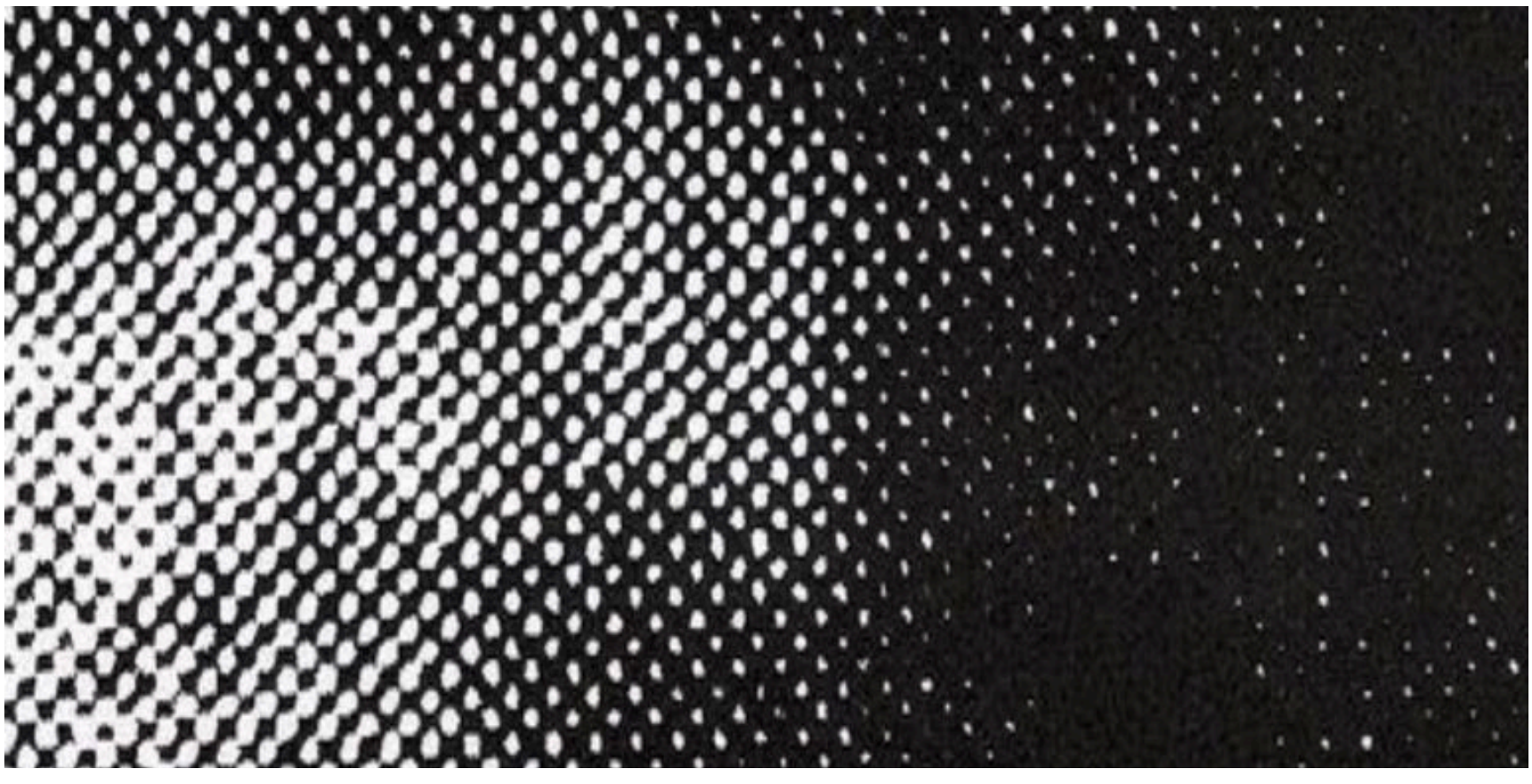

Title: Large Language Models and Artificial Intelligence in Healthcare: A Comprehensive Review

Authors:

- Debargha Brahma , B.Tech- CSE (AI & ML), Section-H, UG/SOET/30/24/293, Adamas University, Email: [debargha1.brahma@stu.adamasuniversity.ac.in]
- Devlina Dutta , B.Tech- CSE (AI & ML), Section-H, UG/SOET/30/24/036, Adamas University, Email: [devlina1.dutta@stu.adamasuniversity.ac.in]

Keywords:

Large language models, artificial intelligence, healthcare, clinical decision support, medical documentation, ethics, multimodal AI

Word Count:

Words	771 of 771
Characters	6350 of 6350
Characters excluding spaces	5624 of 5624

Plain Language Summary

Large language models (LLMs)—advanced AI systems that understand and generate human language—are rapidly transforming healthcare. These tools help doctors make better decisions, write medical notes, enhance patient communication, and accelerate medical research. Despite their promise, challenges remain around accuracy, privacy, and integration with physician workflows. This review examines how LLMs are being adopted in healthcare, their benefits, current limitations, and future possibilities.

Abstract

Background:

Large language models (LLMs) are transforming healthcare through their ability to analyze vast amounts of text and support clinical practice.

Objectives:

To assess the applications, technical evolution, impact, challenges, and future directions of LLMs in healthcare.

Methods:

A narrative review of recent literature, technological advances, and clinical studies evaluating LLM integration in diagnostics, workflow, documentation, and education.

Results:

LLMs improve diagnostic accuracy, streamline workflows, enhance documentation, and transform education and patient communication. Challenges persist with hallucinations, privacy, contextual reasoning, ethics, and workflow integration.

Conclusions:

While LLMs offer transformative potential, responsible deployment requires robust oversight, technical validation, human-AI collaboration models, and attention to ethics and equity.

1. Introduction

Large language models represent a revolutionary advance in artificial intelligence, fundamentally transforming healthcare delivery and medical practice. Their ability to process clinical, research, and patient data enables unprecedented support for diagnostics, workflow efficiency, and research acceleration. The rise of tools like ChatGPT in 2022 spurred further development and research, highlighting both opportunities and critical challenges around privacy, accuracy, and ethical deployment.

2. Methods

This review synthesizes published research, technology reports, and real-world LLM deployments in healthcare. We extracted applications, technical details, and performance metrics from peer-reviewed journals, clinical trials, and grey literature. Ethical and regulatory considerations were drawn from official guidelines and ongoing policy debates.

3. Results

Technical Foundations and Multimodal Capabilities

LLMs—primarily transformer-based architectures—have evolved from millions to billions of parameters, improving natural language processing in clinical tasks. Domain-specific models (such as GatorTron, trained on 90+ billion tokens of medical text) outperform generic models for clinical concept extraction, relation identification, and medical question answering.

Multimodal models now process medical images, structured data, and audio alongside text. While promising, true synergy between modalities (e.g., Med-Flamingo, LLaVA-Med) remains early-stage, with text-only baselines still highly competitive.

Clinical Decision Support and Diagnostic Applications

LLMs analyze patient symptoms, records, and literature to support diagnosis. Diagnostic accuracy varies (25–97.8% across contexts and specialties), sometimes matching or exceeding physicians—notably in ophthalmology and emergency care. For instance, GPT-4 outperformed resident physicians on standardized internal medicine vignettes, while models like GatoTron aid pharmacovigilance by identifying drug interactions in electronic health records.

Medical Writing, Documentation, and Education

LLMs streamline medical writing by drafting manuscripts, summaries, and operative notes. They simplify complex data and improve medical communication clarity. In education, LLMs simulate realistic patient encounters, act as virtual tutors, and personalize learning for trainees. For patients, LLMs explain radiology reports and triage urgency, matching physician-level accuracy in certain tasks.

Diagnostic Imaging and Workflow Automation

LLMs and AI-driven computer vision have achieved human-level—and often superior—performance in classifying radiological images and pathology slides. Automated AI reporting reduces documentation workload and increases efficiency for clinicians and staff.

AI models accelerate outbreak detection and diagnosis (e.g., rapid differentiation of respiratory viruses, COVID-19).

Challenges, Limitations, and Ethical Issues

- **Hallucinations and Reliability:** High-stakes environments demand rigorous validation to prevent misleading clinicians or patients.
- **Interpretability:** Black-box decisions complicate accountability and integration into evidence-based practice.

- **Bias and Data Quality:** Limited or non-diverse data may embed social or demographic biases.
- **Privacy and Regulation:** Handling sensitive health data under HIPAA, GDPR, and other regulations brings risks of data breaches.
- **Workflow Integration:** Physician collaboration with LLMs does not always improve diagnostic reasoning without careful design.

Geographic and Specialty Trends

Most research originates from the United States, India, and Germany, potentially limiting generalizability. Specialties like genitourinary, digestive, and cardiovascular medicine lead in LLM adoption, while psychiatry and neurology remain underexplored.

---

4. Discussion

LLMs are reshaping multiple aspects of healthcare. Evidence shows these models can rival or surpass human experts in defined diagnostic tasks, streamline documentation, and enable new educational methods. Yet limitations persist: hallucinations may compromise patient safety, interpretability remains challenging, and equity concerns arise from biased training data.

Successful implementation depends on multidisciplinary governance, transparent workflows, and ongoing evaluation. Multimodal and domain-specialized LLMs offer promising next steps. Human-AI collaboration—rather than replacement—will maximize benefits and minimize risks.

---

5. Conclusion

LLMs are poised to transform healthcare delivery, with applications spanning clinical support, administration, education, and beyond. Yet responsible deployment must balance innovation with caution—ensuring accuracy, privacy, fairness, and clinician oversight remain paramount. Future progress requires improved technical solutions, regulatory frameworks, and sustained interdisciplinary collaboration so these technologies augment rather than replace the human core of healthcare.

Disclosures

Author Contributions:

Debargha Brahma designed research, drafted manuscript ; Devlina Dutta analysed data; both the authors approved the final version.

Conflict of Interest:

The authors declare no conflicts of interest.

Data Access Statement:

All data are available upon reasonable request.

Ethics Statement:

Not applicable.

References

1. **Large Language Models in Healthcare and Medical Applications: A Review**

2. **The application of large language models in medicine: A scoping review**

3. **Large Language Models in Medicine: Applications, Challenges, and Future Directions**

4. **Impact of Artificial Intelligence (AI) Technology in Healthcare Sector: A Critical Evaluation of Both Sides of the Coin**

5. **Frontiers | Editorial: Large Language Models for medical applications**

6. **Artificial intelligence in healthcare: transforming the practice of medicine**

7. **Large language models forecast patient health trajectories enabling digital twins - npj Digital Medicine**

8. **An automated framework for assessing how well LLMs cite relevant medical references - Nature Communications**

9. **Large Language Models Could Revolutionize Health Care, but Technical Hurdles May Limit Their Applications**

10. **Tandfonline**

11. **Jamanetwork**

12. **7 ways AI is transforming healthcare**

13. **Leveraging Large Language Models and AI for Intelligent Healthcare Systems (HealthLLM) – AICCSA 2025**

14. **Mednexus**

15. **Comparing Diagnostic Accuracy of Clinical Professionals and Large Language Models: Systematic Review and Meta-Analysis**

16. **ChatGPT With GPT-4 Outperforms Emergency Department Physicians in Diagnostic Accuracy: Retrospective Analysis**

17. **A large language model for electronic health records - npj Digital Medicine**

18. **Impact of Artificial Intelligence (AI) Technology in Healthcare Sector: A Critical Evaluation of Both Sides of the Coin**

19. **Jamanetwork**

20. **Assessing ChatGPT 4.0's test performance and clinical diagnostic accuracy on USMLE STEP 2 CK and clinical case reports - Scientific Reports**

21. **Natural Language Processing in Electronic Health Records in relation to healthcare decision-making: A systematic review - PubMed**

22. **Artificial intelligence in healthcare: transforming the practice of medicine**