QUALITY IN SPORT
eISSN 2450-3118 · Open Access · Peer-reviewed
apcz.umk.pl/QS Nicolaus Copernicus University in Toruń

Cite as: KLAMKA, Antoni, KAWALEC, Paulina, BRONIKOWSKI, Kamil, RYSZKOWSKI, Maksymilian, FORNAL, Krystian, BĘDKOWSKA-KUŚMIEREK, Natalia, ZABOROWSKA, Sabina, PAWLAK, Paulina, WOJCIESZUK, Olaf, and STARCZEWSKI, Łukasz. Large Language Models in Medicine: Opportunities, Limitations, and Future Directions – A Scoping Review. Quality in Sport. 2026;61:73065. https://doi.org/10.12775/QS.2026.61.73065

ARTICLE TIMELINE
Received: 04.06.2026. Revised: 29.06.2026. Accepted: 03.07.2026. Published: 08.07.2026.

The journal has been awarded 20 points in the parametric evaluation by the Polish Ministry of Higher Education and Science (Annex to the announcement of 05.01.2024, No. 32553). Unique Journal Identifier: 201398. Scientific disciplines assigned: Economics and finance (Field of social sciences); Management and Quality Sciences (Field of social sciences).
Punkty Ministerialne z 2019 – aktualny rok 20 punktów. Załącznik do komunikatu Ministra Szkolnictwa Wyższego i Nauki z dnia 05.01.2024 Lp. 32553. Posiada Unikatowy Identyfikator Czasopisma: 201398. Przypisane dyscypliny naukowe: Ekonomia i finanse (Dziedzina nauk społecznych); Nauki o zarządzaniu i jakości (Dziedzina nauk społecznych).

Large Language Models in Medicine: Opportunities, Limitations, and Future Directions – A Scoping Review

Antoni Klamka

https://orcid.org/0009-0009-4278-0168

antonklamka@wp.pl

Medical University of Warsaw

Paulina Kawalec

https://orcid.org/0009-0009-5896-3444

kawalecpaulina01@gmail.com

Medical University of Warsaw

Kamil Bronikowski

https://orcid.org/0009-0000-2402-0821

kamil.bronikowski5@gmail.com

Cardinal Stefan Wyszynski University in Warsaw Faculty of Medicine Collegium Medicum

Maksymilian Ryszkowski
https://orcid.org/0009-0000-4309-6522
maksymilianryszkowski@gmail.com
Medical University of Warsaw

Krystian Fornal
https://orcid.org/0009-0002-5714-9554
krystianf@windowslive.com
Medical University of Warsaw

Natalia Będkowska-Kuśmierek
https://orcid.org/0009-0003-6385-1427
nbedkowskakusmierek@gmail.com
Medical University of Warsaw, Warsaw, Poland

Sabina Zaborowska
https://orcid.org/0009-0004-4742-5046
sab.zaborowska@gmail.com
Medical University of Warsaw, Warsaw, Poland

Paulina Pawlak
https://orcid.org/0009-0000-6102-1727
paulinaa379@gmail.com
Medical University of Warsaw, Warsaw, Poland

Olaf Wojcieszuk
https://orcid.org/0009-0000-7006-3494
olaf@wojcieszuk.pl
Medical University of Warsaw, Warsaw, Poland

Łukasz Starczewski
https://orcid.org/0009-0000-2258-4885
lukasz.starczewski7@gmail.com
Faculty of Medicine, Medical University of Warsaw, Warsaw, Poland

Corresponding Author

Antoni Klamka, antonklamka@wp.pl

Abstract

Introduction: Large language models (LLMs) have emerged as one of the most significant developments in artificial intelligence, demonstrating considerable potential across various healthcare domains. Their ability to process and generate natural language has led to growing interest in applications ranging from clinical decision support and personalized medicine to biomedical research and healthcare management. However, concerns regarding reliability, transparency, and clinical safety continue to limit their widespread implementation.

Material and Methods: A scoping review was conducted to explore the current applications, opportunities, limitations, and future directions of LLMs in medicine. A targeted literature selection approach was adopted, and 24 peer-reviewed articles indexed in PubMed and published between 2023 and 2026 were included. Data regarding study characteristics, application domains, reported benefits, limitations, and future perspectives were extracted and synthesized using a qualitative thematic approach.

Results: The reviewed studies demonstrated that LLMs are increasingly being applied in clinical decision support, diagnostic assistance, personalized healthcare, biomedical research, and multimodal medical systems. Evidence suggests that LLMs can improve information retrieval, support clinical reasoning, facilitate literature synthesis, and enhance personalized health monitoring. Emerging developments include knowledge-grounded models, retrieval-augmented generation systems, multimodal architectures, and AI agents designed for specialized medical tasks. Despite these advances, significant challenges remain, including hallucinations, limited explainability, data privacy concerns, and the lack of standardized clinical validation frameworks.

Conclusions: Current evidence indicates that LLMs have substantial potential to enhance healthcare delivery, research, and personalized medicine. However, they should currently be regarded as supportive tools rather than autonomous clinical decision-makers. Future progress will likely depend on the development of specialized, evidence-grounded, and multimodal systems, together with robust validation frameworks, regulatory oversight, and effective human–AI collaboration.

Keywords: large language models; artificial intelligence; healthcare; clinical decision support; personalized medicine; medical education; biomedical research; scoping review

Introduction

Artificial intelligence (AI) has become an increasingly important component of modern healthcare. Recent advances in natural language processing have led to the development of large language models (LLMs), such as ChatGPT, Gemini, Claude, and Med-PaLM, which can understand, generate, and analyze human language with remarkable accuracy. Their ability to process large volumes of medical information has attracted significant interest across healthcare systems.

LLMs have shown potential in a wide range of medical applications, including clinical decision support, differential diagnosis, treatment planning, medication management, and interpretation of diagnostic findings. They are also increasingly used in medical education, patient communication, health coaching, and biomedical research. By analyzing unstructured data such as clinical notes, scientific literature, and electronic health records, LLMs may improve efficiency and support healthcare professionals in everyday practice.

Despite these opportunities, several challenges limit the widespread adoption of LLMs in medicine. Key concerns include hallucinations, algorithmic bias, limited explainability, data privacy risks, and the need for rigorous clinical validation. These issues raise important ethical, legal, and patient safety considerations that must be addressed before routine implementation.

Given the rapid growth of research in this field, a synthesis of current evidence is needed. Therefore, the aim of this scoping review is to map the current applications of LLMs in medicine, identify their benefits and limitations, and discuss future directions for their responsible integration into healthcare systems.

Materials and Methods

This scoping review aimed to explore the current applications, opportunities, limitations, and future directions of large language models (LLMs) in medicine. The review was conducted in accordance with the general principles of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR). A targeted literature selection approach was adopted, and the review was based on 24 peer-reviewed articles indexed in PubMed and published between 2023 and 2026. Studies were selected according to their relevance to medical and healthcare applications of LLMs, including clinical decision support, medical education, patient communication, biomedical research, and emerging AI-driven healthcare technologies. Information regarding study characteristics, application domains, reported benefits, limitations, and future directions was extracted and analyzed. The findings were synthesized using a qualitative thematic approach and organized into six major categories:

clinical decision support and diagnosis, medical education and professional support, patient communication and personalized healthcare, biomedical research applications, emerging multimodal and agent-based systems, and limitations and future perspectives. As this review was based exclusively on previously published literature, ethical approval was not required.

AI was utilized for two specific purposes in this research. Text analysis of clinical reasoning narratives to identify linguistic patterns associated with specific logical fallacies. Assistance in refining the academic English language of the manuscript, ensuring clarity, consistency, and adherence to scientific writing standards. AI were used for additional linguistic refinement of the research manuscript, ensuring proper English grammar, style, and clarity in the presentation of results. It is important to emphasize that all AI tools were used strictly as assistive instruments under human supervision. The final interpretation of results, classification of errors, and conclusions were determined by human experts in clinical medicine and formal logic. The AI tools served primarily to enhance efficiency in data processing, pattern recognition, and linguistic refinement, rather than replacing human judgment in the analytical process.

Results

Characteristics of Included Studies

A total of 24 publications were included in this scoping review. The selected literature comprised original research articles, scoping reviews, systematic reviews, methodological studies, and narrative reviews. The included studies covered a broad range of medical and healthcare applications of large language models (LLMs), reflecting the rapidly expanding interest in generative artificial intelligence within the healthcare sector. [1-24]

The identified publications addressed several major domains, including clinical decision support, diagnostic assistance, medical education, patient communication, personalized health coaching, biomedical research, drug discovery, surgical applications, and healthcare administration. Several studies focused on the evaluation of LLM performance and reliability in clinical settings, whereas others explored emerging technologies such as retrieval-augmented generation (RAG), multimodal large language models, federated learning frameworks, and agent-based AI systems.[1-24]

The reviewed literature demonstrated increasing diversification of LLM applications, ranging from support for healthcare professionals and patients to advanced biomedical research and precision medicine initiatives. Despite the reported benefits, many studies highlighted concerns regarding hallucinations, transparency, clinical validation, data privacy, and ethical governance, emphasizing the need for further evaluation before widespread clinical implementation.[1-24]

From Information Retrieval to Clinical Reasoning

One of the most significant developments in the medical application of large language models (LLMs) is their transition from information retrieval systems to tools capable of supporting clinical reasoning. Early implementations primarily focused on answering medical questions, summarizing scientific literature, and facilitating access to health information. However, recent advances suggest that LLMs are increasingly being evaluated as decision-support systems capable of assisting healthcare professionals in diagnostic and therapeutic processes. Previous reviews have highlighted the growing interest in integrating LLMs into clinical workflows, particularly in tasks requiring interpretation of complex medical information and communication of clinical knowledge. [1, 13, 21]

Evidence from diagnostic applications indicates that human–AI collaboration may improve clinical performance in selected settings. In a usability study involving radiology residents, LLM-assisted support significantly improved the accuracy of brain MRI differential diagnosis compared with conventional information-search methods. These findings suggest that LLMs may serve as effective cognitive support tools by helping clinicians generate and refine diagnostic hypotheses, particularly in complex cases requiring extensive knowledge retrieval and pattern recognition. Nevertheless, the authors emphasized that expert supervision remains essential due to the risk of inaccurate or misleading outputs. [1, 21, 24]

Another rapidly developing area involves pharmacological decision support. Recent studies have demonstrated that specialized healthcare-oriented LLMs can integrate information from multiple evidence-based resources to support drug recommendation, dosage assessment, adverse drug reaction identification, and drug–drug interaction analysis. DrugGPT represents an example of a knowledge-grounded system designed to reduce hallucinations and improve the transparency of generated recommendations through the incorporation of curated medical knowledge bases. Such approaches highlight a broader trend toward domain-specific LLMs that prioritize reliability and traceability over general-purpose language generation. [13,17]

The literature also suggests that future clinical applications will increasingly combine LLMs with external knowledge sources, retrieval systems, and specialized reasoning frameworks. Rather than functioning as autonomous decision-makers, current evidence supports their role as assistive technologies that augment physician expertise, improve information accessibility, and facilitate clinical decision-making. Consequently, the most promising direction appears to be the development of collaborative human–AI models in which clinical judgment remains under the responsibility of healthcare professionals while LLMs provide structured, evidence-informed support. [1, 8]

Personalized and Preventive Medicine

Beyond traditional clinical applications, large language models are increasingly being explored in personalized and preventive healthcare. Recent developments suggest that LLMs can integrate patient-generated data, wearable sensor information, and behavioral patterns to provide individualized health insights and recommendations. A notable example is the Personal Health Large Language Model (PH-LLM), which demonstrated promising performance in sleep and fitness coaching, achieving results comparable to or exceeding those of human experts in selected tasks. [1, 3, 14]

The growing interest in AI-driven health coaching further highlights the potential of LLMs to support lifestyle modification, physical activity promotion, and long-term health monitoring. However, current evidence indicates substantial heterogeneity in evaluation methods and limited clinical validation, emphasizing the need for standardized assessment frameworks before widespread implementation. [1,24]

Additionally, emerging studies suggest that LLMs may contribute to the early identification of health-related conditions through the analysis of patient language and communication patterns. Such approaches may expand the role of LLMs from information delivery toward proactive and personalized health management. [24]

Large Language Models as Research Assistants

Large language models are increasingly being used to support scientific research and knowledge synthesis. Recent evidence indicates that LLMs can assist researchers throughout multiple stages of the research process, including literature searching, study screening, data extraction, evidence synthesis, and manuscript preparation. A systematic review of LLM-assisted reviews identified a growing number of studies employing GPT-based models to automate and accelerate traditionally time-consuming research tasks. [4, 5]

To improve the reliability of generated outputs, recent research has focused on retrieval-augmented generation (RAG) systems, which combine LLMs with external knowledge sources. By grounding responses in retrieved evidence, these approaches aim to reduce hallucinations and improve the accuracy of information synthesis, making them particularly relevant for medical and biomedical research. [4, 11]

Beyond literature analysis, LLMs have also been applied to knowledge extraction, knowledge graph construction, and scientific discovery. Emerging frameworks suggest that future systems may function as autonomous research assistants capable of integrating diverse sources of biomedical information and supporting complex research workflows. [4, 5, 11]

Toward Multimodal and Specialized Medical AI

Recent research suggests that the future development of medical LLMs is moving beyond general-purpose conversational systems toward specialized and multimodal architectures. Rather than relying solely on language-based reasoning, emerging models increasingly integrate external knowledge sources, domain-specific databases, and advanced retrieval mechanisms to improve reliability and clinical relevance. Knowledge-grounded systems such as DrugGPT have demonstrated the potential of combining LLMs with curated pharmacological resources, enabling more accurate and transparent clinical recommendations. [13, 17, 22]

Another important trend is the emergence of multimodal large language models capable of processing both textual and visual information. EndoChat represents an example of this approach, integrating language understanding with endoscopic image analysis to support surgical scene interpretation, training, and decision-making. Such systems illustrate how future medical AI may combine multiple data modalities to provide more comprehensive clinical support. [20, 22]

In parallel, recent studies have highlighted the development of LLM agents and retrieval-augmented architectures that can perform complex, multi-step tasks while accessing external knowledge sources. These approaches may enable more autonomous and context-aware healthcare applications, particularly in biomedical research, precision medicine, and clinical decision support.[13, 17, 20, 22]

Current Limitations and Barriers

Despite their growing capabilities, large language models face several important limitations that currently restrict their widespread adoption in clinical practice. One of the most frequently reported concerns is the generation of hallucinations, where models produce plausible but factually incorrect information. In healthcare settings, such inaccuracies may compromise patient safety and reduce trust in AI-assisted decision-making.

Another major challenge involves transparency and explainability. Many LLMs function as “black-box” systems, making it difficult for clinicians to verify the reasoning behind generated

recommendations. Recent approaches, including retrieval-augmented generation (RAG), attempt to improve reliability by grounding responses in external evidence sources; however, their integration into routine clinical workflows remains limited.
Data privacy, regulatory compliance, and ethical considerations represent additional barriers to implementation. Healthcare applications frequently involve sensitive patient information, creating challenges related to data security, governance, and accountability. Furthermore, the lack of standardized clinical validation frameworks makes it difficult to determine whether LLM-generated recommendations are sufficiently safe and effective for real-world medical use. [1, 11, 13, 17, 18]

Overall, current evidence suggests that while LLMs offer considerable potential, further improvements in reliability, transparency, validation, and regulatory oversight are required before their routine integration into healthcare systems can be achieved. [1, 11, 13, 17, 18]

Discussion

This scoping review highlights the rapidly expanding role of large language models (LLMs) across multiple domains of healthcare, including clinical decision support, personalized medicine, biomedical research, and multimodal medical applications. The reviewed literature suggests that LLMs have evolved beyond simple conversational systems and are increasingly being integrated into workflows that require information synthesis, reasoning, and decision support.

One of the most important findings of this review is that the greatest value of LLMs appears to lie in augmentation rather than replacement of healthcare professionals. Studies evaluating clinical applications consistently demonstrate improved performance when LLMs are used in collaboration with clinicians rather than as autonomous decision-makers. This observation was particularly evident in diagnostic and pharmacological applications, where human oversight remains essential for ensuring patient safety and appropriate interpretation of AI-generated outputs.

Another important trend identified in the reviewed literature is the transition from general-purpose language models toward specialized and knowledge-grounded systems. Emerging approaches such as retrieval-augmented generation (RAG), domain-specific medical models, and multimodal architectures seek to address the limitations of conventional LLMs by incorporating external knowledge sources and additional data modalities. These developments may improve reliability, transparency, and clinical usefulness, particularly in complex healthcare environments.

The findings also suggest that LLMs may play an increasingly important role in preventive and personalized medicine. The integration of wearable device data, patient-generated health information, and continuous monitoring systems has the potential to support individualized healthcare strategies that extend beyond traditional episodic clinical encounters. Such applications may become particularly relevant in chronic disease management, lifestyle medicine, and population health interventions.

Despite these promising developments, several barriers continue to limit the widespread implementation of LLMs in healthcare. Hallucinations, insufficient explainability, data privacy concerns, regulatory uncertainty, and the lack of standardized validation frameworks remain significant challenges. The reviewed studies consistently emphasize that current LLMs cannot yet be considered fully reliable for independent clinical decision-making. Consequently, robust evaluation standards and evidence-based implementation strategies will be essential before broader adoption can occur.

Taken together, the available evidence suggests that the future of medical LLMs will likely involve collaborative human–AI systems that combine the efficiency of artificial intelligence with the expertise and accountability of healthcare professionals. Rather than replacing clinicians, LLMs are more likely to function as intelligent assistants capable of enhancing decision-making, improving access to medical knowledge, and supporting more personalized healthcare delivery.

Conclusions

Large language models are rapidly transforming multiple areas of medicine, including clinical decision support, personalized healthcare, biomedical research, and medical knowledge management. The reviewed literature demonstrates that LLMs can improve information access, support clinical reasoning, facilitate research activities, and enable more individualized healthcare approaches.

Despite their considerable potential, important challenges remain, including hallucinations, limited explainability, data privacy concerns, and the need for rigorous clinical validation. Current evidence suggests that LLMs should be viewed as supportive tools rather than autonomous clinical decision-makers.

Future developments are likely to focus on specialized, knowledge-grounded, and multimodal systems that integrate external evidence sources and diverse healthcare data. As these technologies continue to evolve, their successful implementation will depend on robust evaluation frameworks, appropriate regulatory oversight, and effective human–AI collaboration. Overall, LLMs represent a promising advancement in modern medicine, with the potential to enhance healthcare delivery while maintaining the central role of healthcare professionals in clinical decision-making.

Disclosure

Author Contributions

Conceptualization: Antoni Klamka, Paulina Kawalec

Methodology: Antoni Klamka, Maksymilian Ryszkowski

Software: Paulina Kawalec, Natalia Będkowska-Kuśmierek

Validation: Kamil Bronikowski, Maksymilian Ryszkowski

Formal Analysis: Antoni Klamka, Kamil Bronikowski

Investigation: Krystian Fornal, Natalia Będkowska-Kuśmierek

Resource: Sabina Zaborowska, Paulina Pawlak

Data Curation: Krystian Fornal, Olaf Wojcieszuk

Writing – Original Draft Preparation: Antoni Klamka

Writing – Review and Editing: Sabina Zaborowska, Łukasz Starczewski

Visualization: Paulina Pawlak, Olaf Wojcieszuk

Supervision: Antoni Klamka

Project Administration: Antoni Klamka, Łukasz Starczewski

All authors have read and agreed to the published version of the manuscript.

Funding Statement: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sina Shool, Sara Adimi, Reza Saboori Amleshi, Ehsan Bitaraf, Reza Golpira & Mahmood Tara “A systematic review of large language model (LLM) evaluations in clinical medicine” PMCID: PMC11889796 DOI: 10.1186/s12911-025-02954-4
2. Florian Leiser, Richard Guse, Ali Sunyaev “Large Language Model Architectures in Health Care: Scoping Review of Research Perspectives” PMCID: PMC12226782 DOI: 10.2196/70315
3. Justin Khasentino, Anastasiya Belyaeva, Xin Liu, Zhun Yang, Nicholas A Furlotte, Chace Lee, Erik Schenck, Yojan Patel, Jian Cui, Logan Douglas Schneider, Robby Bryant, Ryan G Gomes, Allen Jiang, Roy Lee, Yun Liu, Javier Perez, Jameson K Rogers, Cathy Speed, Shyam Tailor, Megan Walker, Jeffrey Yu, Tim Althoff, Conor Heneghan, John Hernandez, Mark Malhotra, Leor Stern, Yossi Matias, Greg S Corrado, Shwetak Patel, Shravya Shetty, Jiening Zhan, Shruthi Prabhakara, Daniel McDuff, Cory Y McLean “A personal health large language model for sleep and fitness coaching” PMCID: PMC12532688 DOI: 10.1038/s41591-025-03888-0
4. Dmitry Scherbakov, PhD , Nina Hubig, PhD , Vinita Jansari, PhD , Alexander Bakumenko, MSc , Leslie A Lenert, MD, MS “The emergence of large language models as tools in literature reviews: a large language model-assisted systematic review”
   PMCID: PMC12089777 DOI: 10.1093/jamia/ocaf063
5. Wenyu Li , Zhitao Mao, Zhengyang Xiao, Xiaoping Liao, Mattheos Koffas, Yixin Chen, Hongwu Ma, Yinjie J Tang “Large language model for knowledge synthesis and AI-enhanced biomanufacturing” PMID: 40050157 DOI: 10.1016/j.tibtech.2025.02.008
6. Haw Hwai, Yi-Ju Ho, Chih-Hung Wang, Chien-Hua Huang “Large language model application in emergency medicine and critical care” PMID: 39198112 DOI: 10.1016/j.jfma.2024.08.032
7. Yuan Gao, Lihao Fu, Tong Si “Large language model agents as experimental orchestrators in data-driven enzyme engineering” PMID: 41175605 DOI: 10.1016/j.copbio.2025.103376
8. Pawel Pratyush, Suresh Pokharel, Stefan Schulze, Lisa Bramer, Robert H. Newman & Dukka B. KC “Large Language Model (LLM)-Based Advances in Prediction of Post-

translational Modification Sites in Proteins” PMID: 40601266 DOI: 10.1007/978-1-0716-4623-6_19

9. Shan Gao, Kaixian Yu, Yue Yang, Sheng Yu, Chenglong Shi, Xueqin Wang, Niansheng Tang, Hongtu Zhu “Large language model powered knowledge graph construction for mental health exploration” PMCID: PMC12350947 DOI: 10.1038/s41467-025-62781-z

10. Yijun Wang, Wuping Tan, Siyi Cheng, Chen Peng, Jin Peng, Fanglin Qin, Long Tang, Tongjian Zhu, Bing Wu, Jinjun Liu, Jun Wang “Large Language Model Agent for Managing Patients With Suspected Hypertension” PMID: 41064862 DOI: 10.1161/HYPERTENSIONAHA.125.25305

11. Yiqun Miao, Yuhan Zhao, Yuan Luo, Huiying Wang, Ying Wu “Improving Large Language Model Applications in the Medical and Nursing Domains With Retrieval-Augmented Generation: Scoping Review” PMCID: PMC12587015 DOI: 10.2196/80557

12. Zhihong Zhang, Mohamad Javad Momeni Nezhad, Seyed Mohammad Bagher Hosseini, Ali Zolnour, Zahra Zonour, Seyedeh Mahdis Hosseini, Maxim Topaz, Maryam Zolnoori “A Scoping Review of Large Language Model Applications in Healthcare” PMID: 40776319 DOI: 10.3233/SHTI251302

13. David Guo, Kim-Kwang Raymond Choo “Applications of Federated Large Language Model for Adverse Drug Reactions Prediction: Scoping Review” PMCID: PMC12516295 DOI: 10.2196/68291

14. Xiangxun Lai, Yue Lai, JiaCheng Chen, Shengqi Huang, Qi Gao, Caihua Huang “Evaluation Strategies for Large Language Model-Based Models in Exercise and Health Coaching: Scoping Review” PMCID: PMC12520646 DOI: 10.2196/79217

15. Takamitsu Watanabe, Katsuma Inoue, Yasuo Kuniyoshi, Kohei Nakajima, Kazuyuki Aihara “Comparison of Large Language Model with Aphasia” PMCID: PMC12165151 DOI: 10.1002/advs.202414016

16. Tao Chen, Hongyi Mo, Tao Wang, Chenlu Jiang, Zihan Liu, Fengzhen Hou, Jue Gan “A lightweight large language model for regulatory affairs translation in pharmaceutical industry” PMCID: PMC12575748 DOI: 10.1038/s41598-025-21867-w

17. Hongjian Zhou, Fenglin Liu, Jinge Wu, Wenjun Zhang, Guowei Huang, Lei Clifton, David Eyre, Haochen Luo, Fengyuan Liu, Kim Branson, Patrick Schwab, Xian Wu,

Yefeng Zheng, Anshul Thakur & David A. Clifton "A collaborative large language model for drug analysis" PMCID: PMC13190338 DOI: 10.1038/s41551-025-01471-z

18. Haelynn Gim, Benjamin Cook, Jasmin Le, Brandon Stretton, Christina Gao, Aashray Gupta, Joshua Kovoor, Christina Guo, Matthew Arnold, Galina Gheihman, Stephen Bacchi "Large language model-supported interactive case-based learning: a pilot study" PMID: 40125598 DOI: 10.1111/imj.70030

19. Koichiro Yasaka, Naoaki Nishimura, Takahiro Fukushima, Takatoshi Kubo, Shigeru Kiryu, Osamu Abe "Fine-tuned large language model for classifying CT-guided interventional radiology reports" PMID: 40551586 DOI: 10.1177/02841851251349495

20. Guankun Wang, Long Bai, Junyi Wang, Kun Yuan, Zhen Li, Tianxu Jiang, Xiting He, Jinlin Wu, Zhen Chen, Zhen Lei, Hongbin Liu, Jiazheng Wang , Fan Zhang, Nicolas Padoy, Nassir Navab, Hongliang Ren "EndoChat: Grounded multimodal large language model for endoscopic surgery" PMID: 40929920 DOI: 10.1016/j.media.2025.103789

21. Su Hwan Kim, Jonas Wihl, Severin Schramm, Cornelius Berberich, Enrike Rosenkranz, Lena Schmitzer, Kerem Serguen, Christopher Klenk, Nicolas Lenhart, Claus Zimmer, Benedikt Wiestler & Dennis M. Hedderich "Human-AI collaboration in large language model-assisted brain MRI differential diagnosis: a usability study" PMCID: PMC12350562 DOI: 10.1007/s00330-025-11484-6

22. Tiantian Yang , Yihang Xiao , Zhijie Bao , Jianye Hao , Jiajie Peng "The rise and potential opportunities of large language model agents in bioinformatics and biomedicine" PMCID: PMC12602188 DOI: 10.1093/bib/bbaf601

23. Weixin Liang, Yaohui Zhang, Zhengxuan Wu, Haley Lepp, Wenlong Ji, Xuandong Zhao, Hancheng Cao, Sheng Liu, Siyu He, Zhi Huang, Diyi Yang, Christopher Potts, Christopher D. Manning & James Zou "Quantifying large language model usage in scientific papers" PMID: 40760036 DOI: 10.1038/s41562-025-02273-8

24. Sofia Gaiduchenko, Eiji Aramaki, Seiji Shimidzu, Takako Fujimaki, Kyoko Kawabata, Shoko Wakamiya, Mizuki Shimadzu Kato, Isao Miyashiro "Large Language Model Challenges to Detect Cancer-Related Cognitive Impairment from Patient Short Speech" PMID: 40776231 DOI: 10.3233/SHTI251214