Journal of Multi Disciplinary Engineering Technologies
Volume 18, Issue 02

The Role of Large Language Models in Healthcare: Evolution, Applications, and Ethical Considerations

Gargi Mishra[1], Nupur Chugh[1], Rakhi Joon[1], Supriya Bajpai[2*]

[1]Department of Computer Science and Engineering, Bharati Vidyapeeth's College of Engineering, New Delhi, Delhi, 110063, India.
[2]IITB Monash Research Academy, IIT Bombay, Mumbai, Maharashtra, 400076, India.

*Corresponding author(s). E-mail(s): supriyamishra39@gmail.com;
Contributing authors: gargi.mishra.eee@gmail.com;
nupur.chugh@bharatividyapeeth.edu; rakhi.cse@bharatividyapeeth.edu;

Abstract

Medical and other fields experience transformative impacts because of Large Language Models (LLMs) which make natural language understanding as well as generation advanced. The analysis explores Pretrained Language Models (PLMs) development into LLMs alongside their healthcare applications and ethical concerns from implementation. The paper examines medical-specific LLMs and their operational effectiveness across healthcare procedures before discussing future hurdles. The widespread use of LLMs depends on maintaining both patient safety through fair clinical decision support and maintaining data protection standards during their applications in patient communication and medical research.

Keywords: Large Language Models, Transformer Architecture, Natural Language Processing (NLP), Generative AI, Pretraining and Fine-tuning, Ethical and Societal Implications

1 Introduction

Medical professionals now benefit from Artificial Intelligence tools through Large Language Models (LLMs) which have transformed healthcare by supporting medical professionals in diverse tasks [1]. Large language models display effective potential to enhance medical diagnosis while they facilitate automated documentation work as they assist clinical decision processes [2]. Using extensive medical publications together with extensive patient records allows LLMs to produce informed patient solutions that benefit medical results alongside decreasing professional healthcare requirements [3].

Healthcare institutions encounter multiple barriers when implementing LLMs despite their demonstrated advantages. Multiple ethics problems related to data security and system prejudice and decision-holder responsibility must be resolved to enable broad deployment of these models [4].

Published on: May, 2025 ISSN(P) 0974-1771 ISSN(O) 2581-9372

Reliability in high-stakes medical applications depends on rigorous testing of LLMs to validate both their accuracy and their explaining power regarding medical text processing and generation.

Pretrained Language Models (PLMs) to LLMs allowed progress toward creating AI models that understand intricate medical settings. The development of early PLMs such as BERT and RoBERTa established natural language processing capabilities in medicine yet newer LLMs including GPT-4 and Med-PaLM 2 present better capabilities through their contextual understanding ability and their ability to process multiple data sources [5]. These developments enabled AI to become operational in medical diagnosis along with assisting patient communication while creating customized treatment solutions.

The growing significance of ethically sound AI deployment practices becomes evident while the field expands. Healthcare professionals and patients need to trust LLM decisions which requires handling bias issues and maintaining transparent system operations. Responsible AI implementation in medical settings requires medical personnel to fulfill requirements under HIPAA and GDPR regulations. Researchers aiming to create valuable healthcare contributions through LLMs must focus on improving model accuracy and interpretability to maintain ethical standards [6].

The research examines LLMs in healthcare including their creation along with functional applications and their responsible utilization in medical practices. This paper assesses the impact of LLMs on clinical operations while reviewing their benefits and limitations thus proposing directions for similar artificial intelligence models in medical settings. The analysis will provide insights about LLM utilization for improved patient care alongside a discussion of related difficulties in their deployment.

2 Development of LLMs in Healthcare

Large Language Models in healthcare developed through advanced technology provide superior natural language understanding and contextual reasoning abilities together with multimodal integration capabilities. Early Pretrained Language Models (PLMs) including BERT and ClinicalBERT underwent major development until the arrival of highly advanced generative models Med-PaLM and BioMistral. The present generation of LLMs demonstrates sophisticated medical query processing while analyzing both organized and unstructured patient data to provide accurate help in medical choices [5].

Primary technological advances enabled by the change from PLMs to LLMs consist of self-supervised learning alongside reinforcement learning with human feedback and the ability to combine diverse medical data types. Modern healthcare relies heavily on LLMs because they bring together their processing strength for clinical content and their learning efficiency using minimal examples. Medical applications need responsible deployment of LLMs despite the need to resolve issues with model interpretability and regulatory compliance and ethical concerns [4].

The historical development and essential breakthroughs in LLMs for healthcare along with their broad medical applications will be examined while discussing the barriers which need resolution to optimize their true potential.

2.1 Transition from PLMs to LLMs in Healthcare

Advanced natural language understanding capabilities combined with better contextual reasoning capabilities emerge from Large Language Models (LLMs) development in healthcare. Early Pretrained Language Models (PLMs) including BERT and ClinicalBERT underwent major development until the arrival of highly advanced generative models Med-PaLM and BioMistral. The present generation of LLMs demonstrates sophisticated medical query processing while analyzing both organized and unstructured patient data to provide accurate help in medical choices.

Technical advancements in the transition from PLMs to LLMs introduced self-supervised learning along with reinforcement learning with human feedback (RLHF) as well as ability to integrate various medical data modalities. Modern healthcare relies heavily on LLMs because they bring together their processing strength for clinical content and their learning efficiency using minimal examples. Medical applications of LLMs require both confirmations of model interpretability and resolutions of ethical worries together with compliance with regulations to remain responsible.

The historical development and essential breakthroughs in LLMs for healthcare along with their broad medical applications will be examined while discussing the barriers which need resolution to optimize their true potential.

Table 1: Differences Between PLMs and LLMs in Healthcare

Feature	PLMs (e.g., BERT, BioBERT)	LLMs (e.g., GPT-4, Med-PaLM)
Task Type	Single-task, classification-based	Multi-task, generative-based
Training Data	Biomedical literature, structured data	Unstructured patient data, multimodal datasets
Reasoning Capability	Limited, relies on explicit training	Advanced, capable of complex medical reasoning
Adaptability	Requires extensive fine-tuning	Few-shot and zero-shot learning capabilities
Multimodal Processing	No	Yes, integrates text, images, and structured data

3 Key Technological Advancements in Healthcare LLMs

3.1 Self-Supervised Learning and Reinforcement Learning

LLMs of today employ automatic learning methods to gain knowledge from massive unmarked medical text collections. The technique allows these models to understand clinical concepts in context through unlabeled medical text without needing large labeled datasets.

The performance improvement of LLM systems depends heavily on reinforcement learning from human feedback (RLHF). Expert feedback incorporation enables models to match clinical reasoning which decreases the possibility of producing wrong medical recommendations.

3.2 Multimodal Integration

Unlike early PLMs, contemporary LLMs can process and integrate multiple types of medical data, including:

- Electronic Health Records (EHRs) – Extracting structured insights from unstructured patient data.
- Radiology and Pathology Images – LLMs like LLaVA-Med and Med-Flamingo can analyze medical images alongside clinical notes.
- Genomic Data – Personalized treatment recommendations based on genetic information. This multimodal capability enhances diagnostic accuracy and facilitates more comprehensive patient assessments.

3.3 Few-Shot and Zero-Shot Learning

One of the most significant advantages of LLMs is their ability to generalize across tasks with minimal labeled data. Few-shot learning allows models to perform complex medical NLP tasks without requiring extensive retraining, making them highly adaptable to new healthcare challenges.

3.4 Knowledge Graphs and Retrieval-Augmented Generation (RAG)

Healthcare LLMs incorporate structured medical knowledge from databases such as:

- Unified Medical Language System (UMLS)
- PubMed and ClinicalTrials.gov
- Drug Databases (e.g., DrugBank)

By combining pre-trained models with retrieval-augmented generation, LLMs can generate more accurate and evidence-based medical responses.

4 Applications and Challenges of Healthcare LLMs

Brief summarization of existing LLMs for Healthcare is given in Table 2. The table provides a brief summarization of existing LLMs for Healthcare, categorizing them based on model name, size (in billions of parameters), and key features. The models are listed in chronological order of their development, highlighting their unique capabilities such as multimodal integration, reinforcement learning, knowledge retrieval, and specific adaptations for medical tasks. For example, Med-PaLM (540B) and HealAI (540B) are among the largest models designed for medical applications, while smaller models like ChatDoctor (7B) and DoctorGLM (6B) focus on doctor-patient interactions. Some models, such as BioMistral and BiMediX, support multilingual processing, while others, like LLM-CXR and Med-Flamingo, specialize in medical imaging. This categorization provides insights into how LLMs are evolving to address different healthcare challenges, from general medical reasoning to domain-specific applications like oncology (OncoGPT) and mental health (SoulChat).

4.1 Current Applications

LLMs have demonstrated significant improvements in various healthcare domains, as shown in Figure 1:

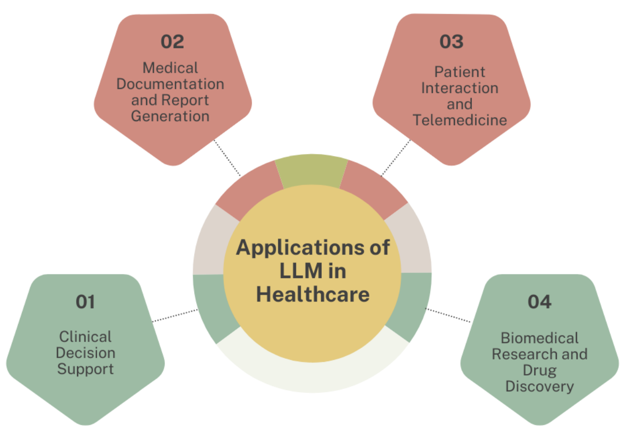

Fig. 1: Applications of LLM in Healthcare.

- Clinical Decision Support: LLMs aid healthcare professionals by summarizing patient histories, extracting insights from vast medical datasets, and recommending personalized treatment plans. These models help in reducing diagnostic errors and improving patient outcomes.
- Medical Documentation and Report Generation: By automating the generation of medical reports, including discharge summaries and radiology interpretations, LLMs minimize administrative workload for healthcare providers and enhance documentation consistency.

Table 2: Brief summarization of existing LLMs for Healthcare. Sorted in chronological order of publication.

Model Name	Size (B)	Features
GatorTron [7]	8.9	Training from scratch
Galactica [8]	120	Reasoning, Multidisciplinary
Med-PaLM [9]	540	CoT, Self-consistency
ChatDoctor [10]	7	Retrieve online, External knowledge
DoctorGLM [11]	6	Extra prompt designer
MedAlpaca [12]	13	Adapt to Medicine
BenTsao [13]	7	Knowledge graph
PMC-LLaMA [14]	7	Adapt to Medicine
Visual Med-Alpaca [15]	7	Multimodal generative model, Self-Instruct
BianQue [16]	6	Chain of Questioning
Med-PaLM 2 [17]	340	Ensemble refinement, CoT, Self-consistency
GatorTronGPT [18]	20	Training from scratch for medicine
LLM-CXR [19]	3	Multimodal, Chest X-rays
HuatuoGPT [20]	7	Reinforced learning from AI feedback
ClinicalGPT [21]	7	Multi-round dialogue consultations
MedAGI [22]	–	Multimodal
LLaVA-Med [23]	13	Multimodal, Self-instruct, Curriculum learning
OphGLM [24]	6	Multimodal, Ophthalmology LLM
SoulChat [25]	6	Mental Healthcare
Med-Flamingo [26]	80	Multimodal, Few-Shot medical VQA
Zhongjing [27]	13	Multi-turn Chinese medical dialogue
MedChatZH [28]	7	Traditional Chinese Medicine, Bilingual
JMLR [29]	13	RAG, LLM-Rank loss
BioMistral [30]	7	Multilingual, Model merging emphasis
BiMediX [31]	47	English and Arabic language
OncoGPT [32]	7	Real-world doctor-patient oncology dialogue
Polaris [33]	–	Several specialized support agents
HealAI [34]	540	RAG, Interactive Editing
Apollo [35]	7	Multilingual, Lightweight, Proxy tuning
Medical mT5 [36]	3	Multilingual
Qilin-Med [37]	7	Domain-specific pre-training, RAG
Me LLaMA [38]	70	Catastrophic Forgetting
EpiSemoGPT [39]	7	Predicting epileptogenic zones
Aloe-Alpha [40]	8	Synthetic CoT
CancerLLM [41]	7	Specifically for cancer

- Patient Interaction and Telemedicine: AI-driven virtual assistants improve healthcare accessibility by providing preliminary medical advice, answering patient queries, and offering mental health support. These applications enhance patient engagement while reducing dependency on human medical professionals.
- Biomedical Research and Drug Discovery: LLMs expedite the research process by analyzing large-scale scientific literature, identifying potential therapeutic targets, and facilitating data-driven drug discovery. By integrating structured and unstructured biomedical data, these models contribute to the rapid advancement of medical science.

4.2 Ethical and Regulatory Challenges

The deployment of LLMs in healthcare must adhere to ethical guidelines and regulatory frameworks. The complexity of integrating AI-driven solutions into medical practice presents a range of ethical concerns, including bias, accountability, and patient privacy.

- Bias and Fairness: One of the most significant ethical challenges is the presence of bias in AI models. Since LLMs are trained on vast datasets that may include historical biases, there is a risk that these biases could influence medical decisions. This can result in disparities in health-care outcomes, particularly for underrepresented populations. Ensuring fairness in AI requires

careful selection of training data, bias-mitigation techniques, and continuous monitoring of model performance.
- Transparency and Interpretability: LLMs operate as "black-box" models, meaning their decision-making processes are often opaque. This lack of transparency raises concerns for healthcare professionals, who need clear justifications for AI-generated recommendations. Developing explainable AI (XAI) techniques is essential for fostering trust and enabling clinicians to understand how a model arrived at its conclusions. Methods such as attention visualization, rule-based reasoning, and counterfactual explanations can improve interpretability.
- Accountability and Liability: The question of accountability in AI-assisted medical decisions remains unresolved. If an AI model provides incorrect or harmful recommendations, determining liability is challenging. Should the responsibility lie with the model developers, healthcare institutions, or the practitioners using the AI system? Establishing legal frameworks to define AI accountability in clinical practice is necessary to protect patients and medical professionals.
- Data Privacy and Security: Healthcare data is highly sensitive, and the use of LLMs in clinical settings must comply with stringent data protection laws such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe. Data leakage or unauthorized access to patient records can have severe consequences. Therefore, implementing robust encryption, access control mechanisms, and federated learning techniques can enhance data security while preserving patient confidentiality.
- Regulatory Compliance: The adoption of LLMs in healthcare requires regulatory approval to ensure patient safety and ethical AI use. Government agencies, such as the FDA and EMA, are working to develop guidelines for AI-driven medical applications. Compliance with these regulations is critical for gaining acceptance within the healthcare industry.

Overall, addressing these ethical and regulatory challenges is crucial for the responsible deployment of LLMs in healthcare. Future efforts should focus on creating fair, transparent, and accountable AI systems that adhere to legal and ethical standards while maximizing patient benefits.

5 Conclusion and Future Directions

The development of LLMs in healthcare has transformed medical AI applications, from clinical decision support to patient interaction. As models continue to evolve, the integration of ethical AI principles, improved transparency, and multimodal capabilities will be crucial for their widespread adoption. Future research should focus on:

Enhancing explainability and interpretability of AI models. Developing lightweight, resource-efficient LLMs for real-time medical applications. Ensuring fairness and minimizing bias in AI-driven healthcare systems. By addressing these challenges, LLMs can play a pivotal role in advancing precision medicine, improving healthcare accessibility, and enhancing patient outcomes worldwide.

LLMs have the potential to transform healthcare by enhancing diagnostic accuracy, automating documentation, and improving patient interactions. However, addressing ethical concerns, ensuring transparency, and optimizing model efficiency are crucial for their successful integration into clinical practice.

Declarations

- The authors received no specific funding for this study.
- The authors declare that they have no conflicts of interest to report regarding the present study.
- No Human subject or animals are involved in the research.
- All authors have mutually consented to participate.
- All the authors have consented the Journal to publish this paper.
- Authors declare that all the data being used in the design and production cum layout of the manuscript is declared in the manuscript.

References

[1] Plum, T., Binzberger, R., Thiele, R., Shang, F., Postrach, D., Fung, C., Fortea, M., Stakenborg, N., Wang, Z., Tappe-Theodor, A., *et al.*: Mast cells link immune sensing to antigen-avoidance behaviour. Nature **620**(7974), 634–642 (2023)

[2] Lee, B.T., Kwon, J.-m., Cho, J., Bae, W., Park, H., Seo, W.-W., Cho, I., Lee, Y., Park, J., Oh, B.-H., *et al.*: Usefulness of deep-learning algorithm for detecting acute myocardial infarction using electrocardiogram alone in patients with chest pain at emergency department: Dami-ecg study. Journal of Cardiovascular Intervention **2**(2), 100–112 (2023)

[3] Chatterjee, S., Fruhling, A., Kotiadis, K., Gartner, D.: Towards new frontiers of healthcare systems research using artificial intelligence and generative AI. Taylor & Francis (2024)

[4] Price, W.N., Cohen, I.G.: Privacy in the age of medical big data. Nature medicine **25**(1), 37–43 (2019)

[5] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (long and Short Papers), pp. 4171–4186 (2019)

[6] Tjoa, E., Guan, C.: A survey on explainable artificial intelligence (xai): Toward medical xai. IEEE transactions on neural networks and learning systems **32**(11), 4793–4813 (2020)

[7] Yang, X., PourNejatian, N., Shin, H.C., Smith, K.E., Parisien, C., Compas, C., Martin, C., Flores, M.G., Zhang, Y., Magoc, T., et al.: Gatortron: A large language model for clinical natural language processing. MedRxiv, 2022–02 (2022)

[8] Taylor, R., Kardas, M., Cucurull, G., Scialom, T., Hartshorn, A., Saravia, E., Poulton, A., Kerkez, V., Stojnic, R.: Galactica: A large language model for science. arXiv preprint arXiv:2211.09085 (2022)

[9] Singhal, K., Azizi, S., Tu, T., Mahdavi, S.S., Wei, J., Chung, H.W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., et al.: Large language models encode clinical knowledge. arXiv preprint arXiv:2212.13138 (2022)

[10] Li, Y., Li, Z., Zhang, K., Dan, R., Jiang, S., Zhang, Y.: Chatdoctor: A medical chat model fine-tuned on a large language model meta-ai (llama) using medical domain knowledge. Cureus **15**(6) (2023)

[11] Xiong, H., Wang, S., Zhu, Y., Zhao, Z., Liu, Y., Huang, L., Wang, Q., Shen, D.: Doctorglm: Fine-tuning your chinese doctor is not a herculean task. arXiv preprint arXiv:2304.01097 (2023)

[12] Han, T., Adams, L.C., Papaioannou, J.-M., Grundmann, P., Oberhauser, T., Löser, A., Truhn, D., Bressem, K.K.: Medalpaca–an open-source collection of medical conversational ai models and training data. arXiv preprint arXiv:2304.08247 (2023)

[13] Wang, H., Liu, C., Xi, N., Qiang, Z., Zhao, S., Qin, B., Liu, T.: Huatuo: Tuning llama model with chinese medical knowledge. arXiv preprint arXiv:2304.06975 (2023)

[14] Wu, C., Lin, W., Zhang, X., Zhang, Y., Xie, W., Wang, Y.: Pmc-llama: toward building open-source language models for medicine. Journal of the American Medical Informatics Association

31(9), 1833–1843 (2024)

[15] Li, C., Wong, C., Zhang, S., Usuyama, N., Liu, H., Yang, J., Naumann, T., Poon, H., Gao, J.: Llava-med: Training a large language-and-vision assistant for biomedicine in one day. Advances in Neural Information Processing Systems **36**, 28541–28564 (2023)

[16] Ye, Q., Liu, J., Chong, D., Zhou, P., Hua, Y., Liu, F., Cao, M., Wang, Z., Cheng, X., Lei, Z., et al.: Qilin-med: Multi-stage knowledge injection advanced medical large language model. arXiv preprint arXiv:2310.09089 (2023)

[17] Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Amin, M., Hou, L., Clark, K., Pfohl, S.R., Cole-Lewis, H., et al.: Toward expert-level medical question answering with large language models. Nature Medicine, 1–8 (2025)

[18] Peng, C., Yang, X., Chen, A., Smith, K.E., PourNejatian, N., Costa, A.B., Martin, C., Flores, M.G., Zhang, Y., Magoc, T., *et al.*: A study of generative large language model for medical research and healthcare. NPJ digital medicine **6**(1), 210 (2023)

[19] Lee, S., Kim, W.J., Chang, J., Ye, J.C.: Llm-cxr: instruction-finetuned llm for cxr image understanding and generation. arXiv preprint arXiv:2305.11490 (2023)

[20] Zhang, H., Chen, J., Jiang, F., Yu, F., Chen, Z., Li, J., Chen, G., Wu, X., Zhang, Z., Xiao, Q., et al.: Huatuogpt, towards taming language model to be a doctor. arXiv preprint arXiv:2305.15075 (2023)

[21] Wang, G., Yang, G., Du, Z., Fan, L., Li, X.: Clinicalgpt: large language models finetuned with diverse medical data and comprehensive evaluation. arXiv preprint arXiv:2306.09968 (2023)

[22] Zhou, J., Chen, X., Gao, X.: Path to medical agi: Unify domain-specific medical llms with the lowest cost. arXiv preprint arXiv:2306.10765 (2023)

[23] Li, C., Wong, C., Zhang, S., Usuyama, N., Liu, H., Yang, J., Naumann, T., Poon, H., Gao, J.: Llava-med: Training a large language-and-vision assistant for biomedicine in one day. Advances in Neural Information Processing Systems **36**, 28541–28564 (2023)

[24] Gao, W., Deng, Z., Niu, Z., Rong, F., Chen, C., Gong, Z., Zhang, W., Xiao, D., Li, F., Cao, Z., et al.: Ophglm: Training an ophthalmology large language-and-vision assistant based on instructions and dialogue. arXiv preprint arXiv:2306.12174 (2023)

[25] Chekuri, A., Johal, A.S., Allen, M.R., Ayers, J.W., Hogarth, M., Farcas, E.: Towards optimizing llm use in healthcare: Identifying patient questions in mychart messages. In: AMIA Annual Symposium Proceedings, vol. 2024, p. 232 (2025)

[26] Moor, M., Huang, Q., Wu, S., Yasunaga, M., Dalmia, Y., Leskovec, J., Zakka, C., Reis, E.P., Rajpurkar, P.: Med-flamingo: a multimodal medical few-shot learner. In: Machine Learning for Health (ML4H), pp. 353–367 (2023). PMLR

[27] Yang, S., Zhao, H., Zhu, S., Zhou, G., Xu, H., Jia, Y., Zan, H.: Zhongjing: Enhancing the chinese medical capabilities of large language model through expert feedback and real-world multi-turn dialogue. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 19368–19376 (2024)

[28] Tan, Y., Li, M., Huang, Z., Yu, H., Fan, G.: Medchatzh: a better medical adviser learns from better instructions. arXiv preprint arXiv:2309.01114 (2023)

[29] Wang, J., Yang, Z., Yao, Z., Yu, H.: Jmlr: Joint medical llm and retrieval training for enhancing reasoning and professional question answering capability. arXiv preprint arXiv:2402.17887 (2024)

[30] Labrak, Y., Bazoge, A., Morin, E., Gourraud, P.-A., Rouvier, M., Dufour, R.: Biomistral: A collection of open-source pretrained large language models for medical domains. arXiv preprint arXiv:2402.10373 (2024)

[31] Pieri, S., Mullappilly, S.S., Khan, F.S., Anwer, R.M., Khan, S., Baldwin, T., Cholakkal, H.: Bimedix: Bilingual medical mixture of experts llm. arXiv preprint arXiv:2402.13253 (2024)

[32] Jia, F., Liu, X., Deng, L., Gu, J., Pu, C., Bai, T., Huang, M., Lu, Y., Liu, K.: Oncogpt: A medical conversational model tailored with oncology domain expertise on a large language model meta-ai (llama). arXiv preprint arXiv:2402.16810 (2024)

[33] Mukherjee, S., Gamble, P., Ausin, M.S., Kant, N., Aggarwal, K., Manjunath, N., Datta, D., Liu, Z., Ding, J., Busacca, S., et al.: Polaris: A safety-focused llm constellation architecture for healthcare. arXiv preprint arXiv:2403.13313 (2024)

[34] Goyal, S., Rastogi, E., Rajagopal, S.P., Yuan, D., Zhao, F., Chintagunta, J., Naik, G., Ward, J.: Healai: A healthcare llm for effective medical documentation. In: Proceedings of the 17th ACM International Conference on Web Search and Data Mining, pp. 1167–1168 (2024)

[35] Wang, X., Chen, N., Chen, J., Wang, Y., Zhen, G., Zhang, C., Wu, X., Hu, Y., Gao, A., Wan, X., et al.: Apollo: A lightweight multilingual medical llm towards democratizing medical ai to 6b people. arXiv preprint arXiv:2403.03640 (2024)

[36] García-Ferrero, I., Agerri, R., Salazar, A.A., Cabrio, E., Iglesia, I., Lavelli, A., Magnini, B., Molinet, B., Ramirez-Romero, J., Rigau, G., et al.: Medical mt5: an open-source multilingual text-to-text llm for the medical domain. arXiv preprint arXiv:2404.07613 (2024)

[37] Ye, Q., Liu, J., Chong, D., Zhou, P., Hua, Y., Liu, F., Cao, M., Wang, Z., Cheng, X., Lei, Z., et al.: Qilin-med: Multi-stage knowledge injection advanced medical large language model. arXiv preprint arXiv:2310.09089 (2023)

[38] Xie, Q., Chen, Q., Chen, A., Peng, C., Hu, Y., Lin, F., Peng, X., Huang, J., Zhang, J., Keloth, V., et al.: Me llama: foundation large language models for medical applications. 2024. URL https://arxiv. org/abs/2402.12749 (2023)

[39] Yang, S., Luo, Y., Jiao, M., Fotedar, N., Rao, V.R., Ju, X., Wu, S., Xian, X., Sun, H., Karakis, I., et al.: Episemollm: A fine-tuned large language model for epileptogenic zone localization based on seizure semiology with a performance comparable to epileptologists. MedRxiv, 2024–09 (2024)

[40] Gururajan, A.K., Lopez-Cuena, E., Bayarri-Planas, J., Tormos, A., Hinjos, D., Bernabeu-Perez, P., Arias-Duart, A., Martin-Torres, P.A., Urcelay-Ganzabal, L., Gonzalez-Mallo, M., et al.: Aloe: A family of fine-tuned open healthcare llms. arXiv preprint arXiv:2405.01886 (2024)

[41] Li, M., Huang, J., Yeung, J., Blaes, A., Johnson, S., Liu, H., Xu, H., Zhang, R.: Cancerllm: A large language model in cancer domain. arXiv preprint arXiv:2406.10459 (2024)