Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

The Impact of Artificial Intelligence (ChatGPT) on Medical Decision-Making in the Healthcare System of Afghanistan

Mohammad Reza Cov Andish

mohammadrezacovandish@gmail.com

Kabul University

Mustafa Muradi

Kabul University

Ahmad Sorosh Hamdard

Kabul University

Research Article

Keywords: Clinical Decision Support Systems, Large Language Models, ChatGPT, Healthcare Infrastructure, Medical Ethics, Quantitative Analysis

Posted Date: October 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11287099/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: The authors declare no competing interests.

The Impact of Artificial Intelligence (ChatGPT) on Medical Decision-Making in the Healthcare System of Afghanistan

Mohammad Reza Cov Andish, **Mustaffa Muradi**, and **Ahmad Srosh Hamdard**

Department of Information Systems, Kabul University, Kabul, Afghanistan

mohammadrezacovandish@gmail.com
muradim412@gmail.com
sroshhamdard2002@gmail.com

***Abstract*—The integration of Large Language Models (LLMs) into clinical practice is altering medical decision-making. While many studies evaluate the theoretical accuracy of models like ChatGPT in high-resource settings, data on their real-world deployment in developing nations remains scarce. This research investigates the practical use, perceived accuracy, and safety of ChatGPT in Afghanistan's healthcare system, a sector facing severe resource limits and a shortage of specialists. We conducted a cross-sectional quantitative survey of 42 medical professionals in Kabul's public and private clinics. We evaluated four main clinical metrics: Appropriateness, Patient Usefulness, Safety, and Alignment with Guidelines. Survey results show a high adoption rate among early-career physicians (66.7% having under six years of experience), who mainly use the tool for general medical information and interpreting lab tests. While ChatGPT scored highly for patient education, practitioners reported significant reservations about clinical safety. Specialists, especially surgeons and internists, raised concerns about AI hallucinations during complex diagnoses and treatments. Despite these risks, 59.5% of the surveyed doctors believe AI will improve the Afghan healthcare system by filling information gaps. We conclude that while ChatGPT is a helpful information tool, its safe use requires strict human oversight and the development of clear clinical guidelines.**

***Index Terms*—Clinical Decision Support Systems, Large Language Models, ChatGPT, Healthcare Infrastructure, Medical Ethics, Quantitative Analysis.**

I. Introduction

Clinical decision-making traditionally relies on empirical evidence, medical education, and experience. Advanced natural language processing (NLP), specifically Large Language Models (LLMs), adds a new dimension to this process [1], [2]. Models like ChatGPT can analyze and output vast amounts of medical literature instantly, acting as dynamic Clinical Decision Support Systems (CDSS).

The impact of this technology depends heavily on the setting. In wealthy nations, hospitals use LLMs to reduce administrative work and prevent burnout [3], [4]. In contrast, developing nations often rely on these models as primary medical references. Decades of instability, a lack of specialists, and poor diagnostic tools severely limit the Afghan healthcare system [5], [15]. General practitioners often face high patient volumes alone, with limited access to specialists or updated drug databases.

Here, ChatGPT provides an accessible, low-cost way to bridge these gaps. Physicians query the model for differential diagnoses, check drug interactions, or translate clinical guidelines. Still, using a general text model for critical medical choices is risky [6]. Since LLMs are designed to generate plausible text rather than retrieve verified facts, they can produce convincing but incorrect medical advice, known as hallucinations [7], [8]. Acting on these errors without regulatory oversight can directly compromise patient safety.

So far, very few studies examine how Afghan healthcare workers use AI in real clinics [17]. This study addresses that gap. By conducting a quantitative field survey among active medical personnel in Kabul, we measure how often doctors use AI, evaluate its accuracy across different specialties, and determine its safety limits. The goal is to provide a foundational dataset to help build future digital health policies in Afghanistan.

II. Literature Review

A. Evolution of Clinical Decision Support Systems

Using computers to help with medical diagnosis is not a new idea. Early expert systems required doctors to manually input patient data to get diagnostic probabilities. These systems were rigid and hard to scale because human biology and medical research constantly change. Machine learning later improved pattern recognition, especially in radiology [9]. However, these models were limited to specific tasks. Today's AI, powered by LLMs, is fundamentally different. It offers a general, conversational interface that handles diverse medical questions without needing task-specific training.

B. Capabilities of LLMs in Medicine

Researchers have thoroughly tested ChatGPT's medical knowledge. Studies show the model can pass the United States Medical Licensing Examination (USMLE) without special training [10]. Other tests prove it can summarize electronic health records, write clinical notes, and explain complex diseases to patients in simple terms [11]. These features

indicate that LLMs can speed up paperwork and improve patient education [4].

C. Safety Constraints and AI Hallucinations

The biggest hurdle to using ChatGPT in healthcare is its core design. As a predictive language model, it guesses the next logical word rather than pulling facts from a secure medical database. This causes hallucinations. In a hospital, a hallucination might look like a wrong drug dosage, a fake medical study, or bad advice for a patient with mixed symptoms [12], [8]. Research confirms that while LLMs are good at suggesting possible conditions, their specific treatment recommendations are unreliable and require a human doctor's review [6], [14].

D. Digital Health in Developing Nations

The technology gap changes how we view AI's risks and benefits. In strong healthcare systems, doctors double-check AI answers against official databases. In resource-limited settings like Afghanistan, these backup systems often do not exist. Studies on digital health in conflict zones show that while internet access is growing, digital health literacy remains low [5], [17], [13]. Trusting unverified AI models in these areas creates unique ethical problems, which is why local surveys are essential.

III. Methodology

A. Research Design and Setting

We used a cross-sectional, quantitative design to gather primary data directly from active medical practitioners. We chose Kabul, Afghanistan, as the study setting because it has a wide mix of public hospitals and private clinics.

B. Population and Sampling

The target group included doctors, surgeons, and specialists working in clinics. We used purposive sampling to get a mix of experience levels and medical departments. We collected 42 valid responses. To participate, respondents had to hold a medical degree, work in a healthcare facility, and have basic experience using AI.

C. Data Collection Instrument

We collected data using a structured questionnaire distributed through physical copies and Google Forms. The survey had two sections.

The first section recorded demographics: age, gender, years of experience, medical specialty, and workplace type.

The second section measured AI performance. Participants graded ChatGPT on four clinical pillars:

1) **Clinical Appropriateness:** How valid the AI's advice is for specific symptoms.
2) **Usefulness to the Patient:** How well the AI explains medical issues to patients.
3) **Safety of the Recommendation:** The risk of harm from following the AI's advice.
4) **Alignment with Clinical Guidelines:** How well the AI matches established medical rules.

Participants scored these metrics on a 5-point Likert scale, from 1 (Strongly Negative/Unsafe) to 5 (Strongly Positive/Safe).

D. Statistical Formalization

To analyze the Likert scores objectively, we aggregated the data mathematically. Let $x_{i,m}$ denote the response of participant i for metric m. The sample mean $\bar{X}_m$ for each metric is:

$$\bar{X}_m = \frac{1}{N}\sum_{i=1}^{N} x_{i,m} \tag{1}$$

where $N = 42$. To measure the agreement or disagreement among the doctors, we computed the standard deviation σ_m:

$$\sigma_m = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_{i,m} - \bar{X}_m)^2} \tag{2}$$

We also defined an overall predictive impact ratio (I_R) regarding the future of the healthcare system:

$$I_R = \left(\frac{C_{improve} + 0.5 \times C_{mixed}}{N}\right) \times 100 \tag{3}$$

where $C_{improve}$ is the number of positive predictions, and C_{mixed} is the number of neutral or conditional predictions.

E. Ethical Considerations

We followed strict ethical protocols. Every questionnaire included an informed consent form explaining the academic purpose of the study. Participation was voluntary, and all data was kept anonymous. We did not record any personal identifiers, such as names or specific hospital locations.

IV. Results

A. Demographic Profile

The demographics of the 42 respondents provide context for our analysis. The sample includes 54.8% male and 45.2% female practitioners. The age data points to a young workforce: 71.4% are between 25 and 35 years old, 16.7% are between 36 and 45, and 11.9% are 45 or older.

Table I shows clinical experience. Exactly 66.7% of the participants have under six years of experience. The group includes specialists in General Practice, Internal Medicine, Surgery, Obstetrics and Gynecology, and Laboratory Sciences.

Table I
Clinical Experience Distribution

Experience Range	Percentage (%)
Less than 1 Year	4.8%
1 – 3 Years	35.7%
4 – 6 Years	26.2%
7 – 10 Years	14.3%
More than 10 Years	19.0%

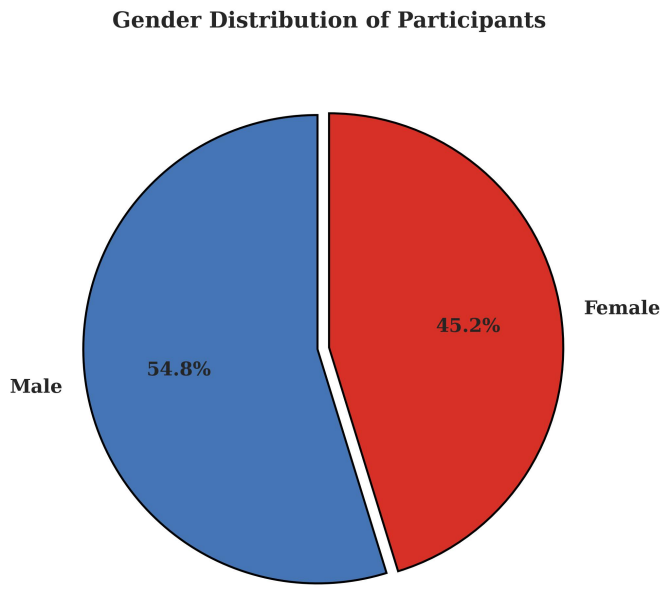

Figure 1. Gender distribution of the surveyed medical professionals.

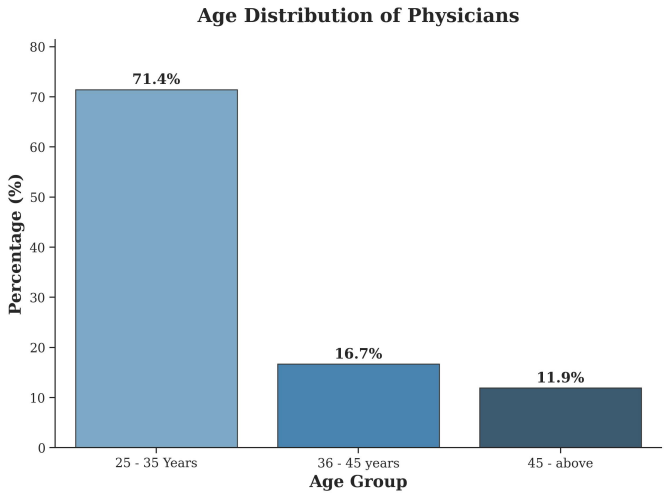

Figure 2. Age distribution highlighting a predominantly young workforce.

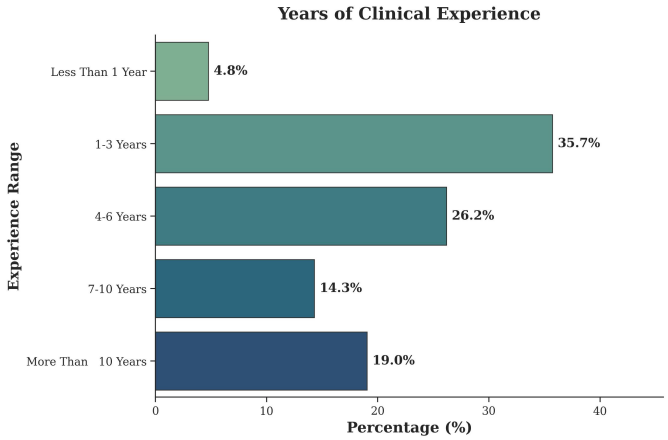

Figure 3. Clinical experience of the participants, showing early-career dominance.

B. Familiarity and AI Usage Patterns

Most sampled physicians were already familiar with ChatGPT. About 38% rated themselves as "Moderately Familiar," 22% as "Very Familiar," and 16% as "Extremely Familiar."

When asked how they use the tool at work, over 80% said they look up general medical information. The next most common uses were interpreting complex lab results and checking drug side effects. A smaller group of general practitioners also uses it to generate initial ideas for patient diagnoses.

C. Evaluation of Clinical Metrics

Our main goal was to test the four clinical pillars. The scores reveal a clear divide between how useful AI is for administration versus how safe it is for patient care.

1) Appropriateness and Usefulness: ChatGPT scored highest in "Usefulness to the Patient." Doctors noted that the AI easily translates difficult medical terms into simple language, making it great for educating patients after a check-up. The model scored moderately in "Clinical Appropriateness." Doctors found it reliable for simple, single-issue illnesses, but noted its advice gets worse when dealing with complex cases involving multiple diseases.

2) Safety and Clinical Guidelines: The "Safety of the Recommendation" metric received the lowest scores and the most disagreement among specialists. Surgeons and internists warned that using AI for surgical plans or complex drug doses is highly risky. They argued that ChatGPT cannot see physical variables like a patient's exact weight or local medical history. Regarding "Alignment with Clinical Guidelines," the model generally follows WHO protocols but often recommends treatments or drugs that are simply unavailable in Afghanistan.

D. Predictive Impact on the Healthcare System

Finally, we asked doctors how AI will impact the Afghan healthcare system overall. Table II shows these results.

Table II
PREDICTED IMPACT OF CHATGPT ON THE AFGHAN HEALTHCARE SYSTEM

Opinion Category	Percentage (%)
Improve (Positive Impact)	59.5%
Both / Mixed (Conditional Impact)	21.4%
Not Sure (Undecided)	14.3%
Harm (Negative Impact)	4.8%

A clear majority (59.5%) believe AI will have a positive impact because it gives doctors quick access to global medical data. However, 21.4% have a mixed view, noting that AI will only help if clinics enforce strict rules on how to use it.

V. DISCUSSION

Our results provide clear evidence about LLMs in resource-constrained clinics. The data shows that young, early-career doctors are the main users of AI in Afghanistan. While senior doctors rely on years of experience, younger doctors use digital

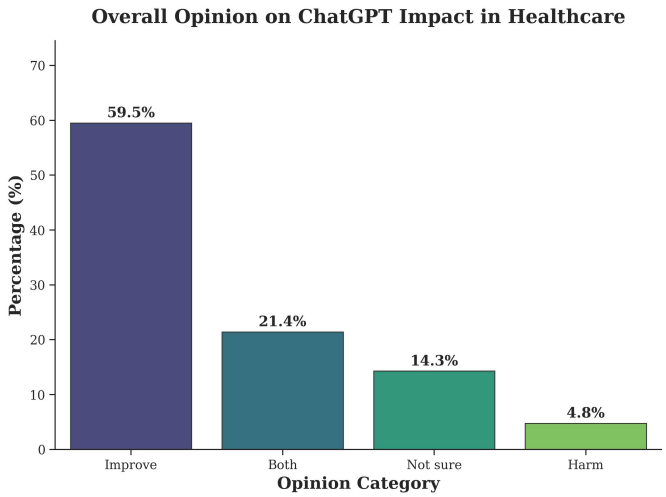

Figure 4. Overall opinion of doctors regarding the future impact of ChatGPT on the Afghan healthcare system.

tools to make up for the lack of mentors and updated medical books in their clinics.

The gap between patient usefulness and clinical safety matches the known limits of current AI models. ChatGPT is an excellent search engine and translator, which makes it perfect for patient education. However, writing prescriptions and diagnosing illnesses require exact facts and an understanding of local limits. The low safety scores from specialists confirm that AI is not ready to make medical decisions on its own. The harder the medical specialty, the less doctors trust the AI.

The 59.5% positive prediction makes sense in context. This optimism comes from the severe lack of information in Afghan hospitals, not from blind trust in the AI. When a doctor cannot easily access a medical journal or call a specialist, a fast AI answer feels like a massive upgrade.

But this creates the risk of automation bias—the habit of trusting a machine's answer and ignoring common sense [16]. If Afghan doctors start accepting AI diagnoses without double-checking them, medical errors will increase. Clinics must manage these tools carefully.

VI. Limitations

This study has a few limits. First, we only surveyed 42 participants, all located in Kabul. Medical work in the capital is very different from rural areas, so our data does not capture the severe tech isolation in remote provinces. Second, we relied on doctors' opinions rather than tracking actual patient recovery rates. Finally, we only tested ChatGPT. Newer AI models built specifically for medicine might score better on safety and accuracy in the future.

VII. Conclusion

This study is the first quantitative analysis of AI use in Afghan healthcare. The data confirms that Afghan doctors actively use ChatGPT to look up medical literature, read lab results, and educate patients. While the AI provides fast access to information, it still has serious flaws regarding clinical safety, precision, and understanding local drug shortages.

Overall, the medical community is cautiously optimistic. To ensure AI helps rather than harms patients in Afghanistan, doctors must use it only as a backup tool, not a replacement for medical judgment. Healthcare leaders and the Ministry of Public Health need to create clear, standard guidelines on how to use AI safely in local clinics. Looking forward, we plan to expand this research to public hospitals across all provinces to build a national framework for digital health in Afghanistan.

References

[1] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019.

[2] K. Singhal *et al.*, "Large language models encode clinical knowledge," *Nature*, vol. 620, pp. 172–180, 2023.

[3] A. J. Thirunavukarasu *et al.*, "Large language models in medicine," *Nature Medicine*, vol. 29, no. 8, pp. 1930–1940, 2023.

[4] M. Sallam, "ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns," *Healthcare*, vol. 11, no. 6, p. 887, 2023.

[5] World Medical Program, "Healthcare in Afghanistan: Systemic challenges and the need for digital integration," *Global Health Metrics*, Tech. Rep., 2021.

[6] Z. Zhao *et al.*, "Evaluating the safety of ChatGPT in clinical decision support," *NPJ Digital Medicine*, vol. 7, p. 23, 2024.

[7] D. Dash, A. Sharma, and P. Singh, "Evaluation of ChatGPT in medical diagnosis: Accuracy, hallucinations, and safety implications," *Artificial Intelligence in Medicine*, vol. 143, p. 102613, 2023.

[8] A. Haupt and M. Marks, "AI-generated medical advice—GPT and beyond," *JAMA*, vol. 329, no. 16, pp. 1349–1350, 2023.

[9] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in health and medicine," *Nature Medicine*, vol. 28, no. 1, pp. 31–38, 2022.

[10] T. H. Kung *et al.*, "Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models," *PLOS Digital Health*, vol. 2, no. 2, p. e0000198, 2023.

[11] J. Liu, H. Wang, and S. Chen, "Evaluating the utility of ChatGPT for patient education and medical information retrieval," *Journal of Medical Internet Research*, vol. 25, p. e4567, 2023.

[12] B. Mesko, "The impact of AI on the future of healthcare and the ethical considerations," *Digital Medicine*, vol. 6, no. 1, pp. 102–115, 2023.

[13] A. Hussain and M. Ali, "The digital divide in global healthcare: AI adoption in developing nations," *International Journal of Health Informatics*, vol. 12, no. 4, pp. 210–225, 2023.

[14] S. Chen *et al.*, "Limitations of large language models in specialized medical domains," *Journal of Clinical AI*, vol. 4, pp. 45–59, 2023.

[15] R. Ahmed, "Healthcare infrastructure and policy challenges in post-conflict regions," *Global Public Health Review*, vol. 18, no. 2, pp. 134–148, 2022.

[16] V. Patel and L. Smith, "Automation bias in medical diagnostics: The psychological impact of AI assistance," *Medical Decision Making*, vol. 43, no. 5, pp. 560–572, 2023.

[17] S. A. H. Zaidi *et al.*, "Adoption of Artificial Intelligence in Healthcare of Developing Countries," *Journal of Global Health*, vol. 14, p. 04052, 2024.

Mohammad Reza Cov Andish received his B.Sc. degree in Computer Science from Kabul University in 2026. His research interests include lifelong machine learning, natural language processing, deep learning, semantic modeling, and continuous learning frameworks for intelligent systems.

Mustaffa Muradi received the B.S. degree in Information Systems from Kabul University in 2026. His research interests include artificial intelligence in healthcare, data analysis, medical decision support systems, and the ethical integration of digital health technologies in developing nations.

Ahmad Srosh Hamdard received the B.S. degree in Information Systems from Kabul University in 2026. His research interests include health informatics, natural language processing, the evaluation of large language models in clinical settings, and intelligent systems.