each stage carefully. The prompt was run twice for every report, to ensure stability.

During the prompt optimisation process abdominal radiologists evaluated the generated reports and provided feedback. Based on their evaluations, both the prompt as well as the tool's parameters were adjusted to optimise performance. A low temperature setting of 0.1 was used to reduce variability and maintain consistency in the output. The top_p parameter, governing nucleus sampling and the range of possible words the tool considers, was fine-tuned to 0.5 based on trial-and-error testing.

2.2. Assessment by radiologists

Two abdominal radiologists, with 5 years and 12 years of experience as radiologists, participated in the assessment of different iterations of the simplified radiology reports. Each radiologist reviewed ten example sets, which included an original report from abdominal CT or MRI scans and a simplified version generated using GPT-4 with optimised prompts. The reports included positive, negative and intermediate (neither clearly positive, nor clearly negative) outcomes for scans in the context of colorectal carcinoma (CRC). Similar to the approach described by Amin et al. [15], radiologists rated each simplified report using a four-point Likert scale (1 = strongly disagree to 4 = strongly agree) based on three criteria: 1) absence of inaccurate or misleading information, 2) inclusion of all relevant/actionable information from the original report, and 3) the radiologist's comfort in sharing the simplified output with patients without supervision. Based on the feedback provided by the two radiologists, iterative refinements were made to the prompt design until an optimal version was achieved. To ensure an unbiased evaluation, a third independent radiologist, with 6 years of experience as a radiologist, subsequently assessed ten example reports to provide an objective final evaluation of the approach.

2.3. Assessment by patients

Patients with either a history of CRC or a current CRC diagnosis were recruited through convenience sampling from outpatients' clinics at two public hospitals in the Netherlands. Eligible patients had to be at least 18 years old and proficient in Dutch. Patients trained to interpret medical terminology and/or radiology reports, were excluded. The study aimed to include 12–20 patients to ensure data saturation for qualitative insights. All participants provided written informed consent. Semi-structured interviews were conducted in November of 2024, either in person or through Microsoft Teams.

At the beginning of each interview, a short assessment of patient characteristics was performed. All interviews were conducted by one researcher (N.B.). Each interview lasted approximately 45 min, and the conversations were led by a pre-defined script (Appendix A). During the interview, each patient was asked to review three sets of reports consisting of an original radiology report, an AI-generated simplified version and a short AI-generated simplified summary (consisting of 2–3 sentences). The radiology report outcomes included one negative outcome, one positive outcome and one intermediate result (neither clearly positive nor negative) presented in a random order to minimise bias.

Table 1

Assessment of the final version of the simplified reports by 2 abdominal radiologists.

Report	Statement 1			Statement 2			Statement 3		
	Radiologist 1	Radiologist 2	Mean	Radiologist 1	Radiologist 2	Mean	Radiologist 1	Radiologist 2	Mean
1	2	3	2.5	3	3	3	2	3	2.5
2	3	3	3	3	3	3	3	3	3
3	4	3.5	3.75	4	3.5	3.75	4	3	3.5
4	3	3	3	3	3	3	3	3	3
5	4	3.5	3.75	4	4	4	4	3	3.5
6	4	4	4	4	4	4	4	3	3.5
7	3	3.5	3.25	3	4	3.5	3	3	3
8	4	4	4	4	4	4	4	3	3.5
9	3	3	3	3	4	3.5	3	3	3
10	4	2	3	4	3	3.5	4	2	3

Patients were first asked to explain the content of the original report in their own words, followed by the simplified reports. Their understanding was graded on a five-point Likert scale, indicating the percentage of the report that patients understood: 1 = Understood less than 10 %, 2 = understood 10–40 %, 3 = understood 40–60 %, 4 = understood 60–90 %, 5 = understood more than 90 %.

Furthermore, patients were asked to identify whether the report outcome was positive or negative. The percentage of correctly identified outcomes was calculated. Each interview concluded with an assessment of patient satisfaction and qualitative assessment of their preferences and feedback. Patients were asked to provide qualitative feedback on the AI-generated simplified reports and to indicate their preferences regarding the formatting of the report, the long or short version.

2.4. Data analysis

Recordings of the semi-structured interviews were transcribed both manually and using Microsoft Teams [16]. Qualitative data were coded and structured using ATLAS.ti_23.1 software (ATLAS.ti Scientific Software Development GmbH, Berlin, Germany) [17]. Statements were summarised through open analysis. Quantitative analysis was performed in IBM SPSS statistics 28.0.0.0 (IBM Corp., Armonk, NY, USA) [18]. Categorical variables were presented as frequencies with percentages, and continuous outcomes were presented as means with standard deviation (SD).

3. Results

3.1. Tool optimisation

The final prompt design for GPT-4 was refined through three iterative rounds, incorporating feedback from abdominal radiologists to optimise its structure and functionality. The optimised approach comprised a structured one-shot prompt flow with four sequential steps:

Table 2

Assessment of the final version of the simplified reports by an external radiologist.

Report	Statement 1		Statement 2		Statement 3	
	External radiologist	External radiologist	External radiologist	External radiologist	External radiologist	External radiologist
1	3	3	3	3	3	3
2	3	2	3	3	3	3
3	3	3	3	3	3	3
4	3	3	3	3	3	3
5	3	4	4	4	4	4
6	4	4	4	4	4	4
7	3	4	4	4	4	4
8	3	4	4	4	4	4
9	3	3	3	3	3	3
10	3	4	4	4	4	4

summarising, simplifying and restructuring the report, and a final step of reviewing all previous actions (Fig. 1).

The simplified reports served solely for informational purposes, focusing exclusively on the radiological findings. Any details on treatment or follow-up steps based on the findings were explicitly left out of the simplified reports. A disclaimer was added to each simplified report, stating: “This simplification tool has been evaluated by a panel of medical professionals; however, errors may still occur. For questions and information about next steps, please consult your treating physician.”.

3.2. Assessment by radiologists

The final AI-generated simplified reports were evaluated by two radiologists involved in tool optimisation, yielding a mean overall score of 3.33 ($\pm$ 0.41 SD) on a four-point Likert scale (Table 1). The scores indicate that the simplified reports contained minimal misleading or incorrect information, were complete and they felt comfortable providing these reports to patients. Scores for earlier versions of the simplified reports, can be found in Appendix 2 and 3. An independent abdominal radiologist who had not been not involved in previous iterations provided a similar assessment (3.33 ± 0.44 SD) (Table 2). Again no major errors or critical omissions were identified in the simplified reports.

3.3. Assessment by patients

Twelve patients (Table 3) participated in semi-structured interviews, with data saturation achieved. Their mean age was 66.3 years ($\pm$ 9.3 SD). Patient comprehension improved after reading the simplified version of the report. Mean comprehension scores increased from 2.00

Table 3

Baseline characteristic of patients included in the study.

	Mean	Sd
Age in years	66.3	9.3
Education	Frequency	Percentage
- HBO	4	33.3 %
- University	3	25.0 %
- Other	5	41.7 %
Frequency of helping others with medical letters		
- Never	7	58.3 %
- Occasionally	3	25.0 %
- Frequently	2	16.7 %
Degree of certainty in completing own medical information		
- Reasonably	3	25.0 %
- Quite	6	50.0 %
- Very	3	25.0 %
Use of digital health application		
- Yes	11	91.7 %
- No	1	8.3 %

Table 4
Comprehension results from semi-structured interviews with patients.

	Comprehension	Standard deviation	Correct nature	Percentage
	Mean score		Frequency	
Original report	2.00	0.91	23/36	63.9 %
AI simplified version	3.28	0.87	29/36	80.6 %
Short AI simplified summary	3.50	0.90	30/36	83.3 %

Note * Grading based on a five-point Likert scale.

(± 0.91 SD) on a four-point Likert scale for original reports (indicating an understanding of 10 %-40 % of the report) to 3.28 (± 0.87 SD) and 3.50 (± 0.90 SD) for the long simplified report and the short AI-generated summary (indicating an understanding of 40 %-60 % of the simplified report), respectively. The ability to correctly classify report findings (positive or negative) increased from 63.9 % for original reports to 80.6 % and 83.3 % for the long simplified report and short summary, respectively. Notably, 80.6 % of misclassifications involved intermediate results (Table 4).

Patient satisfaction with AI-generated reports was high, averaging 8.30 (± 0.56 SD) on a 10-point scale (Table 5). Most participants (83.3 %) preferred the long simplified report. When patients had the option to access to both of the AI-generated simplified versions, 8 out of 12 (66.7 %) preferred that. None opted for the original report alone. Qualitative insights highlighted the improved readability of the simplified reports. Patients suggested including a description of follow-up steps based on the radiology findings, allowing them to understand and anticipate the next steps in their care pathway.

4. Discussion

This study demonstrates that LLM-based simplification of radiology reports improves patient understanding and satisfaction while maintaining a high accuracy and without the increased workload of manual simplification. These results highlight the potential for LLMs to enhance patient-centred communication in clinical practice in other languages than English. The simplification of complex medical language plays a crucial role in fostering greater patient literacy, which is an essential step toward empowering patients to actively engage in their care. This responsible use of AI, by providing accessible and understandable information, supports informed decision-making and enhances the overall quality of healthcare communication.

Unlike previous approaches requiring manual simplification, RADiANT automates the process, reducing clinician burden. Radiologists rated the accuracy and completeness of the generated simplified reports highly, with a mean overall score of 3.33 on a four-point Likert scale. The accuracy scores align with prior studies, such as Lyu et al., who reported a mean score of 4.27 on a five-point Likert scale for AI-generated radiology summaries [19]. The slightly lower scores in this

Table 5
Satisfaction results from semi-structured interviews with patients.

AI mark	Simplification preference	Report preference
Mean (+/- Sd)	Frequency (percentage)	Frequency (percentage)
8.30 (+/- 0.56 Sd)	Short AI 2 (16.7 %)	Original: 0 (0 %)
	Long AI 10 (83.3 %)	Short AI: 0 (0 %)
		Long AI: 4 (33.3 %)
		Both: 8 (66.7 %)

Note * Simplification preference opts for either the long or short AI-generated simplified report.

Report preference opts for the original report, the long or short AI-generated simplified report, or, both AI-generated simplifications.

study may reflect the use of a four-point scale, which eliminated neutral responses.

The simplified radiology reports generated by the optimised RADiANT tool substantially improved patient comprehension, with mean scores on a five-point Likert-scale increasing from 2.00 (patients understanding 10–40 % of the content) for original reports to 3.28 (patients understanding 40–60 % of the content). This improvement aligns with previous research emphasising the effectiveness of plain language summaries in improving accessibility to complex medical information [7]. Moreover, a recent study that utilised LLMs to translate discharge summary notes demonstrated substantial gains in both objective and subjective comprehension, particularly among patients with limited health literacy [11].

The omission of actionable follow-up steps was a notable patient concern. However, this was necessary to prevent AI from making clinical decisions, a task GPT-4 cannot reliably do yet [19]. The simplified reports are designed to complement, not replace, physician communication, as emphasised in the accompanying disclaimer. This approach ensures compliance with the Medical Device Regulation (MDR) definition of a non-medical device [20,21].

Integrating AI in clinical practice raises concerns about accountability, especially if errors occur. Studies highlight that while AI can improve standardisation, it also introduces risks related to data reliability [22]. A recent validation study on AI-assisted medical transcription found that 70 % of generated notes contained errors, including omission errors, with potential safety implications [23]. To mitigate these risks, the RADiANT tool includes a clear disclaimer advising patients to consult their healthcare provider. Furthermore, we have planned for initial close monitoring of AI-generated reports, which will later transition into random audits.

We acknowledge the debate on patient access timing for medical results. While some advocate delaying access to align with physician availability, current Dutch policies grant immediate access. Though there are pros and cons for either strategy, this discussion falls outside the scope of this study. The RADiANT tool does not alter this timeline but instead focuses solely on enhancing the interpretability of medical results. By offering both a concise and a detailed summary, the system provides multiple levels of complexity, catering to diverse patient needs. Future adaptations could extend its use to other radiology modalities, medical disciplines, and languages, since this project is both generalisable and scalable.

Challenges remain in evaluating patient comprehension. Although Likert scales provide quantifiable insights, subjective interpretation may introduce variability. However, the responses were so unequivocal that this subjectivity did not appear to significantly influence the results. The study's limited sample size and higher education level of participants may limit generalisability. Future research should include more diverse populations and larger samples. Additionally, refining standardised AI parameter tuning protocols would improve reproducibility.

5. Conclusion

The RADiANT tool demonstrates that LLMs, specifically GPT-4, can effectively and safely simplify radiology reports, significantly improving patient comprehension and satisfaction. This study is the first to show the potential of AI-driven simplification in a non-English patient population, underscoring the broader applicability of LLMs in enhancing patient understanding across diverse linguistic and cultural contexts. Future steps include EHR integration for seamless clinical workflow incorporation. Continuous validation, monitoring, and ethical oversight will be essential to ensure safe, effective, and scalable deployment across healthcare systems. By enhancing accessibility to medical information, while addressing its limitations, AI-powered solutions like RADiANT can empower patients and promote more inclusive, informed healthcare communication.

CRediT authorship contribution statement

M.H.Elise van Driel: Writing – original draft, Visualization, Project administration, Methodology, Formal analysis. **Noa Blok:** Writing – original draft, Investigation, Data curation. **Jan A.J.G. van den Brand:** Writing – review & editing, Software. **Davy van de Sande:** Writing – review & editing, Software. **Marianne de Vries:** Writing – review & editing, Investigation. **Bram Eijlers:** Writing – review & editing, Investigation. **Fokko Smits:** Writing – review & editing, Investigation. **Jacob J. Visser:** Writing – review & editing, Supervision, Resources. **Diederik Gommers:** Writing – review & editing, Supervision, Resources. **Cornelis Verhoef:** Writing – review & editing, Supervision, Resources. **Michel E. van Genderen:** Writing – review & editing, Software. **Dirk J. Grünhagen:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Denise E. Hilling:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study received no funding. The study was conducted using internal resources from Erasmus Medical Centre.

Author contributions.

D.J.G. and D.E.H. contributed to the conceptualization and methodology of the study, provided overall supervision, and served as joint senior authors. M.H.E.D. developed the study methodology, performed the formal analysis, managed project administration, and created visualisations. N.B. curated the data through patient interviews and contributed to the investigation alongside M.V., B.E., and F.S., who validated the simplified reports. J.A.J.G.B., D.S. and M.G. provided software support, including the development of the prompt and guidance on Microsoft Azure. J.J.V., D.G., and C.V. contributed resources and provided supervision. M.H.E.D. and N.B. wrote the original draft, while D.J.G., D.E.H., J.B., D.S., M.V., B.E., F.S., J.J.V., D.G., and C.V. contributed to reviewing and editing the manuscript. All authors reviewed and approved the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejrad.2025.112111>.

Data availability

Data will be made available on request.

References

- [1] Rijksoverheid, Gegevens medisch dossier inzien, aanpassen of vernietigen., 2025. <https://www.rijksoverheid.nl/wetten-en-regelingen/productbeschrijvingen/gegevens-medisch-dossier-inzien-aanpassen-of-vernietigen>. (Accessed 16 January 2025).
- [2] RIVM, E-healthmonitor 2021: Stand van zaken digitale zorg, 2021. <https://rivm.nl/documenten/e-healthmonitor-2021-stand-van-zaken-digitale-zorg>. (Accessed 16 January 2025).
- [3] M. Gerard, A. Fossa, P.H. Folcarelli, J. Walker, S.K. Bell, What Patients Value About Reading Visit Notes: A Qualitative Inquiry of Patient Experiences With Their Health Information, *J Med Internet Res* 19 (7) (2017) e237.
- [4] L. Daraz, A.S. Morrow, O.J. Ponce, W. Farah, A. Katabi, A. Majzoub, M.O. Seisa, R. Benkhadra, M. Alsawas, P. Larry, M.H. Murad, Readability of Online Health Information: A Meta-Narrative Systematic Review, *Am J Med Qual* 33 (5) (2018) 487–492.
- [5] O.E. Imoisili, E. Levinsohn, C. Pan, B.A. Howell, S. Streiter, J.R. Rosenbaum, Discrepancy Between Patient Health Literacy Levels and Readability of Patient Education Materials from an Electronic Health Record, *Health Lit Res Pract* 1 (4) (2017) e203–e207.
- [6] Low Health Literacy and Health Outcomes: An Updated Systematic Review, *Annals of Internal Medicine* 155(2) (2011) 97–107.
- [7] J. Kemp, R. Short, S. Bryant, L. Sample, N. Befera, Patient-Friendly Radiology Reporting-Implementation and Outcomes, *J Am Coll Radiol* 19(2 Pt b) (2022) 377–383.
- [8] B. Salam, D. Kravchenko, S. Nowak, A.M. Sprinkart, L. Weinhold, A. Odenthal, N. Mesrobian, L.M. Bischoff, U. Attenberger, D.L. Kuetting, J.A. Luetkens, A. Isaak, Generative Pre-trained Transformer 4 makes cardiovascular magnetic resonance reports easy to understand, *J Cardiovasc Magn Reson* 26 (1) (2024) 101035.
- [9] OpenAI, ChatGPT (version 4) [Large language model], OpenAI, 2024.
- [10] E. Goh, R. Gallo, J. Hom, E. Strong, Y. Weng, H. Kerman, J.A. Cool, Z. Kanjee, A. S. Parsons, N. Ahuja, E. Horvitz, D. Yang, A. Milstein, A.P.J. Olson, A. Rodman, J. H. Chen, Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial, *JAMA Netw. Open* 7 (10) (2024) e2440969–e.
- [11] K. Anivarya, W. Huanfei, W.M. Kelly, M. Vishala, E. Matthew, A Cross-Sectional Study of GPT-4-Based Plain Language Translation of Clinical Notes to Improve Patient Comprehension of Disease Course and Management, *Nejm Ai* 2 (2) (2025) A10a2400402.
- [12] G. Leroy, J.E. Endicott, O. Mouradi, D. Kauchak, M.L. Just, Improving perceived and actual text difficulty for health information consumers using semi-automated methods, *AMIA Annu Symp Proc* 2012 (2012) 522–531.
- [13] Vital, Vital's Doctor-to-Patient medical note translator, 2023. <https://vital.io/translate>. (Accessed 16 January 2025).
- [14] DeepLearning.AI, ChatGPT Prompt Engineering for Developers, 2022. <https://learn.deeplearning.ai/courses/chatgpt-prompt-eng/lesson/1/introduction>. (Accessed September 2024).
- [15] K.S. Amin, M.A. Davis, R. Doshi, A.H. Haims, P. Khosla, H.P. Forman, Accuracy of ChatGPT, Google Bard, and Microsoft Bing for Simplifying Radiology Reports, *Radiology* 309 (2) (2023) e232561.
- [16] Microsoft, Microsoft Teams (Version 1.6.00.376), Microsoft Corporation, 2024.
- [17] ATLAS.ti Scientific Software Development GmbH, ATLAS.ti (version 23.1), ATLAS.ti Scientific Software Development GmbH, 2024.
- [18] I.B.M. Corporation, IBM SPSS Statistics (Version 29.0). [Computer software], IBM (2024).
- [19] Q. Lyu, J. Tan, M.E. Zapadka, J. Ponnatapura, C. Niu, K.J. Myers, G. Wang, C. T. Whitlow, Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential, *Visual Comput. Ind., Biomed., Art* 6 (1) (2023) 9.
- [20] Pantón, Is uw apparaat of software een medisch hulpmiddel?, 2025. <https://cetool.nl/medisch-hulpmiddel>. (Accessed December 2024).
- [21] Pantón, Is uw apparaat of software een medisch hulpmiddel? - Quickscan, 2025. <https://cetool.nl/medisch-hulpmiddel/quickscan>. (Accessed December 2024).
- [22] E. Sacoransky, B.Y.M. Kwan, D. Soboleski, ChatGPT and assistive AI in structured radiology reporting: A systematic review, *Curr Probl Diagn Radiol* 53 (6) (2024) 728–737.
- [23] J. Biro, J.L. Handley, N.K. Cobb, V. Kottamasu, J. Collins, S. Krevat, R.M. Ratwani, Accuracy and Safety of AI-Enabled Scribe Technology: Instrument Validation Study, *J Med Internet Res* 27 (2025) e64993.