BMJ Open

Large language model-based information extraction from free-text radiology reports: a scoping review protocol

Daniel Reichenpfader [ID],[1] Henning Müller [ID],[2,3] Kerstin Denecke [ID] [1]

[1]Institute for Patient-centered Digital Health, Bern University of Applied Sciences, Bern, Switzerland
[2]Department of Radiology and Medical Informatics, Université de Genève, Genève, Switzerland
[3]Informatics Institute, HES-SO Valais-Wallis, Sierre, Switzerland

Correspondence to
Daniel Reichenpfader;
daniel.reichenpfader@bfh.ch

ABSTRACT

Introduction Radiological imaging is one of the most frequently performed diagnostic tests worldwide. The free-text contained in radiology reports is currently only rarely used for secondary use purposes, including research and predictive analysis. However, this data might be made available by means of information extraction (IE), based on natural language processing (NLP). Recently, a new approach to NLP, large language models (LLMs), has gained momentum and continues to improve performance of IE-related tasks. The objective of this scoping review is to show the state of research regarding IE from free-text radiology reports based on LLMs, to investigate applied methods and to guide future research by showing open challenges and limitations of current approaches. To our knowledge, no systematic or scoping review of IE from radiology reports based on LLMs has been published. Existing publications are outdated and do not comprise LLM-based methods.

Methods and analysis This protocol is designed based on the *JBI Manual for Evidence Synthesis*, chapter 11.2: 'Development of a scoping review protocol'. Inclusion criteria and a search strategy comprising four databases (PubMed, IEEE Xplore, Web of Science Core Collection and ACM Digital Library) are defined. Furthermore, we describe the screening process, data charting, analysis and presentation of extracted data.

Ethics and dissemination This protocol describes the methodology of a scoping literature review and does not comprise research on or with humans, animals or their data. Therefore, no ethical approval is required. After the publication of this protocol and the conduct of the review, its results are going to be published in an open access journal dedicated to biomedical informatics/digital health.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This scoping review protocol strictly adheres to standardised guidelines for scoping review conduction, including the *JBI Manual for Evidence Synthesis* and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) guideline.

⇒ The search strategy comprises four databases: PubMed, IEEE Xplore, Web of Science Core Collection and ACM Digital Library.

⇒ This scoping review will close the knowledge gap present in the field of information extraction (IE) from radiology reports caused by the recent rapid technical progress.

⇒ According to the nature of a scoping review, identified sources of evidence are not critically appraised.

⇒ The results of the scoping review will serve as a basis for defining further research directions regarding IE from radiology reports.

INTRODUCTION

Diagnostic tests, such as many types of radiological imaging, are the basis for decision-making in modern medicine[1]: For example, 74.5% of Austrian women aged 50–69 years have received bilateral mammography during the timeframe of 2 years according to the Austrian Health Interview Survey in 2019.[2] With breast cancer being the 'second most common malignancy in the world',[3] mammography shows to reduce the risk of breast cancer mortality of women aged 50–69 years with high certainty. This risk reduction is based on the treatment decisions that are in turn based on radiology reports, where experts describe the findings from the images. Traditionally, radiologists create semistructured free-text radiology reports describing findings and their interpretation based on acquired images. Structured reporting, on the other hand, aims at improving clinical outcomes and standardisation by providing frameworks for report layouts and contents. However, implementing structured reporting often requires changes to existing clinical processes. A consequent temporary increase in workload for radiologists makes it difficult to transfer structured reporting into clinical practice due to resistance among clinicians.[4] Existing information could be made available by extracting clinically relevant information including its semantics and relations by applying natural language processing (NLP)

methods. NLP is defined as the 'tract of artificial intelligence and linguistics, devoted to making computers understand the statements or words written in human languages'.[5] The extracted information could be made available for secondary use purposes, for example, for prediction or research, based on methods related to information extraction (IE).

IE is a subfield within NLP to extract relevant information from text. Subtasks of IE include among others, named entity recognition, relation extraction and template filling. To solve these subtasks, different approaches might be applied: Basic approaches are based on heuristics. Machine learning-based approaches, on the contrary, include traditional methods (eg, support vector machine, Naïve Bayes), or methods based on deep learning. Deep learning, in turn, comprises, among others, recurrent and convolutional neural networks as well as—most recently developed—large language models (LLMs).[6]

LLMs are 'deep learning models with a huge number of parameters trained in an unsupervised way on large volumes of text'.[7] We narrow this definition and only regard models with at least one million parameters as LLMs. Most of today's models are based on the transformer architecture, which was first described in 2017.[8] Since then, new LLMs have been published on an ongoing basis, being trained on growing datasets and surpassing state-of-research performance regularly. Well-known models include BERT (2018)[9], Megatron-ML (2019)[10], GPT-3 (2020)[11], GPT-4 and PaLM 2 (2023)[12 13].

Regarding existing literature concerning IE from radiology reports, several reviews are available, although these sources either miss to include current developments or only focus on a specific aspect or clinical domain. The application of NLP to radiology reports for IE has already been subject to two systematic reviews in 2016[14] and 2021.[15] While the former is not freely available, the latter searches only Google Scholar and includes only one study based on LLMs. More recent reviews include a specific scoping review on the application of NLP to reports, specifically related to breast cancer[16] and a systematic review on the application of deep learning-based NLP methods in radiology, although this only includes sources of evidence (SOE) up to 2019.[17] Recent studies have been conducted on deep learning methods, including those based on architectures such as bidirectional long short-term memory, recurrent neural network and gated recurrent unitarchitectures.[18–20]

A search in PROSPERO, conducted on 30 May 2023, with the search query 'Natural Language Processing AND radiology' yielded 12 results. Eleven results are not related to IE from radiology reports. One registered review describes named entity recognition and relation extraction in clinical documents using NLP. However, this review is neither focused on radiology reports nor LLM, and the search process was last updated on 7 July 2021, potentially missing many of the recently published articles regarding the application of LLMs.[21] Therefore, as LLMs have only recently gained momentum, a research gap exists and there is no overview of LLM-based approaches to IE from radiology reports available.

As compared with a systematic review, a scoping review usually does not include a critical appraisal of the identified SOE. On the other hand, conducting a scoping review takes fewer resources to perform and is therefore especially suitable for the dynamically changing research area focused on IE from radiology reports. With this protocol for a scoping review, we therefore intend to fill the identified research gap and answer the following research question:

What is the state of research regarding information extraction from free-text radiology reports based on LLMs?

Table 2	Inclusion criteria
C.01	The full-text SOE is retrievable.
C.02	The SOE was published after 31 December 2017.
C.03	The SOE is published in a peer-reviewed journal or conference proceeding.
C.04	The SOE describes original research, excluding reviews, comments, patents and white papers.
C.05	The SOE describes the application of NLP methods for the purpose of IE from free-text radiology reports.
C.06	The described approach is LLM-based (defined as deep learning models with more than one million parameters, trained on unlabelled text data).

IE, information extraction; LLM, large language model; NLP, natural language processing; SOE, sources of evidence.

Table 1	Research subquestions to be answered based on the scoping review	
RQ.01—Performance	What is the performance of LLMs for information extraction from radiology reports?	
RQ.02—Training and modelling	Which models are used and how is the pretraining and fine-tuning process designed?	
RQ.03—Use cases	Which modalities and anatomical regions do the analysed reports correspond to?	
RQ.04—Data and annotation	How much data was used to train the model, how was the annotation process designed and is the data publicly available?	
RQ.05—Challenges	What are open challenges and common limitations of existing approaches?	

LLMs, large language models.

Table 3	Primary search terms
PubMed	(information extraction) AND (radiolog*) AND (report*) AND (large language model)
Google Scholar	'information extraction radiology reports large language model'

Specifically, we are interested in the subquestions that arise from the posed research question, see table 1.

The objective of this scoping review protocol is to answer the above-mentioned aspects, give an overview of recent developments, identify main trends and guide future research by showing open challenges and limitations of current approaches.

METHODS AND ANALYSIS

The scoping review will adhere to the *JBI Manual for Evidence Synthesis*, chapter 11: Scoping reviews.[22] This manual in turn complies with the specifications of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA) Extension for Scoping Reviews, which provides a guideline on the design and methodology of a scoping review.[23] The completed PRISMA Protocols checklist is available as an online supplemental file.

This protocol is designed specifically based on chapter 11.2 of the *JBI Manual for Evidence Synthesis*: 'Development of a scoping review protocol'. The manual defines sections and their contents to be included in the protocol, comprising inclusion criteria, search strategy, source of evidence selection, data extraction, analysis of the evidence and presentation of results. These aspects are described in the following chapters.

Inclusion criteria

In table 2, we describe the criteria to be applied in selecting SOE. Focus was put on aligning these criteria with the title as well as the research question and subquestions of the scoping review.

Search strategy

The chosen search strategy comprises three steps: First, a limited search of at least two databases (PubMed and Google Scholar) is used to obtain a list of relevant index terms and keywords, see table 3. Next, based on this list of terms, a comprehensive and systematic search query is developed iteratively.

We include four databases to be searched using the developed query: PubMed, IEEE Xplore, Web of Science Core Collection and ACM Digital Library. The search query will be initially designed for use in PubMed and then adapted for compatibility with the other three databases, with the aim of automating this process wherever possible.[24] The full draft search strategy for this database (PubMed) is documented in an online supplemental file. Each search will include all indexed SOE from the inception of the database to 1 August 2023. Each of the four search strings, including the number of retrieved records, date coverage and date of search, will be documented using a standardised template provided by the Karolinska Institutet.[25]

As a third and last step, after the selection process, reference lists of studies that are included in the review are searched for additional SOE ('forward search'). This process might be supported by automation tools.

Source of evidence selection

The SOE selection process will be conducted by two reviewers individually. The review process is performed and managed using the software platform Rayyan.[26] Before screening, duplicate records are removed semi-automatically (manual check of automatically identified duplicates) and a pilot testing procedure is carried out to ensure agreement of both reviewers on inclusion criteria: A random sample of 25 SOE entries is selected and assessed by both reviewers. Then, decisions are compared. In case of any differences, inclusion criteria are clarified and/or adapted. Screening is started only when an agreement of >75% is achieved—otherwise, additional batches of 10 SOE entries are assessed similarly until the specified level of agreement is reached.

Next, all records, consisting of titles and abstracts, are screened by both reviewers and included if they fulfil all inclusion criteria. After completion, disagreements
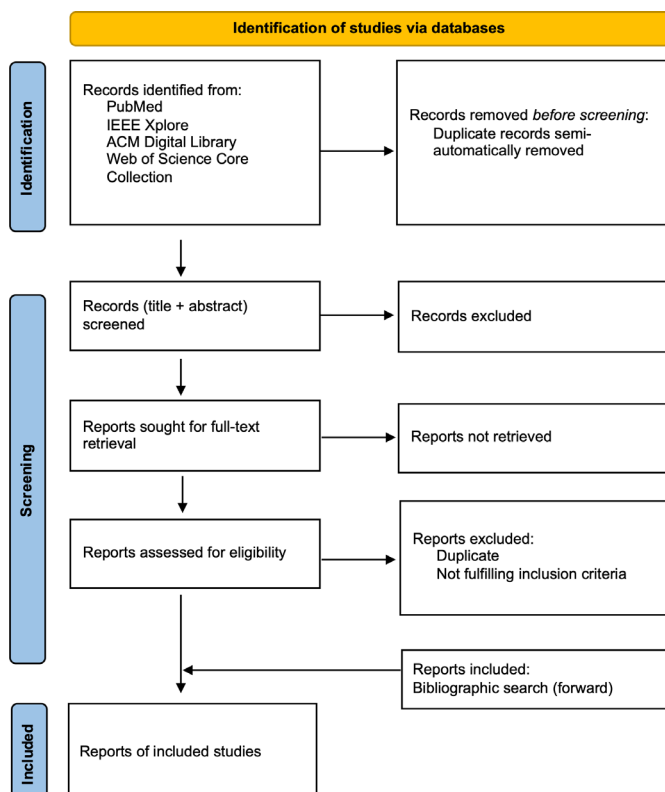
Figure 1 Source of evidence selection process.

Box 1 Data charting table

Scoping review details
⇒ Scoping review title.
⇒ Review objective.
⇒ Review question and subquestions.

Evidence source: details and characteristics
⇒ Citation details (eg, author/s, date, title, journal, volume, issue, pages).
⇒ Origin/country of origin.

Details extracted from source of evidence (according to subquestions)

Extracted information
⇒ Information model (description of entities and/or relations).
⇒ Information model development process.
⇒ Structuring of results (eg, mapping to ontology).

Model
⇒ Model design.
⇒ Pretraining and further pretraining process.
⇒ Fine-tuning process.
⇒ Described performance measures.
⇒ Baseline.

Data set
⇒ Amount.
⇒ Split training/test/validation.
⇒ Availability.
⇒ Modality.
⇒ Anatomical region.
⇒ Origin.
⇒ Language.

Annotation process
⇒ Process description.
⇒ Approach (automated, semiautomated, manual, mixed).
⇒ Number of annotators.
⇒ Annotation guideline.
⇒ Inter-annotator agreement.
⇒ Tools used.

Data availability (source code)

Open challenges

Limitations

are solved by the decision of a third reviewer. Then, full-text retrieval is performed for all included records. Records that cannot be retrieved are excluded. Retrieved full texts are assessed for eligibility: Sources that do not comply with all defined inclusion criteria are excluded. Last, a forward search is performed using reference lists of remaining SOE. See figure 1 for an illustration of the described process.

Data extraction
As a next step, key information is extracted from the final set of included studies. A charting table was created based on the *JBI Manual for Evidence Synthesis*, Appendix 11.1, and adapted as well as augmented in accordance with the research question and subquestions this scoping review addresses, see box 1.[22] Before extraction, a pilot test is conducted first to ensure the validity of the data charting table: Two SOE are extracted by two reviewers each. Results and possible adaptions of the charting table are discussed and agreed on. Upon agreement, data extraction is performed by one reviewer.

Analysis of the evidence and presentation of results
Analysis of evidence is limited to descriptive mapping and does not include synthesis or critical appraisal. Aspects described in the data charting table are described by frequency counts where possible. These frequencies provide the basis to answer the research subquestions described in table 1. The results are presented using either tables, lists, cross tabulations, bar charts, pie charts or other diagram types. Diagrams and tables are accompanied by descriptive texts.

Patient and public involvement statement
None.

ETHICS AND DISSEMINATION
This scoping review protocol does not include any research with or related to humans, animals or their data, hence no ethical approval is sought for. After the publication of the protocol, the scoping review itself is carried out. Its results are then published in an open access journal dedicated to the field of biomedical informatics. Any changes to this protocol will be documented in the outcome document, including the date of the change and a justification for it.

ORCID iDs
Daniel Reichenpfader http://orcid.org/0000-0002-8052-3359
Henning Müller http://orcid.org/0000-0001-6800-9878
Kerstin Denecke http://orcid.org/0000-0001-6691-396X

REFERENCES

1 Müskens JLJM, Kool RB, van Dulmen SA, *et al*. Overuse of diagnostic testing in healthcare: a systematic review. *BMJ Qual Saf* 2022;31:54–63.

2 Holzer M. *Österreichische Gesundheitsbefragung 2019*. 2020.

3 Canelo-Aybar C, Ferreira DS, Ballesteros M, *et al*. Benefits and harms of breast cancer Mammography screening for women at average risk of breast cancer: a systematic review for the European Commission initiative on breast cancer. *J Med Screen* 2021;28:389–404.

4 Nobel JM, van Geel K, Robben SGF. Structured reporting in radiology: a systematic review to explore its potential. *Eur Radiol* 2022;32:2837–54.

5 Khurana D, Koli A, Khatter K, *et al*. Natural language processing: state of the art, current trends and challenges. *Multimed Tools Appl* 2023;82:3713–44.

6 Practical NLP. Practical natural language processing. 2022. Available: https://github.com/practical-nlp/practical-nlp-code

7 Birhane A, Kasirzadeh A, Leslie D, *et al*. Science in the age of large language models. *Nat Rev Phys* 2023;5:277–80.

8 Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. In: *Advances in neural information processing systems*. Curran Associates, Inc, 2017.

9 Devlin J, Chang MW, Lee K, *et al*. Pre-training of deep Bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; Minneapolis, Minnesota: Association for Computational Linguistics, 2019:4171–86

10 Shoeybi M, Patwary M, Puri R, *et al*. Training multi-billion parameter language models using model parallelism. *arXiv* 2020. Available: http://arxiv.org/abs/1909.08053

11 Brown TB, Mann B, Ryder N, *et al*. Language models are few-shot learners. Proceedings of the 34th International Conference on Neural Information Processing Systems; Red Hook, NY, USA: Curran Associates Inc, 2020:1877–901

12 OpenAI. GPT-4 technical report. *arXiv* 2023. Available: http://arxiv.org/abs/2303.08774

13 Google [Internet]. Introducing palm 2. 2023. Available: https://blog.google/technology/ai/google-palm-2-ai-large-language-model/

14 Pons E, Braun LMM, Hunink MGM, *et al*. Natural language processing in radiology: a systematic review. *Radiology* 2016;279:329–43.

15 Casey A, Davidson E, Poon M, *et al*. A systematic review of natural language processing applied to radiology reports. *BMC Med Inform Decis Mak* 2021;21:179.

16 Saha A, Burns L, Kulkarni AM. A Scoping review of natural language processing of radiology reports in breast cancer. *Front Oncol* 2023;13:1160167.

17 Sorin V, Barash Y, Konen E, *et al*. Deep learning for natural language processing in radiology-fundamentals and a systematic review. *J Am Coll Radiol* 2020;17:639–48.

18 Sugimoto K, Takeda T, Oh J-H, *et al*. Extracting clinical terms from radiology reports with deep learning. *J Biomed Inform* 2021;116:103729.

19 Ong CJ, Orfanoudaki A, Zhang R, *et al*. Machine learning and natural language processing methods to identify ischemic stroke, acuity and location from radiology reports. *PLOS ONE* 2020;15:e0234908.

20 Steinkamp JM, Chambers C, Lalevic D, *et al*. Toward complete structured information extraction from radiology reports using machine learning. *J Digit Imaging* 2019;32:554–64.

21 Navarro DF, Berkovsky S, Dras M. Named entity recognition and relation extraction in clinical documents using natural language processing: a systematic review. PROSPERO 2020 CRD42020178037. 2022. Available: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020178037

22 Peters M, Godfrey C, McInerney P, *et al*. Chapter 11: Scoping reviews. In: Aromataris E, Munn Z, eds. *JBI manual for evidence synthesis*. JBI, 2020.

23 Tricco AC, Lillie E, Zarin W, *et al*. PRISMA extension for Scoping reviews (PRISMA-SCR): checklist and explanation. *Ann Intern Med* 2018;169:467–73.

24 Kibbee M. LibGuides: a guide to evidence synthesis: 6. Translate search strategies [Internet]. 2023. Available: https://guides.library.cornell.edu/evidence-synthesis/translate

25 Karolinska Institutet. University library. Presenting a search strategy. 2022. Available: https://kib.ki.se/en/search-evaluate/searching-information/presenting-search-strategy

26 Ouzzani M, Hammady H, Fedorowicz Z, *et al*. Rayyan—a web and mobile App for systematic reviews. *Syst Rev* 2016;5:210.