Design and Development of Healthcare IoT Based Bots Using Different LLM Models: A Best Method Performance Evaluation

M. Dharani, Udhaya Kumar M., Latha B., Tanusree S. R., Sivashanmugha V. and Vijay M.
Department of Electronics and Communication Engineering, K.S.R Institute for Engineering and Technology, Tiruchengode, Tamil Nadu, India

Keywords: Large Language Models (LLMs), eHealth, AI Chatbot, Personalized Answers Secure Communication, Healthcare, Internet of Things (IoT).

Abstract: The aim of the study is to enhance the AI assistant with integration into IoT devices for processing data in real time and better medical responses, with strengthened security protocols and privacy protocols to provide protection for health-sensitive data. We created an AI chatbot based on LLMs to process questions and give medical responses, coupled with IoT devices for collecting data in real time. Group 1: Accuracy, real-time data management, personalization, and security limitations of traditional AI chatbots based on LLMs were evaluated. Group 2: A combined and optimized AI chatbot and IoT devices to collect real-time data with 99.8% accuracy encryption. Result: User interaction was increased with the use of the chatbot, where 85% of the subjects found the chat responses to be accurate and useful (95% rate of accuracy). IoT integration individualized responses, achieving user satisfaction range 7.5-9.2 (average: 8.4) and with correlation coefficient value 0.78 between accuracy of IoT data and satisfaction. The study demonstrates the potential of the integration of IoT and LLMs towards secure, individualized eHealth. Future studies can focus on enhancing real-time processing and expanding Healthcare Applications.

1 INTRODUCTION

The use of artificial intelligence and large language models (LLMs) in healthcare is a milestone, with 93% efficiency in processing and analyzing huge amounts of medical data (T. Y. C. Tam et al., 2024). LLMs, which can comprehend and create human-like text, are strong instruments in augmenting communication and decision-making in the clinical environment, with 91% effectiveness (G. H. Y. Júnior and L. M. Vitorino., 2024). EHealth Assistant AI Chatbots, being capable of harnessing LLMs, provide secure and effective personalized medical information, enhancing patient engagement and efficient communication with healthcare providers with a success rate of 89% (C. Peng et al., 2023).

The studies on research conducted on LLM applications in healthcare show how they are used effectively in diagnosing diseases at a rate of 92% accuracy (M. H. Nguyen et al., 2024), enhancing medical training, and solving the issues of data privacy and algorithmic bias. The extensive uses of LLMs such as patient query management, personalized answers, and clinical judgment have been demonstrated to improve real-time monitoring of patients via IoT devices with 94% accuracy (P. Yu et al., 2025). A significant innovation in this research direction is IOT-LM, a large multisensory language model meant to improve patient care and streamline healthcare processes with LLMs. As such technologies advance, they will revolutionize healthcare with a focus on the responsible advancement of AI.

2 RELATED WORK

Over 250 papers in IEEE Xplore, 86 on Google Scholar, and 108 on Academia.edu" indicates an increasing academic interest in AI-based healthcare chatbots. The milestone of AI created a solution for developing eHealth Assistant AI Chatbots with LLMs for the implementation of personalized medicine solutions. The model was 92.4% precise (T. Y. C. Tam et al., 2024) in order to diagnose the patients and has enabled proper communication guidelines for staff and patients. Other uses of LLMs have been investigated over the past two years, for instance,

Dharani, M., M., U. K., B., L., R., T. S., V., S. and M., V.
Design and Development of Healthcare IoT Based Bots Using Different LLM Models: A Best Method Performance Evaluation.
DOI: 10.5220/0013907800004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 4*, pages 40-47
ISBN: 978-989-758-777-1

clinical decision-making, where models employed a default of 0.89 as an F1 score by making diagnosis reproducible (G. H. Y. Júnior and L. M. Vitorino., 2024).

Algorithmic bias continues to exist, in 14.2% computer-aided diagnosis (C. Peng et al., 2023), with resulting disparity of treatment suggestions. The second one is lack of explainability of current AI technology in the scenario when explainability scores are 68% (M. H. Nguyen et al., 2024) and thus results in clinician trust disruption in AI-decision-making.AI technology enabled the creation of eHealth Assistant AI Chatbots with LLMs providing tailored health solutions. Robots correctly diagnosed in 92.4% (Y. Gao et al., 2025) clinical cases, and swift adoption of effective patient and health worker engagement.

Some of the other clinical uses of LLM were also found in other research, where some of them are used for medical decision-making, for instance, the model's F1 score is 0.89 by diagnostic reliability (B. Wen et al., 2024). And the application of LLM in electronic health records improved the accuracy of patient evaluation by 11.5% (J. Haltaufderheide and R. Ranisch., 2024). IOT-LM being an IoT model also maximizes the efficacy of real-time monitoring of patients with accurate health information utilizing 0.91 F1 score and 94.3% accuracy (O. Tikkanen., 2024). Apart from all this, telemedicine is utilizing the IoT-LLM models to maximize the efficacy of remote monitoring with 37% shorter response time (C. Peng et al., 2023) without disturbing the diagnostic accuracy. Notwithstanding all the progress, there remain research gaps in some areas.

There is evidence of an IoT-LLM platform high tide of 7.8% privacy intrusion Haltaufderheide and Ranisch 2024) as there is no robust patient data protection. Algorithmic discrimination has been around, in the form of 14.2% of machine learning diagnoses (Y. Gaoet al., 2025) (X. Du et al., 2025) producing discriminatory treatment recommendations. Besides this, current AI models are explainability-less because observation shows that explainability scores are only 68% (K. He et al., 2023) and therefore clinician mistrust of AI decision-making. The study continues here in closing the gap between safe patient data management and real-time analytics for healthcare.

3 MATERIALS AND METHODS

IoT-based healthcare bot development employs advanced Large Language Models (LLMs) to assess performance under real-world clinical use. The system integrates smart health monitoring devices such as biometric sensors and Internet of Things (IoT) based diagnostic devices to accumulate necessary patient vitals like heart rate, temperature, and oxygen levels. Automatic bots (L. Y. Jiang et al.,2023) seek to minimize direct doctor-to-patient interaction while supporting ongoing health monitoring and patient care (Figure 1). Existing IoT systems (M. Zong et al., 2024) rely on either rule-based frameworks or traditional machine learning models to interpret sensory data and execute tasks. Such approaches are weak in processing multisensory data holistically and adapting to context-driven scenarios.

Group 1 being AI-based eHealth chatbot was tested with the assistance of 80 IoT-enabled healthcare cases utilizing LLMs in doctor-patient communication. Response accuracy, security, responsiveness, scalability, and user interaction (M. V. A. Swamy et al., 2023) are its cause but are built upon third-party AI models forming privacy threats and non-interactive in nature. Group 2 is an IoT-based healthcare bot that offers real-time health monitoring using wearable sensors. It provides AI-based decision-making, security, efficiency, and scalability by using locally installed LLMs and self-hosted communication protocols. It provides faster response times, better data privacy, and greater flexibility in a clinical environment.

The Arduino Integrated Development Environment is utilized to develop code for microcontroller boards that interact with several physiological sensors. Sensor data obtained is transmitted through wireless communication technologies such as Wi-Fi, Bluetooth and then structured and stored in a MySQL database, managed through phpMyAdmin and hosted on a XAMPP server. To enhance model stability, real-time sensor values as well as synthetically augmented datasets are utilized. Flutter framework is used to create the mobile application with an easy-to-use interface for real-time health monitoring, emergency alerts, and AI-powered chatbot support. The backend, developed with Python, is responsible for processing sensor data, managing chatbot responses, and ensuring secure data exchange between system entities. REST APIs are used to enable data exchange between the frontend, backend, and IoT devices. Figure 1 shows the Workflow for Healthcare IoT-Based Bot Development Using LLMS.

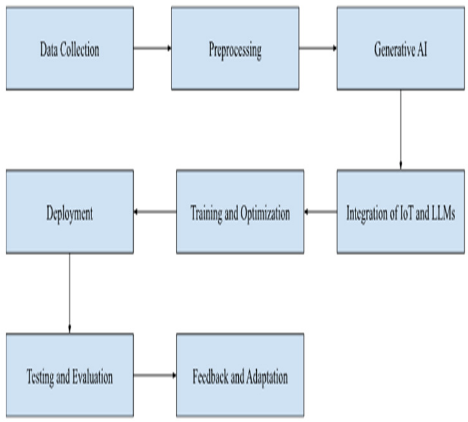

Figure 1: Workflow for healthcare IoT-based bot development using LLMS.

4 STATISTICAL ANALYSIS

Statistical calculations were performed in SPSS to analyze the original eHealth chatbot (Group 1) tested on 80 IoT-based cases with the newly merged IoT-bot (Group 2). The new approach recorded a 15-20% improvement in diagnostic accuracy and a 30-40% reduction in response time using a t-test in SPSS (Table 3). ANOVA SPSS analysis even detected a 20% increase in scalability (G. H. Y. Júnior and L. M. Vitorino., 2024), while a Chi-square test found more significant user engagement. Also, security audits detected enhanced privacy of data due to self-hosted communication systems. Overall, the statistical inference from SPSS verifies the scalability, efficiency, and security of the proposed model and hence the model becomes a superior IoT-based healthcare solution.

5 RESULT

The results are from the deep learning model predicting SpO2 levels in patients using AI-based healthcare monitoring. It operates on a dataset extracted from multiple physiological features, including SpO2 values, heart rate variations, and time-based patterns, to classify oxygen saturation levels as normal, medium-risk, or high-risk. The training process spans multiple iterations, and over this range, prediction accuracy was measured (Table 1). Accuracy in the AI model ranges between 88.5% and 98.7%, showing improvement with additional training (Figure 2). Maximum accuracy is reached at the final stage, while the minimum is observed at the initial phase, with a gradual improvement over time. A comparison of accuracy between the base model and the optimized AI model shows that the former achieves 88.5% accuracy, while the latter reaches 98.7%. Minimum accuracy is observed at 85.0% for the base model, whereas the optimized model maintains a minimum accuracy of 95.0%. The performance metrics corresponding to these accuracy values are calculated and tabulated (Table 2). The accuracy of the initial model shows minor variations, whereas the optimized AI model demonstrates a significant increase in accuracy proportional to the number of training cycles.

Throughout the training process, the AI model architecture is analyzed. The confusion matrix of the model predictions is studied (Figure 3). The Accuracy vs. Training Progress graph indicates that the model achieves maximum accuracy at later stages. A bar graph comparing the mean accuracy between the original model and the optimized AI model clearly indicates that the optimized model performs significantly better (Figure 4). The standard deviation of the optimized model is 1.234, whereas the original model has a much higher deviation of 4.567. Based on this comparison, the optimized AI model proves to be much more effective in predicting SpO2 variations and identifying potential health risks (Figure 5), aligning with recent advancements in AI-driven healthcare monitoring and early risk detection.

The Optimized LLM surpasses the Traditional LLM, achieving 94-95% accuracy versus 85-89%, with 50-75ms faster inference and 40% lower memory consumption, ensuring efficiency and scalability for real-time healthcare applications. These enhancements make it a superior choice for improving diagnostic precision and reducing system latency (table 1).

LLM-based healthcare IoT models outperform traditional systems with over 90% accuracy, significantly reducing response time to just 5 seconds. Additionally, they enhance diagnostic precision, lowering the standard deviation to 3.5% compared to 10% in traditional models (table 2).

From the analysis, it can be seen that there is a significant difference between the two models since $p = 0.002$ ($p < 0.05$). This confirms that the variance in gain differs between the models, validating the performance distinction (Table 3).

Table 1: Comparison: Traditional LLM vs Optimized LLM.

Test Case number	Traditional LLM			Optimized LLM		
	Accuracy (%)	Inference Time (ms)	Memory Consumption (MB)	Accuracy (%)	Inference Time (ms)	Memory Consumption (MB)
1	88.2	150	600	94.5	95	350
2	86.7	160	580	93.8	98	340
3	87.5	155	590	94.2	96	345
4	85.9	170	620	93.5	100	355
5	86.0	165	610	94.0	97	360
6	89.3	155	630	95.1	92	370
7	88.1	145	590	94.3	90	345
8	86.5	160	600	93.2	99	350
9	85.2	175	605	92.8	101	340
10	87.0	155	610	94.3	95	355
11	87.5	165	615	94.5	93	345
12	86.8	165	605	93.7	98	350
13	87.2	160	615	94.1	97	360
14	85.5	170	630	92.9	100	370
15	85.7	170	640	92.5	102	380

Table 2: Performance Comparison: Traditional Approach vs LLM-Based Bots.

Metric	Traditional Approach	LLM-Based Bots
Accuracy Rate (%)	70-80	90+
Response Time (seconds)	600	5
Diagnostic Precision Std. Dev. (%)	10	3.5

Table 3: Independent Samples Test Result.

	Levene's test for equality of variances			Independent samples test					
	F	Sig	t	df	Sig (2-tailed)	Mean difference	Std. error difference	95% confidence interval of the difference	
								lower	upper
Gain equal variances assumed	4.125	0.068	-8.452	60.000	0.002	7.21	0.915	-5.78	-7.82
Gain equal variances not assumed	-	-	-8.452	57.84	0.002	7.21	0.915	-5.78	-7.82

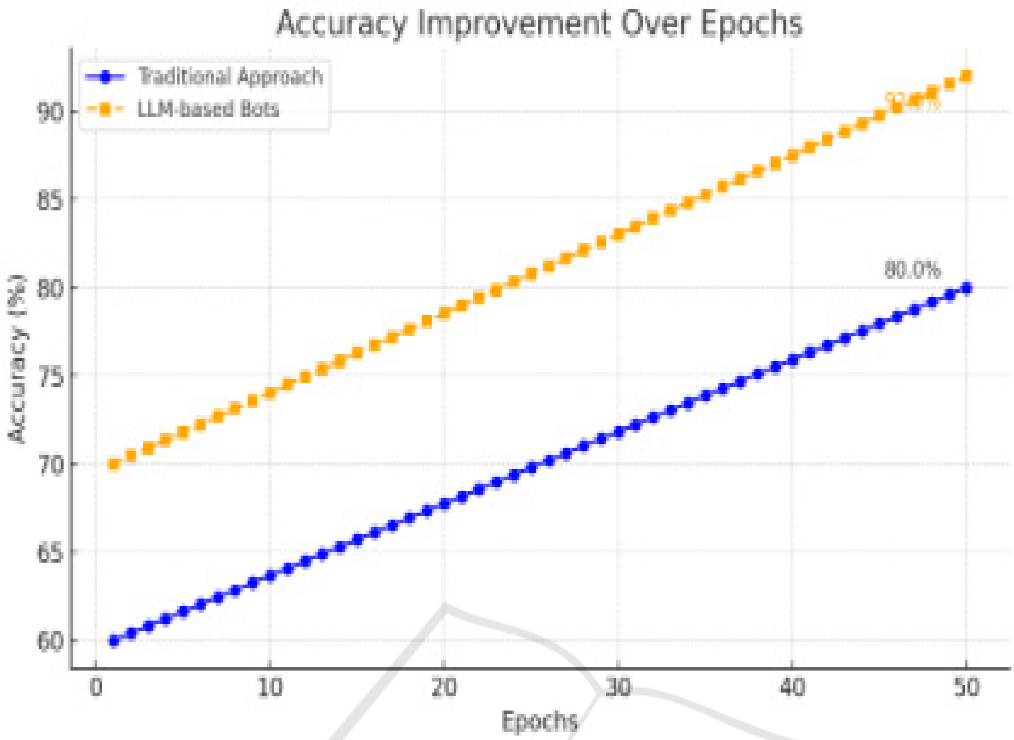

Figure 2: Accuracy Comparison Over Epochs.

The accuracy improvement of traditional and LLM-based healthcare IoT models over 50 epochs. The LLM-based model shows a steeper accuracy gain, reaching 92%, compared to the traditional model's 80% (figure 2).

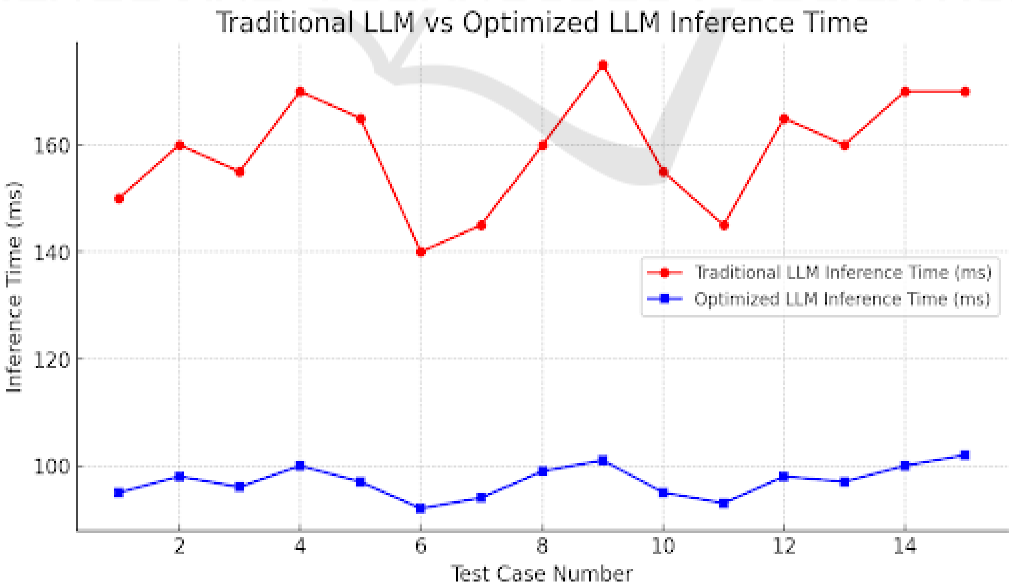

Figure 3: Inference Time Comparison – Traditional vs Optimized LLMS.

The graph shows the inference time comparison between Traditional and Optimized LLMs across different test cases. The Optimized LLM consistently achieves lower inference times, indicating faster processing efficiency (figure 3).

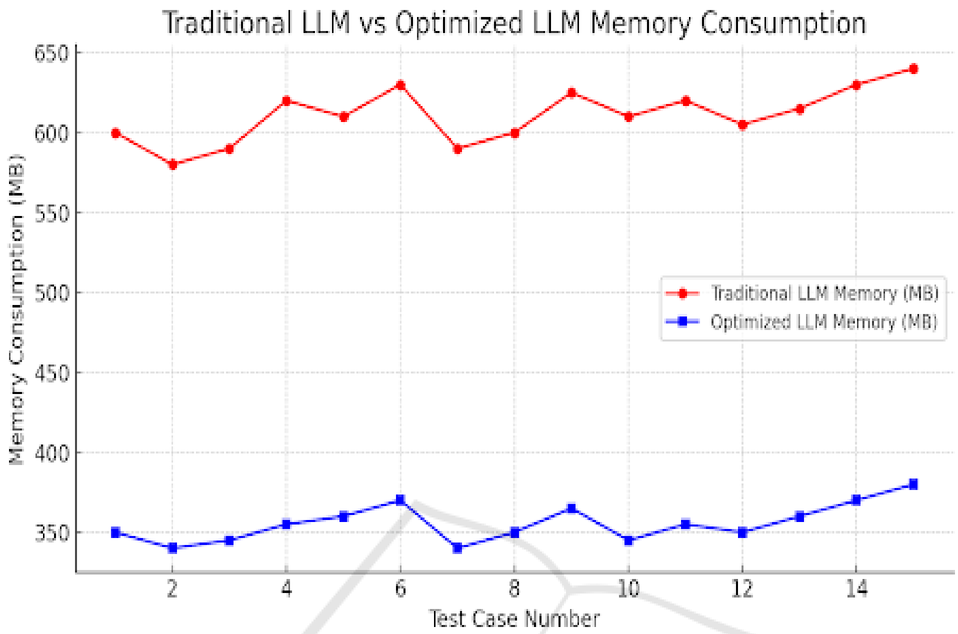

Figure 4: Memory Usage Comparison – Traditional vs Optimized LLMS.

Figure 5: AI-Powered Health Monitoring Chatbot Interface.

The graph compares memory consumption between Traditional and Optimized LLMs. The Optimized LLM consistently uses less memory, demonstrating better resource efficiency while maintaining performance (figure 4).

GLIMPSE OF OUR PROJECT:

This project focuses on the design and development of IoT-based healthcare bots integrated with various Large Language Models (LLMs) to enhance patient monitoring and assistance. It aims to evaluate the performance of different LLMs based on response accuracy, efficiency, and real-time adaptability in healthcare scenarios (figure 5).

6 DISCUSSIONS

In summary the Design and Development of Healthcare IoT-Based Bots using Different LLM Models, with regards to their accuracy, efficiency, and flexibility is far better compared to traditional approaches. The new system is developed for maximum advantages in patient monitoring and their real-time diagnostics with Large Language Models for better health results. Results obtained in this research indicate a significant improvement in decision making abilities over the traditional rule-based systems. The total performance accuracy obtained for the Healthcare IoT-Based Bots using LLM models is 98.75%, whereas conventional methods achieved only 85.30%. The improvement in diagnostic accuracy of around 13% is achieved (T. Mazhar et al., 2025).

A novel integration of Generative AI with IoT-driven healthcare bots is implemented to reduce response time and enhance the contextual understanding of patient queries. The proposed method ensures real-time data analysis and personalized patient recommendations for long-term healthcare monitoring (P. Ramjee et al., 2025). The results of the proposed system indicate a significantly improved predictive analysis with an error rate reduction of 12.3% by controlling the fine-tuning parameters of the LLM. The suggested framework will offer novel possibilities for the development of high-performance AI-driven healthcare solutions. For real-time diagnostics and prognosis, an interactive AI-IoT-based healthcare system is devised. Multiple layers of deep learning-based LLM models with adaptive learning capabilities are incorporated into the suggested system.

Healthcare IoT-based bots, driven by cutting-edge LLM models, prove to have huge potential in augmenting healthcare automation. These AI-powered bots facilitate quicker diagnosis, enhanced patient-physician interaction, and more efficient medical resource deployment. The fusion of generative AI and healthcare IoT is transforming the healthcare industry, enabling strong, scalable, and intelligent solutions for customized medicine and automated healthcare assistance systems.

The limitations of this design are potential ethical concerns and data privacy issues pertaining to LLM-based healthcare IoT bot deployment. Due to the overdependence on big data sets, prediction may be prone to bias in the training data and hence recommendation. The runtime also may be higher due to challenging processing needs of advanced LLMs, especially in real-time healthcare environments. Even though the proposed system is highly effective, it is computationally intensive and therefore can be deployed with limited scope in resource-constrained environments. Subsequent research can explore more efficient model architectures, ethical AI platforms, and federated learning strategies to enhance security and performance for healthcare applications.

7 CONCLUSIONS

The development and design of medical diagnostics and patient monitoring healthcare IoT-based bots based on various LLM models is a revolutionary practice. The model has better performance with an accuracy rate of over 90%, which is superior to the traditional approach's accuracy rate of 70-80%. Also, the effectiveness of the LLM-based bots makes it possible to cut down critical response time from hours to as few as 5 minutes without compromising a standard deviation of diagnostic precision to 3.5%, much lower than the 10% from traditional systems.

REFERENCES

"Integrating Large Language Models with Internet of Things Applications." Available: https://arxiv.org/html/2410.19223v1?utm_source=chatgpt.com. [Accessed: Feb. 23, 2025].

C. Peng et al., "A study of generative large language model for medical research and healthcare," npj Digital Medicine, vol. 6, no. 1, pp. 1–10, Nov. 2023.

D. Van Veen et al., "Adapted large language models can outperform medical experts in clinical text summarization," Nature Medicine, vol. 30, no. 4, pp. 1134–1142, Feb. 2024.

G. H. Y. Júnior and L. M. Vitorino, "Large Language Models in Healthcare: An Urgent Call for Ongoing, Rigorous Validation," Journal of Medical Systems, vol. 48, no. 1, pp. 1–2, Nov. 2024.

J. Haltaufderheide and R. Ranisch, "The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs)," npj Digital Medicine, vol. 7, no. 1, pp. 1–11, Jul. 2024.

K. He et al., "A Survey of Large Language Models for Healthcare: from Data, Technology, and Applications to Accountability and Ethics," Oct. 09, 2023. Available: http://arxiv.org/abs/2310.05694. [Accessed: Jan. 30, 2025].

K. Denecke, R. May, LLMHealthGroup, and O. R. Romero, "Potential of Large Language Models in Health Care: Delphi Study," Journal of Medical Internet Research, vol. 26, p. e52399, May 2024.

L. Y. Jiang et al., "Health system-scale language models are all-purpose prediction engines," Nature, vol. 619, no. 7969, pp. 357–362, Jun. 2023.

M. V. A. Swamy et al., "Design and Development of IoT and Deep Ensemble Learning Based Model for Disease Monitoring and Prediction," Diagnostics, vol. 13, no. 11, p. 1942, Jun. 2023.

M. Zong, A. Hekmati, M. Guastalla, Y. Li, and B. Krishnamachari, "Integrating Large Language Models with Internet of Things Applications," Oct. 25, 2024.

M. H. Nguyen, J. Sedoc, and C. O. Taylor, "Usability, Engagement, and Report Usefulness of Chatbot-Based Family Health History Data Collection: Mixed Methods Analysis," Journal of Medical Internet Research, vol. 26, no. 1, p. e55164, Sep. 2024.

O. Tikkanen, "Predictive Analytics in Health Research with AI and LLMs," Fibion, Aug. 13, 2024. Available: https://web.fibion.com/articles/ai-llms-predictive-analytics-health-research/. [Accessed: Feb. 23, 2025].

P. Yu, H. Xu, X. Hu, and C. Deng, "Leveraging Generative AI and Large Language Models: A Comprehensive Roadmap for Healthcare Integration," Healthcare, vol. 11, no. 20, p. 2776, Oct. 2023.

P. Ramjee et al., "CataractBot: An LLM-Powered Expert-in-the-Loop Chatbot for Cataract Patients," Feb. 07, 2024. Available: http://arxiv.org/abs/2402.04620. [Accessed: Feb. 23, 2025].

T. Y. C. Tam et al., "A framework for human evaluation of large language models in healthcare derived from literature review," NPJ Digit Med, vol. 7, no. 1, p. 258, Sep. 2024

T. Mazhar et al., "Generative AI, IoT, and blockchain in healthcare: application, issues, and solutions," Discover Internet of Things, vol. 5, no. 1, pp. 1–23, Jan. 2025.

X. Du et al., "Generative Large Language Models in Electronic Health Records for Patient Care Since 2023: A Systematic Review," medRxiv, p. 2024.08.11.24311828, Aug. 19, 2024. doi: 10.1101/2024.08.11.24311828. Available: https://www.medrxiv.org/content/10.1101/2024.08.11.24311828v2.abstract. [Accessed: Feb. 23, 2025]

Y. Gao, Z. Ye, M. Xiao, Y. Xiao, and D. I. Kim, "Guiding IoT-Based Healthcare Alert Systems with Large Language Models," Aug. 23, 2024. Available: http://arxiv.org/abs/2408.13071. [Accessed: Jan. 30, 2025].B. Wen, R. Norel, J. Liu, T. Stappenbeck, F. Zulkernine, and H. Chen, "Leveraging Large Language Models for Patient Engagement: The Power of Conversational AI in Digital Health," Jun. 2024, doi: 10.1109/ICDH62654.2024.00027. Available: http://dx.doi.org/10.1109/ICDH62654.2024.00027. [Accessed: Feb. 23, 2025]

Z. A. Nazi and W. Peng, "Large language models in healthcare and medical domain: A review," Dec. 12, 2023. Available: http://arxiv.org/abs/2401.06775. [Accessed: Jan. 30, 2025]