

# Large language models and their impact in medical imaging education

Jiajia Zhu<sup>1,2</sup> and Huanhuan Cai<sup>1,2</sup>

<sup>1</sup> Department of Radiology, The First Affiliated Hospital of Anhui Medical University, Hefei, China

<sup>2</sup> Research Center of Clinical Medical Imaging, Anhui Province, Hefei, China

## ABSTRACT

The recent emergence of large language models (LLMs), which have unprecedented capabilities to process, analyze, and synthesize complex medical information with remarkable proficiency, is poised to have a disruptive impact on health care. In the field of medical imaging, LLMs can be applied with promise in generating radiology reports along with detecting and correcting errors, explaining medical imaging findings, indicating differential diagnoses based on imaging patterns, and providing recommendations on imaging modality and protocol selection. In parallel, LLMs could offer innovative solutions for individualized learning, intelligent tutoring, content generation, and clinical decision support in medical education. However, challenges such as incorrect responses, negative influence on critical thinking, academic integrity concerns, bias, and privacy issues must be addressed to ensure safe and effective implementation of LLMs. This review summarizes the current applications, potential benefits, inherent limitations along with appropriate mitigation strategies, and future directions of LLMs in medical imaging education, emphasizing the need for responsible integration to maximize their utilities while mitigating risks.

**Subjects** Human-Computer Interaction, Computer Education, Computer Vision, Data Mining and Machine Learning

**Keywords** Large language models, Artificial intelligence, Medical imaging, Radiology, Education

## INTRODUCTION

In recent years, the emergence of large language models (LLMs) marks a promising new era in health care technology. A systematic review indicates that LLMs such as OpenAI's GPT-4 and Google's Bard models trained on rich amounts of data represent a transformative development in artificial intelligence (AI) with profound implications for medicine and medical education (*Lucas, Upperman & Robinson, 2024*). These sophisticated models, built on transformer architectures, have demonstrated unprecedented capabilities in natural language understanding and generation, allowing to process, analyze, and synthesize complex medical information with remarkable proficiency (*Zhou et al., 2023*). The ubiquity of these LLMs poses a myriad of potential clinical uses, e.g., providing individualized learning tools and aiding in decision-making.

More recent reviews suggest that the field of medical imaging may be a primary beneficiary of LLMs (*Bradshaw et al., 2025; Liu et al., 2025a*). For example, LLMs can detect and correct common radiology report errors (*Gertz et al., 2024*), explain medical imaging findings at a reading level and in a suitable way for patients (*Berigan et al., 2024*),

Submitted 28 July 2025

Accepted 10 November 2025

Published 3 December 2025

Corresponding author

Huanhuan Cai, [cjh1121@163.com](mailto:cjh1121@163.com)

Academic editor

Nicole Nogoy

Additional Information and  
Declarations can be found on  
page 11

DOI [10.7717/peerj-cs.3433](https://doi.org/10.7717/peerj-cs.3433)

© Copyright

2025 Zhu and Cai

Distributed under

Creative Commons CC-BY 4.0

**OPEN ACCESS**

and indicate differential diagnoses based on medical history and imaging patterns of patients (Kottlors et al., 2023). It is also known that LLMs can outperform medical experts in clinical text summarization and create concise clinical summaries to guide radiologists (Van Veen et al., 2024), as well as assist in the determination of radiologic study and protocol based on request forms (Gertz et al., 2023). These highly promising applications suggest that LLMs could have a considerable impact on many aspects of clinical medical imaging.

The rapid advancement of LLM technologies also come at a time of rapid evolution in medical education (Buja, 2019). In the field of medical imaging education, LLMs hold immense promise to offer innovative solutions to longstanding challenges in radiology training, learning, and practice. The traditional model of medical imaging education, which combines didactic lectures, case-based learning, and hands-on interpretation under supervision, faces increasing pressure from growing imaging volumes, curriculum demands, and the need for continuous professional development. LLMs present opportunities to improve this educational paradigm through offering individualized learning tools and intelligent tutoring systems. It is evident that LLMs like GPT-4 can achieve passing scores on medical licensing examinations and perform comparably to human experts in answering radiology board-style questions. More specialized models fine-tuned on radiology reports have shown even greater promise for domain-specific applications. Nevertheless, integrating these LLMs into medical imaging education might raise concerns about accuracy, ethical implications, and detriments to critical thinking.

This comprehensive review examines the current knowledge on LLM applications in medical imaging education, exploring their potential benefits, inherent limitations, and future directions. We analyze how these models are being implemented across various educational contexts and discuss the technical, ethical, and practical considerations surrounding their adoption. The audience it is intended for radiology trainees, medical educators, medical students, practicing radiologists, and the relevant researchers.

## LITERATURE SEARCH METHODOLOGY

We performed literature searches across four major electronic databases: PubMed (MeSH), MedLine, Google Scholar, and Web of Science. The search strategy combined keywords and Boolean operators to capture relevant studies. Key search terms included: “large language models”, “LLM”, “ChatGPT”, “GPT-4”, “artificial intelligence”, “medical imaging”, “radiology”, “education”, “medical education”, and their combinations. The search covered publications from January 2010 to June 2025, with a focus on recent advances from 2022 onward, coinciding with the widespread adoption of generative LLMs. We included original research articles, review articles, conference proceedings, and influential preprint studies (e.g., from arXiv) that discussed LLM applications in medical imaging or related educational contexts. We excluded studies that: (1) did not focus on LLMs or their applications in medical imaging or education; (2) were not available in English; (3) were purely technical reports without clinical or educational relevance; and (4) focused exclusively on non-medical domains. We adopted a narrative synthesis framework to organize and summarize findings thematically. Given the rapidly evolving

nature of the field and the diversity of study types, a narrative approach was deemed most appropriate to integrate evidence from both empirical studies and conceptual reviews. Key themes were identified iteratively, and findings were critically evaluated to balance both optimistic and cautionary perspectives.

## DEVELOPMENT OF LLMS IN MEDICAL IMAGING

The development of LLMs in medical imaging consists of two stages: self-supervised pretraining and fine-tuning.

During the pretraining stage, massive amounts of unlabeled text data are used to teach the model the rules and patterns of language, achieved *via* self-supervision. For LLMs designed for use in medical imaging, datasets used for self-supervised pretraining should include radiology reports, clinical notes, and structured data from electronic health records (Silva & Matos, 2022). Some specialized LLMs, such as RadBERT (Yan et al., 2022) and Radiology-Llama2 (Liu et al., 2023d), could outperform general-purpose LLMs in clinical radiology tasks (Zhang et al., 2024). Due to the fact that pretraining LLMs requires enormous clinical datasets and large computational resources, most academic groups have been precluded from developing their own state-of-the-art LLMs. After pretraining, a general model, which can produce coherent and relevant text, is created. Then, fine-tuning is necessary for adapting this preliminary pretrained model to specific tasks and domains.

Fine-tuning typically starts with preparing labeled data related to the target task. For example, if the objective is to build an LLM to summarize radiology findings into impressions, datasets of paired findings and impressions should be utilized. Fine-tuning also includes preference alignment, which helps make sure that the LLM is more helpful, ethical, and factual. For instance, after initial training, ChatGPT underwent reinforcement learning with human feedback. This procedure involved humans evaluating the model's responses and offering their preferences, which were utilized to create a reward function to reinforce desirable behaviors and suppress undesirable behaviors. Recent advances in few-shot and zero-shot learning have enabled LLMs to perform novel tasks with minimal examples, while techniques like retrieval-augmented generation have allowed the models to incorporate up-to-date medical knowledge from external databases.

## APPLICATIONS OF LLMS IN RADIOLOGY

In medical imaging, LLMs can be applied with promise in radiology reporting, medical record navigation, and clinical decision-making.

LLMs can automate and improve many reporting-related tasks. There has been extensive research on LLMs-based clinical text (including radiology report) summarization (Liu et al., 2021; Ma et al., 2023; Sun et al., 2023). It has been shown that LLMs can match the performance of experts in impression generation, clinical note summarization, and doctor-patient dialog summarization. This may help radiologists to generate more accurate reports, improving both efficiency and report quality. LLMs can also be leveraged to suggest differential diagnoses on the basis of radiology observations (Kottlors et al., 2023), to detect speech recognition errors in radiology reports (Schmidt et al., 2024), and to transform free-text radiology reports into structured reporting (Adams et al., 2023). While

preliminary and investigative research has yielded promising results, real-world clinical assessment is still warranted. Furthermore, personalizing reporting styles to individual hospital or radiologist (Tie et al., 2024) and integrating LLMs into clinical workflows remain challenging (Kim, 2024), such that careful model monitoring and human oversight are greatly needed.

The additional application of LLMs is to improve decision-making for clinical imaging settings, that said, they may enhance clinicians' abilities and efficiency in decision-making. For instance, LLMs can provide recommendations on imaging modality and protocol selection according to the patient's medical history and the referring clinician's questions (Gertz et al., 2023; Liu et al., 2022). While numerous studies have demonstrated that LLMs are able to achieve performance levels comparable to human experts in recommending imaging strategies for distinct clinical manifestations, some concerns still exist regarding adapting to institution-specific guidelines and handling multiphasic explanations, implying that continued validation in more clinical datasets or customized fine-tuning is necessary.

Considerable effort in the past decade has been directed towards developing LLM tools specializing for medical imaging. For example, CheXbert (Smit et al., 2020) and RadGraph (Jain et al., 2021) could automate the extraction of important clinical information from unstructured radiology reports. In recent years, more updated LLM tools have shown promising results in extracting abnormal observations (Guellec et al., 2024), delineating lesion features (Fink et al., 2023), and assessing treatment efficacy (Huemann et al., 2023), which may facilitate longitudinal studies and retrospective investigations.

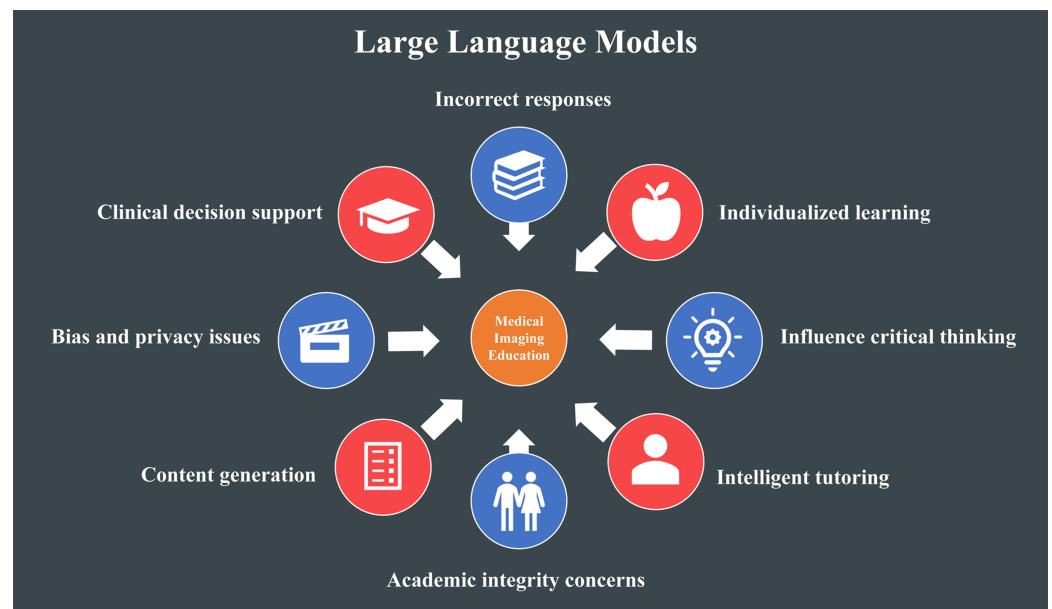
Beyond text-based applications, vision-language models (VLMs) represent a transformative advancement by directly interpreting medical images alongside textual queries. These multimodal systems enable interactive educational applications, such as visual question answering and anomaly localization, allowing trainees to engage in dynamic dialogues about findings within actual imaging studies (Liu et al., 2023a). While challenges like visual hallucinations require careful mitigation, VLMs hold significant potential for creating immersive, image-centric learning environments that closely mimic clinical reasoning workflows.

## APPLICATIONS IN MEDICAL IMAGING EDUCATION

The applications of LLMs in medical imaging education are multifactorial (Fig. 1).

### Performance on standardized examinations

For example, medical education could benefit from LLMs like ChatGPT, which have shown competence in multiple standardized clinical exams involving neurosurgery (Ali et al., 2023), ophthalmology (Antaki et al., 2023), plastic surgery (Humar et al., 2023), general surgery (Oh, Choi & Lee, 2023), clinical toxicology (Sabry Abdel-Messih & Kamel Boulos, 2023), cardiology (Skalidis et al., 2023), Medical Licensing Exam (Kung et al., 2023), and Medical Final Examination (Rosol et al., 2023), rendering them a potentially valuable reference for students and educators (Liu et al., 2023b). For medical students who need to learn a large body of medical knowledge during a short period, LLMs offer direct



**Figure 1** Applications and limitations of large language models (LLMs) in medical imaging education. The applications have a red icon and the limitations have a blue icon.

Full-size DOI: [10.7717/peerj-cs.3433/fig-1](https://doi.org/10.7717/peerj-cs.3433/fig-1)

access to medical resources, research literature, and clinical guidelines. From a high-level point of view, with immediate access to answers on how to manage classical or even complicated clinical scenarios at the fingertips of students and clinicians, it raises the question concerning what factually is the fundamental medical knowledge and understanding required for medical students before advancing to a higher educational level. While it is still not clear how the utilization of LLMs will assist medical educators, LLMs may indeed help educators create tests with improved ability to evaluate medical students' knowledge in realistic settings. Nonetheless, exam performance level alone may not always indicate the clinical knowledge of a student. Likewise, satisfactory exam performance of LLMs does not always suggest the usefulness of these AI techniques in the education of future radiologists.

### AI as interactive teaching assistants

In medical education and training, LLMs have the potential as a powerful tool to advance conventional coursework and teaching, providing medical students with realistic clinical scenarios and feedback ([Webb, 2023](#)). For radiology training, LLMs can be used to detect discrepancy between preliminary trainee reports and final attending reports, which can facilitate the identification of difficult teaching cases. Furthermore, GPT-4 has been demonstrated to yield high-quality case reports and show proficiency in radiology. From a literacy viewpoint, LLMs can assist students and educators in ensuring writing accuracy, standard style and formatting, and appropriate language and terminology, suggested by recent reviews ([Abd-Alrazaq et al., 2023](#); [Gandhi Periaysamy et al., 2023](#)). In addition, LLMs can help medical students in performing comprehensive evaluation of literature and

synthesizing the information in a readily comprehensible way ([Liu et al., 2023c](#); [Parsa & Ebrahimzadeh, 2023](#)). Some LLM functions such as gathering and organizing information can help save time for busy medical students and offer learners real-time feedback on clinical experiences.

### Personalized tutoring and chatbots

With respect to educators, LLMs are beneficial for educating radiology trainees. For instance, LLM-powered chatbots represent useful instruments for interpreting complex medical concepts ([Singhal et al., 2023](#)), simplifying radiology reports ([Amin et al., 2023](#)), and resolving questions about radiologic procedures ([Rahsepar et al., 2023](#)). Nevertheless, there have been several concerns regarding the accuracy and completeness of the provided information, with further studies greatly needed in the future. In general, LLMs facilitate teaching by offering curriculum and evaluation planning as well as support educators and their administration by allowing to allocate resources in a more effective manner ([Shorey et al., 2024](#)). LLMs provide instant feedback and support along with interactive learning experiences and scalability for educators. Moreover, by enhancing education efficiency, LLMs might help address the shortage of medical educators, particularly in resource-limited circumstances, and offer more time for educators to focus their efforts on individual student mentorship.

### Enhancing clinical reasoning and problem-solving skills

LLM systems allow medical students to examine subjects in unprecedented detail and gain insights by exploiting massive amounts of datasets to address current problems. LLMs could improve clinical reasoning and elevate problem-solving abilities by permitting students to raise questions, receive explanations, and give hypotheses regarding disorders as well as their diagnoses and treatments ([Ahn, 2023](#)). Further, LLMs can be used for medical education early in non-clinical environments, combined with present resources like flashcards and practice question banks ([Singh et al., 2023](#)). Despite the promising applications of LLMs ([Sng et al., 2023](#)), they cannot replace the in-person medical education offered by traditional classes and clinical rotations.

Collectively, LLMs have provided large valuable resources for medical imaging students by offering immediate access to medical knowledge and research literature. While LLMs' performance has varied, further specialization could potentially support their functions, including preparing exams, generating case reports, supporting clinical care assessment, and improving clinical reasoning through answering medical questions and producing hypotheses. Nevertheless, some challenges and limitations remain concerning achieving human-like dialogue and interaction in their current forms. [Table 1](#) summarizes the key applications, benefits, and limitations of LLMs across various domains of medical imaging education, providing a comprehensive overview of their current capabilities and constraints.



**Table 1** Applications, benefits, and limitations of large language models in medical imaging education.

Application domain	Specific examples	Key benefits	Main limitations
Individualized learning	<ul style="list-style-type: none"> <li>Customized learning paths</li> <li>Adaptive content delivery</li> <li>Knowledge gap identification</li> </ul>	<ul style="list-style-type: none"> <li>Personalized pacing</li> <li>Tailored to learner's level</li> <li>Continuous availability</li> </ul>	<ul style="list-style-type: none"> <li>Limited personalization depth</li> <li>May reduce peer interaction</li> <li>Requires extensive learner data</li> </ul>
Intelligent tutoring	<ul style="list-style-type: none"> <li>Real-time Q&amp;A support</li> <li>Step-by-step explanations</li> <li>Case-based guidance</li> </ul>	<ul style="list-style-type: none"> <li>Immediate feedback</li> <li>24/7 availability</li> <li>Consistent responses</li> </ul>	<ul style="list-style-type: none"> <li>Potential for incorrect answers</li> <li>Lack of contextual understanding</li> <li>Limited emotional intelligence</li> </ul>
Content generation	<ul style="list-style-type: none"> <li>Case report creation</li> <li>Examination questions</li> <li>Educational summaries</li> </ul>	<ul style="list-style-type: none"> <li>Time-saving for educators</li> <li>Diverse content types</li> <li>Standardized formatting</li> </ul>	<ul style="list-style-type: none"> <li>Quality variability</li> <li>Potential factual errors</li> <li>Requires expert validation</li> </ul>
Assessment & evaluation	<ul style="list-style-type: none"> <li>Automated scoring</li> <li>Performance analytics</li> <li>Progress tracking</li> </ul>	<ul style="list-style-type: none"> <li>Objective assessment</li> <li>Detailed feedback</li> <li>Efficient evaluation</li> </ul>	<ul style="list-style-type: none"> <li>May not capture clinical nuance</li> <li>Limited practical skill assessment</li> <li>Academic integrity concerns</li> </ul>
Academic writing support	<ul style="list-style-type: none"> <li>Literature synthesis</li> <li>Manuscript drafting</li> <li>Language polishing</li> </ul>	<ul style="list-style-type: none"> <li>Improved writing efficiency</li> <li>Standardized terminology</li> <li>Grammar correction</li> </ul>	<ul style="list-style-type: none"> <li>Plagiarism risks</li> <li>Loss of original thought</li> <li>Ethical disclosure requirements</li> </ul>

## CHALLENGES AND LIMITATIONS

Despite their potential, LLMs in medical imaging education face several significant challenges and limitations: incorrect responses, negative influence on critical thinking, academic integrity concerns, bias, and privacy issues (Fig. 1). Although these challenges could limit their usability, appropriate strategies can help mitigate these concerns.

LLMs could yield incorrect responses, referred to as hallucinations or fact fabrications. This is particularly common for applications in the field of medical imaging, where the context of specific medical language is crucial and decisions are high stakes. Thus, it is necessary to validate the accuracy of information provided by LLMs and continue to highlight practical skills. The optimization of LLMs for specific use cases could decrease incorrect responses. For example, LLMs can be connected to specific internet databases (e.g., PubMed) to ground them in up-to-date information (Min et al., 2024). In radiology, domain-specific models such as RadBERT (Yan et al., 2022) and Radiology-GPT (Liu et al., 2025b) have shown greater promise.

Experts have posited concerns regarding the negative influence of overreliance on LLMs on clinical reasoning development and the lack of contextual thinking in LLMs' responses (Abd-Alrazaq et al., 2023). In addition, there have been issues regarding redundancy and a lack of original thoughts, which might adversely impact students' reasoning skills and critical thinking (Arif, Munaf & Ul-Haque, 2023; Gunawardene & Schmuter, 2024; Huh, 2023). It has been traditionally assumed that LLMs show poor complex reasoning. Although more recent LLMs have made progress, complex multistep reasoning is greatly required for better applications in radiology and its education. For instance, during the

model fine-tuning procedure, multistep reasoning can be achieved through giving feedback on each reasoning step in a chain of thoughts rather than giving feedback solely on final answers ([Lightman et al., 2023](#)).

With regard to academic integrity concerns, it is largely known that organizational guidelines have been modified due to LLMs' ability to produce articles ([Loh, 2023](#)). Other concerns involve fabricated references and false information perpetuation ([De Angelis et al., 2023](#); [Gravel, D'Amours-Gravel & Osmanliu, 2023](#); [Perera Molligoda Arachchige, 2023](#)). Several experts even advocate halting improved LLM training in light of societal concerns. Some more tempered suggestions implicate thorough testing, supervised trainee exploration, and open discussions around responsible and appropriate use of LLMs in a recent review ([Bair & Norden, 2023](#)). Therefore, it is critical for medical imaging students to obtain a complete understanding of LLMs' appropriate and effective use as a powerful resource to address conventional challenges, as early as the period of preparing medical school application ([Hashimoto & Johnson, 2023](#); [Munaf, Ul-Haque & Arif, 2023](#)). In both medical research and education, the utilization of LLMs in article or project preparation might raise ethical issues and therefore clarity concerning appropriate methods of disclosure is warranted. Researchers should claim the extent to which LLMs are engaged in their project or article preparation, since transparency is the best approach to combating resistance to their use. Advancements in the knowledge about the LLM techniques would be improved with transparency concerning the utilization of LLMs.

Another major concern for the use of LLMs is the potential for biases in trained models ([Chowdhury, 2023](#)). In case of biased training datasets or algorithms themselves, a risk emerges that LLMs are prone to perpetuate biases present in their training data or adopted algorithms, which may result in inaccurate information delivered to medical students as well as a lack of diversity and representation in medical education ([Temsah et al., 2023](#); [You et al., 2025](#)). These biases could ultimately have adverse impacts on learning and patient care in medical imaging education. The above-described approaches used for the optimization of model performance could also be leveraged to decrease biases. For example, reducing biases in training data and fine-tuning procedures with use of reinforcement learning from human feedback are likely to be able to lower the risks.

Privacy and confidentiality constitute key considerations in the implementation of LLMs into medical imaging education because sensitive information, such as patient clinical and research data, is often shared when using LLMs. It is generally accepted that patient data are protected by the Health Insurance Portability and Accountability Act (HIPAA). The applications of LLMs in medical imaging education settings must ensure that patient information should not be breached and privacy is maintained. Some privacy-preserving LLMs, including those that are used locally, can protect personal health information from third parties ([Cai, 2023](#); [Mukherjee et al., 2023](#)). Researchers, educators, and students can access some LLMs securely through HIPAA-compliant application programming interfaces, whereby users possess the input and output datasets. Entry into sensitive patient data should be avoided and could only be considered after receiving review board approval and consulting institutional guidelines.



While LLMs are largely beneficial in medical imaging education, responsible implementation *via* continued validation, emphasizing practical skills, monitoring for ensuring accuracy, and mitigating potential risks is still critical and necessary.

## CURRENT STATE AND FUTURE OUTLOOK

It is already evident that rapid progress in LLMs has sparked excitement linked to applications in medical imaging and its education, with the adoption of LLMs in medical imaging education evoking a rich range of reactions among educators, learners, and clinicians. This review fills in a missing gap in current literature on LLMs in medical imaging education by extending beyond the primary focus on the possible misuse of LLMs in education to the widespread investigation of their transformative potential. With an increase in size and sophistication, state-of-the-art LLMs are expected to improve in performance. Creating and assessing models optimized for specific applications would aid in shaping how LLMs are employed to improve efficiency, cost-effectiveness, and outcomes in medical imaging education. However, it is notable that LLM systems are by no means a replacement for traditional education and experience, since they still fall short in several aspects of medical knowledge and application in comparison to human performance.

Educators, students, or clinicians who are prone to welcome LLMs typically recognize their potential utility to mitigate concerns in the field of medical education, such as the overload of medical curricula as well as the everchanging knowledge with growing research. The rich range of information might prevent students or clinicians from being experts in all aspects of medicine. It is largely accepted that LLMs could provide an approach to enhancing the learning processes, especially focusing on the most critical aspects needed at each certain level of medical training. Additionally, LLM systems are able to render medical knowledge more accessible since they offer more widespread access to the latest medical research information, thereby decreasing barriers and reconciling disparities in medical education. This is particularly beneficial for individuals who have relatively limited access to costly printed or online up-to-date medical resources, since they could derive updated medical knowledge readily and at minimal cost with use of LLMs.

LLMs could offer tailored explanations, practice questions and feedback that is able to provide more individualized learning approaches suitable for each student. If students obtain and grasp complex medical knowledge more quickly and efficiently in a personalized manner, the overall efficiency of the medical education process would be enhanced. Moreover, LLM systems could impose a beneficial influence on the development of clinical reasoning, critical thinking and decision-making skills, by offering a safe and interactive circumstance for students to engage them in formulating, analyzing and discussing real patient cases.

Despite the above-described applications, refusal to adopt LLMs arises from issues regarding their tendency to provide incorrect information, the ethical concerns about their utilization and reliability. In addition, there is also fear towards the utilization of LLMs resulting in a less emphasis on the development of critical thinking and decision-making skills among medical learners. Furthermore, there is apprehension regarding data privacy

and the lack of personalized education as well. It is possible that the extent of familiarity to the LLM systems would influence comfort and understanding profiles. Individuals who have more exposure to LLM approaches are more prone to understand the tools' present abilities and challenges. For LLMs, the tendency to hallucinate people or generate actually false information gives another element of anxiety among educators and students. Hallucinations are particularly problematic in the field of medical education, as accuracy is paramount for medical education. Thus, it is necessary to conduct critical evaluation and validation of LLMs' outputs. Integration of these LLM techniques in a manner that improves medical education necessitates mitigation of these concerns in a fashion that is forward-thinking and collaborative across multiple disciplines to ultimately transform educational strategies and ethical guidelines.

While educators and students have increasingly recognized LLMs' potential to contribute to the revolutionization of medical education, there is still apprehension regarding their reliability and adverse influence on learners' critical thinking skills. To address these challenges, it seems necessary to develop mechanisms for ensuring information accuracy and implementing LLMs in a manner that complements and extends, rather than replaces, conventional teaching approaches. In order to harness the profits of LLMs in medical imaging education, it is important to use them responsibly, which involves rigorous monitoring to validate the accuracy of information offered by LLMs. As accuracy is paramount in clinical care, the future generations of clinicians must be provided with more reliable tools to improve their learning. The implementation of LLMs into medical imaging education would offer opportunities to improve curricula rather than detract from it. That said, scientific and proper applications of LLMs would make students to place less efforts on memorization and building of mnemonics for test-taking, and instead to focus more on comprehension of knowledge, logical thinking, problem-solving and the application of medical concepts into clinical practice.

Critically, the utilization of LLMs must continue to ensure patient data privacy and confidentiality in health care settings. These priorities should be repeatedly highlighted when learners use these LLM techniques. Collaboration between technical experts and clinicians to align priorities and ensure that the development of these updated tools meets the actual needs of medical education is quite necessary. For the medical education community, learning experiences can be improved while concerns of the LLM technologies can be mitigated if users understand and master the factors that influence the applications of LLMs.

## CONCLUSION

Large language models (LLMs) have rapidly evolved and demonstrate significant potential to enhance performance, improve efficiency, and create value in medical imaging education. They offer innovative solutions for individualized learning, intelligent tutoring, content generation, and clinical decision support. While these applications are promising, key challenges—including ethical concerns, information inaccuracy, privacy issues, and potential impacts on critical thinking—must be responsibly addressed before widespread adoption.

Future research should focus on the following directions:

- **Multimodal integration:** Developing and evaluating vision-language models that combine image interpretation with textual analysis to better support radiology education.
- **Curriculum integration frameworks:** Designing structured approaches for embedding LLMs into existing medical imaging curricula while maintaining educational integrity.
- **Personalized learning pathways:** Creating adaptive LLM systems that respond to individual learner progress, knowledge gaps, and preferred learning styles.
- **Critical thinking and clinical reasoning:** Investigating how LLM use influences the development of diagnostic reasoning and problem-solving skills in trainees.
- **Ethical and bias mitigation strategies:** Establishing guidelines and technical solutions to reduce model bias, ensure fairness, and protect patient privacy.
- **Long-term outcome studies:** Conducting longitudinal and controlled trials to assess the real-world educational impact of LLMs on radiologist training and patient care outcomes.

As LLM technology continues to advance, collaborative efforts among educators, clinicians, and AI developers will be essential to harness its benefits while mitigating risks. With responsible implementation and ongoing research, LLMs are poised to become valuable assets in the future of medical imaging education.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was supported by the National Natural Science Foundation of China (82471952), Anhui Provincial Natural Science Foundation (2308085MH277), Scientific Research Key Project of Anhui Province Universities (2022AH051135) and the Scientific Research Foundation of Anhui Medical University (2022xkj143). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

National Natural Science Foundation of China: 82471952.

Anhui Provincial Natural Science Foundation: 2308085MH277.

Scientific Research Key Project of Anhui Province Universities: 2022AH051135.

Scientific Research Foundation of Anhui Medical University: 2022xkj143.

### Competing Interests

The authors declare that they have no competing interests.

## Author Contributions

- Jiajia Zhu performed the experiments, analyzed the data, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Huanhuan Cai conceived and designed the experiments, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

## Data Availability

The following information was supplied regarding data availability:  
This is a literature review.

## REFERENCES

- Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, Aziz S, Damseh R, Alabed Alrazak S, Sheikh J. 2023. Large language models in medical education: opportunities, challenges, and future directions. *JMIR Medical Education* 9(5):e48291 DOI 10.2196/48291.
- Adams LC, Truhn D, Busch F, Kader A, Niehues SM, Makowski MR, Bressemer KK. 2023. Leveraging GPT-4 for post hoc transformation of free-text radiology reports into structured reporting: a multilingual feasibility study. *Radiology* 307(4):3 DOI 10.1148/radiol.230725.
- Ahn S. 2023. The impending impacts of large language models on medical education. *Korean Journal of Medical Education* 35(1):103–107 DOI 10.3946/kjme.2023.253.
- Ali R, Tang OY, Connolly ID, Fridley JS, Shin JH, Zadnik Sullivan PL, Cielo D, Oyelese AA, Doberstein CE, Telfeian AE, Gokaslan ZL, Asaad WF. 2023. Performance of ChatGPT, GPT-4, and google bard on a neurosurgery oral boards preparation question bank. *Neurosurgery* 93(5):1090–1098 DOI 10.1227/neu.0000000000002551.
- Amin KS, Davis MA, Doshi R, Haims AH, Khosla P, Forman HP. 2023. Accuracy of ChatGPT, Google bard, and Microsoft bing for simplifying radiology reports. *Radiology* 309(2):6 DOI 10.1148/radiol.232561.
- Antaki F, Touma S, Milad D, El-Khoury J, Duval R. 2023. Evaluating the performance of ChatGPT in ophthalmology: an analysis of its successes and shortcomings. *Ophthalmology Science* 3(4):100324 DOI 10.1016/j.xops.2023.100324.
- Arif TB, Munaf U, Ul-Haque I. 2023. The future of medical education and research: is ChatGPT a blessing or blight in disguise? *Medical Education Online* 28(1):2181052 DOI 10.1080/10872981.2023.2181052.
- Bair H, Norden J. 2023. Large language models and their implications on medical education. *Academic Medicine* 98(8):869–870 DOI 10.1097/ACM.0000000000005265.
- Berigan K, Short R, Reisman D, McCray L, Skelly J, Jones K, Befera NT, Ali N. 2024. The Impact of large language model-generated radiology report summaries on patient comprehension: a randomized controlled trial. *Journal of the American College of Radiology* 21(12):1898–1903 DOI 10.1016/j.jacr.2024.06.018.
- Bradshaw TJ, Tie X, Warner J, Hu J, Li Q, Li X. 2025. Large language models and large multimodal models in medical imaging: a primer for physicians. *Journal of Nuclear Medicine* 66(2):173–182 DOI 10.2967/jnumed.124.268072.
- Buja LM. 2019. Medical education today: all that glitters is not gold. *BMC Medical Education* 19(1):110 DOI 10.1186/s12909-019-1535-9.
- Cai W. 2023. Feasibility and prospect of privacy-preserving large language models in radiology. *Radiology* 309(1):e232335 DOI 10.1148/radiol.232335.

- Chowdhury H. 2023.** *Sam altman has one big problem to solve before ChatGPT can generate big cash—making it ‘woke’.* New York, USA: Business Insider.
- De Angelis L, Baglivo F, Arzilli G, Privitera GP, Ferragina P, Tozzi AE, Rizzo C. 2023.** ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Frontiers in Public Health* **11**:1166120 DOI [10.3389/fpubh.2023.1166120](https://doi.org/10.3389/fpubh.2023.1166120).
- Fink MA, Bischoff A, Fink CA, Moll M, Kroschke J, Dulz L, Heuel CP, Kauczor HU, Weber TF. 2023.** Potential of ChatGPT and GPT-4 for data mining of free-text CT reports on lung cancer. *Radiology* **308**(3):9 DOI [10.1148/radiol.231362](https://doi.org/10.1148/radiol.231362).
- Gandhi Periaysamy A, Satapathy P, Neyazi A, Padhi BK. 2023.** ChatGPT: roles and boundaries of the new artificial intelligence tool in medical education and health research—correspondence. *Annals of Medicine & Surgery* **85**(4):1317–1318 DOI [10.1097/MS9.0000000000000371](https://doi.org/10.1097/MS9.0000000000000371).
- Gertz RJ, Bunck AC, Lennartz S, Dratsch T, Iuga AI, Maintz D, Kottlors J. 2023.** GPT-4 for automated determination of radiological study and protocol based on radiology request forms: a feasibility study. *Radiology* **307**(5):e230877 DOI [10.1148/radiol.230877](https://doi.org/10.1148/radiol.230877).
- Gertz RJ, Dratsch T, Bunck AC, Lennartz S, Iuga AI, Hellmich MG, Persigehl T, Pennig L, Gietzen CH, Fervers P, Maintz D, Hahnfeldt R, Kottlors J. 2024.** Potential of GPT-4 for detecting errors in radiology reports: implications for reporting accuracy. *Radiology* **311**(1):e232714 DOI [10.1148/radiol.232714](https://doi.org/10.1148/radiol.232714).
- Gravel J, D’Amours-Gravel M, Osmanliu E. 2023.** Learning to fake it: limited responses and fabricated references provided by ChatGPT for medical questions. *Mayo Clinic Proceedings: Digital Health* **1**(3):226–234 DOI [10.1016/j.mcpdig.2023.05.004](https://doi.org/10.1016/j.mcpdig.2023.05.004).
- Guellec BL, Lefèvre A, Geay C, Shorten L, Bruge C, Hacein-Bey L, Amouyel P, Pruvo JP, Kuchcinski G, Hamroun A. 2024.** Performance of an open-source large language model in extracting information from free-text radiology reports. *Radiology: Artificial Intelligence* **6**(4):w14349 DOI [10.1148/ryai.230364](https://doi.org/10.1148/ryai.230364).
- Gunawardene AN, Schmuter G. 2024.** Teaching the limitations of large language models in medical school. *Journal of Surgical Education* **81**(5):625 DOI [10.1016/j.jsurg.2024.01.008](https://doi.org/10.1016/j.jsurg.2024.01.008).
- Hashimoto DA, Johnson KB. 2023.** The use of artificial intelligence tools to prepare medical school applications. *Academic Medicine* **98**(9):978–982 DOI [10.1097/ACM.00000000000005309](https://doi.org/10.1097/ACM.00000000000005309).
- Huemann Z, Lee C, Hu J, Cho SY, Bradshaw TJ. 2023.** Domain-adapted large language models for classifying nuclear medicine reports. *Radiology: Artificial Intelligence* **5**(6):1188 DOI [10.1148/ryai.220281](https://doi.org/10.1148/ryai.220281).
- Huh S. 2023.** Can we trust AI chatbots’ answers about disease diagnosis and patient care. *Journal of the Korean Medical Association* **66**(4):218–222 DOI [10.5124/jkma.2023.66.4.218](https://doi.org/10.5124/jkma.2023.66.4.218).
- Humar P, Asaad M, Bengur FB, Nguyen V. 2023.** ChatGPT is equivalent to first-year plastic surgery residents: evaluation of ChatGPT on the plastic surgery in-service examination. *Aesthetic Surgery Journal* **43**(12):NP1085–NP1089 DOI [10.1093/asj/sjad130](https://doi.org/10.1093/asj/sjad130).
- Jain S, Agrawal A, Saporta A, Truong SQ, Duong DN, Bui T, Chambon P, Zhang Y, Lungren MP, Ng AY. 2021.** RadGraph: extracting clinical entities and relations from radiology reports (version 1.0.0). PhysioNet RRID:SCR\_007345 DOI [10.13026/hm87-5p47](https://doi.org/10.13026/hm87-5p47).
- Kim W. 2024.** Seeing the unseen: advancing generative AI research in radiology. *Radiology* **311**(2):4 DOI [10.1148/radiol.240935](https://doi.org/10.1148/radiol.240935).
- Kottlors J, Bratke G, Rauen P, Kabbasch C, Persigehl T, Schlamann M, Lennartz S. 2023.** Feasibility of differential diagnosis based on imaging patterns using a large language model. *Radiology* **308**(1):e231167 DOI [10.1148/radiol.231167](https://doi.org/10.1148/radiol.231167).
- Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon I, Elepano C, Madriaga M, Aggabao R, Diaz-Candido G, Maningo J, Tseng V. 2023.** Performance of ChatGPT on USMLE: potential

- for AI-assisted medical education using large language models. *PLOS Digital Health* 2(2): e0000198 DOI 10.1371/journal.pdig.0000198.
- Lightman H, Kosaraju V, Burda Y, Edwards H, Baker B, Lee T, Leike J, Schulman J, Sutskever I, Cobbe K. 2023. Let's verify step by step. In: *The Twelfth International Conference on Learning Representations*.
- Liu Z, Li Y, Shu P, Zhong A, Jiang H, Pan Y, Yang L, Ju C, Wu Z, Ma C. 2025b. Radiology-GPT: a large language model for radiology. *Meta-Radiology* 100153(2):100153 DOI 10.1016/j.metrad.2025.100153.
- Liu Z, Li Y, Shu P, Zhong A, Yang L, Ju C, Wu Z, Ma C, Luo J, Chen C. 2023d. Radiology-Llama2: best-in-class large language model for radiology. ArXiv DOI 10.48550/arXiv.2309.06419.
- Liu X, McDuff D, Kovacs G, Galatzer-Levy I, Sunshine J, Zhan J, Poh MZ, Liao S, Di Achille P, Patel S. 2023c. Large language models are few-shot health learners. ArXiv DOI 10.48550/arXiv.2305.15525.
- Liu F, Yang B, You C, Wu X, Ge S, Liu Z, Sun X, Yang Y, Clifton DA. 2022. Retrieve, reason, and refine: generating accurate and faithful patient instructions. ArXiv DOI 10.48550/arXiv.2210.12777.
- Liu F, You C, Wu X, Ge S, Wang S, Sun X. 2021. Auto-encoding knowledge graph for unsupervised medical report generation. ArXiv DOI 10.48550/arXiv.2111.04318.
- Liu F, Zhou H, Gu B, Zou X, Huang J, Wu J, Li Y, Chen SS, Hua Y, Zhou P. 2025a. Application of large language models in medicine. *Nature Reviews Bioengineering* 3(6):445–464 DOI 10.1038/s44222-025-00279-5.
- Liu J, Zhou P, Hua Y, Chong D, Tian Z, Liu A, Wang H, You C, Guo Z, Zhu L. 2023b. Benchmarking large language models on CMExam—a comprehensive Chinese medical exam dataset. ArXiv DOI 10.48550/arXiv.2306.03030.
- Liu F, Zhu T, Wu X, Yang B, You C, Wang W, Lu L, Liu Z, Zheng Y, Sun X, Yang Y, Clifton L, Clifton DA, LeiClifton D. 2023a. A medical multimodal large language model for future pandemics. *npj Digital Medicine* 6:226 DOI 10.1038/s41746-023-00952-2.
- Loh E. 2023. ChatGPT and generative AI chatbots: challenges and opportunities for science, medicine and medical leaders. *BMJ Lead* 8:e000797 DOI 10.1136/leader-2023-000797.
- Lucas HC, Upperman JS, Robinson JR. 2024. A systematic review of large language models and their implications in medical education. *Medical Education* 58(11):1276–1285 DOI 10.1111/medu.15402.
- Ma C, Wu Z, Wang J, Xu S, Wei Y, Liu Z, Guo L, Cai X, Zhang S, Zhang T. 2023. An iterative optimizing framework for radiology report summarization with ChatGPT. ArXiv DOI 10.48550/arXiv.2304.08448.
- Min D, Xu Z, Qi G, Huang L, You C. 2024. UniHGKR: unified instruction-aware heterogeneous knowledge retrievers. ArXiv DOI 10.48550/arXiv.2410.20163.
- Mukherjee P, Hou B, Lanfredi RB, Summers RM. 2023. Feasibility of using the privacy-preserving large language model Vicuna for labeling radiology reports. *Radiology* 309(1):e231147 DOI 10.1148/radiol.231147.
- Munaf U, Ul-Haque I, Arif TB. 2023. ChatGPT: a helpful tool for resident physicians? *Academic Medicine* 98(8):868–869 DOI 10.1097/ACM.0000000000005250.
- Oh N, Choi GS, Lee WY. 2023. ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. *Annals of Surgical Treatment and Research* 104(5):269–273 DOI 10.4174/astr.2023.104.5.269.



- Parsa A, Ebrahimzadeh MH. 2023. ChatGPT in medicine; a disruptive innovation or just one step forward? *Archives of Bone and Joint Surgery-ABJS* 11:2.
- Perera Molligoda Arachchige AS. 2023. Large language models (LLM) and ChatGPT: a medical student perspective. *European Journal of Nuclear Medicine and Molecular Imaging* 50(8):2248–2249 DOI 10.1007/s00259-023-06227-y.
- Rahsepar AA, Tavakoli N, Kim GHJ, Hassani C, Abtin F, Bedayat A. 2023. How AI responds to common lung cancer questions: ChatGPT vs Google Bard. *Radiology* 307(5):e230922 DOI 10.1148/radiol.230922.
- Rosol M, Gasior JS, Laba J, Korzeniewski K, Mlynczak M. 2023. Evaluation of the performance of GPT-3.5 and GPT-4 on the Polish Medical Final Examination. *Scientific Reports* 13(1):20512 DOI 10.1038/s41598-023-46995-z.
- Sabry Abdel-Messih M, Kamel Boulos MN. 2023. ChatGPT in clinical toxicology. *JMIR Medical Education* 9:e46876 DOI 10.2196/46876.
- Schmidt RA, Seah JCY, Cao K, Lim L, Lim W, Yeung J. 2024. Generative large language models for detection of speech recognition errors in radiology reports. *Radiology: Artificial Intelligence* 6(2):922 DOI 10.1148/ryai.230205.
- Shorey S, Mattar C, Pereira LB, Choolani M. 2024. A scoping review of ChatGPT’s role in healthcare education and research. *Nurse Education Today* 135:11 DOI 10.1016/j.nedt.2024.106121.
- Silva JF, Matos S. 2022. Modelling patient trajectories using multimodal information. *Journal of Biomedical Informatics* 134(1):104195 DOI 10.1016/j.jbi.2022.104195.
- Singh R, Reardon T, Srinivasan VM, Gottfried O, Bydon M, Lawton MT. 2023. Implications and future directions of ChatGPT utilization in neurosurgery. *Journal of Neurosurgery* 139(5):3 DOI 10.3171/2023.3.jns23555.
- Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, Scales N, Tanwani A, Cole-Lewis H, Pfohl S. 2023. Large language models encode clinical knowledge. *Nature* 620:172–180 DOI 10.1038/s41586-023-06291-2.
- Skalidis I, Cagnina A, Luangphiphat W, Mahendiran T, Muller O, Abbe E, Fournier S. 2023. ChatGPT takes on the European exam in core cardiology: an artificial intelligence success story? *European Heart Journal—Digital Health* 4(3):279–281 DOI 10.1093/ehjdh/ztad029.
- Smit A, Jain S, Rajpurkar P, Pareek A, Lungren M. 2020. Combining automatic labelers and expert annotations for accurate radiology report labeling using BERT. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Stroudsburg: ACL, 1500–1519 DOI 10.18653/v1/2020.emnlp-main.117.
- Sng GGR, Tung JYM, Lim DYZ, Bee YM. 2023. Potential and pitfalls of ChatGPT and natural-language artificial intelligence models for diabetes education. *Diabetes Care* 46:e103–e105 DOI 10.2337/dc23-0197.
- Sun Z, Ong H, Kennedy P, Tang L, Chen S, Elias J, Lucas E, Shih G, Peng Y. 2023. Evaluating GPT4 on impressions generation in radiology reports. *Radiology* 307(5):e231259 DOI 10.1148/radiol.231259.
- Temsah O, Khan SA, Chaiah Y, Senjab A, Alhasan K, Jamal A, Aljamaan F, Malki KH, Halwani R, Al-Tawfiq JA. 2023. Overview of early ChatGPT’s presence in medical literature: insights from a hybrid literature review by ChatGPT and human experts. *Cureus* 15(4):e37281 DOI 10.7759/cureus.37281.
- Tie X, Shin M, Pirasteh A, Ibrahim N, Huemann Z, Castellino SM, Kelly KM, Garrett J, Hu J, Cho SY. 2024. Personalized impression generation for PET reports using large language models. *Journal of Imaging Informatics in Medicine* 37(2):471–488 DOI 10.1007/s10278-024-00985-3.

- Van Veen D, Van Uden C, Blankemeier L, Delbrouck JB, Aali A, Bluethgen C, Pareek A, Polacin M, Reis EP, Seehofnerova A, Rohatgi N, Hosamani P, Collins W, Ahuja N, Langlotz CP, Hom J, Gatidis S, Pauly J, Chaudhari AS. 2024.** Adapted large language models can outperform medical experts in clinical text summarization. *Nature Medicine* **30**(4):1134–1142 DOI [10.1038/s41591-024-02855-5](https://doi.org/10.1038/s41591-024-02855-5).
- Webb JJ. 2023.** Proof of concept: using ChatGPT to teach emergency physicians how to break bad news. *Cureus* **15**:e38755 DOI [10.7759/cureus.38755](https://doi.org/10.7759/cureus.38755).
- Yan A, McAuley J, Lu X, Du J, Chang EY, Gentili A, Hsu CN. 2022.** RadBERT: adapting transformer-based language models to radiology. *Radiology: Artificial Intelligence* **4**(4):e210258 DOI [10.1148/ryai.210258](https://doi.org/10.1148/ryai.210258).
- You C, Dai H, Min Y, Sekhon JS, Joshi S, Duncan JS. 2025.** Uncovering memorization effect in the presence of spurious correlations. *Nature Communications* **16**(1):5424 DOI [10.1038/s41467-025-61531-5](https://doi.org/10.1038/s41467-025-61531-5).
- Zhang K, Zhou R, Adhikarla E, Yan Z, Liu Y, Yu J, Liu Z, Chen X, Davison BD, Ren H, Huang J, Chen C, Zhou Y, Fu S, Liu W, Liu T, Li X, Chen Y, He L, Zou J, Li Q, Liu H, Sun L. 2024.** A generalist vision-language foundation model for diverse biomedical tasks. *Nature Medicine* **30**(11):3129–3141 DOI [10.1038/s41591-024-03185-2](https://doi.org/10.1038/s41591-024-03185-2).
- Zhou H, Liu F, Gu B, Zou X, Huang J, Wu J, Li Y, Chen SS, Zhou P, Liu J. 2023.** A survey of large language models in medicine: progress, application, and challenge. *ArXiv* DOI [10.48550/arXiv.2311.05112](https://doi.org/10.48550/arXiv.2311.05112).