THE AI SAFETY GAP: A SCENARIO-BASED STUDY OF RISKY HEALTH DECISIONS AND MISPLACED TRUST

Adriatik KRASNIQI[1], Amira HUSEINI[1], Rihan SAITI[2]

1 Faculty of Medical Sciences, General Medicine
2 Department of Cardiology, Faculty of Medical Sciences
University of Tetova, Republic of North Macedonia

Abstract

Artificial intelligence (AI) is becoming increasingly integrated into healthcare and everyday life, offering rapid and accessible medical information to the public. Despite its benefits, concerns remain regarding the accuracy of AI-generated medical advice and the potential risks associated with excessive reliance on these technologies without professional medical consultation. Understanding how individuals perceive and use AI in healthcare is important for promoting safer digital health practices. This study aimed to evaluate public trust in AI-generated medical advice and explore attitudes and behaviors related to AI use in healthcare decision-making. A cross-sectional survey was conducted among 150 participants from the general population. The questionnaire assessed demographic characteristics, frequency of AI use, trust in AI-generated health information, and healthcare-related decision-making behaviors through hypothetical clinical scenarios. The results demonstrated moderate overall trust in AI-generated medical advice, with frequent use of AI tools reported among participants. While most individuals preferred physician guidance when presented with conflicting medical opinions, a proportion of participants still demonstrated uncertainty or willingness to rely on AI recommendations in potentially serious situations. Participants also expressed strong support for clearer safety measures, physician consultation recommendations, and greater public education regarding the responsible use of AI in healthcare. These findings highlight both the growing role of AI in healthcare decision-making and the importance of ensuring safe and responsible use of these technologies. Strengthening public awareness and implementing safety-oriented AI guidance may help reduce inappropriate reliance on AI-generated medical advice and improve patient safety.

Introduction

The rapid integration of artificial intelligence (AI) into everyday health decision-making raises concerns about its misuse, particularly in high-risk situations. While AI tools offer accessible health information, inappropriate reliance without medical verification may lead to harmful outcomes. This study aims to assess public trust in AI-generated medical advice, identify high-risk behaviors, and explore attitudes toward safer AI use in healthcare.

Materials and Methods

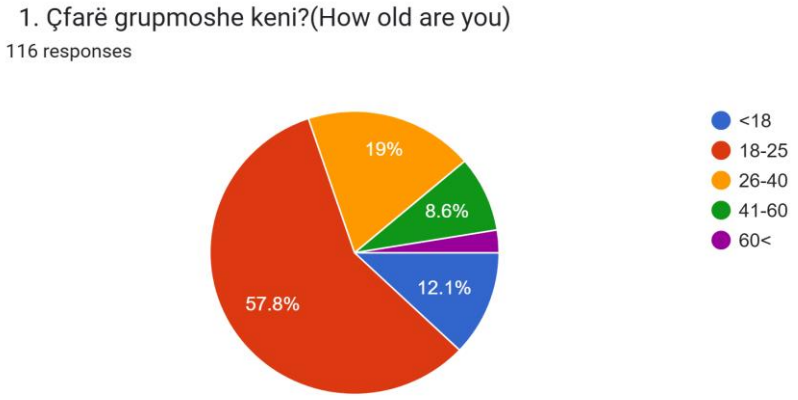
A cross-sectional, scenario-based survey was conducted with 150 participants from the general population. The questionnaire included demographic data, AI usage patterns, trust levels (Likert scale), and verification behaviors. Two clinical scenarios (low-risk: mild headache; high-risk: chest pain with dyspnea) were used to evaluate trust and intended actions. Additional questions explored reasons for avoiding medical consultation and attitudes toward AI safety measures.

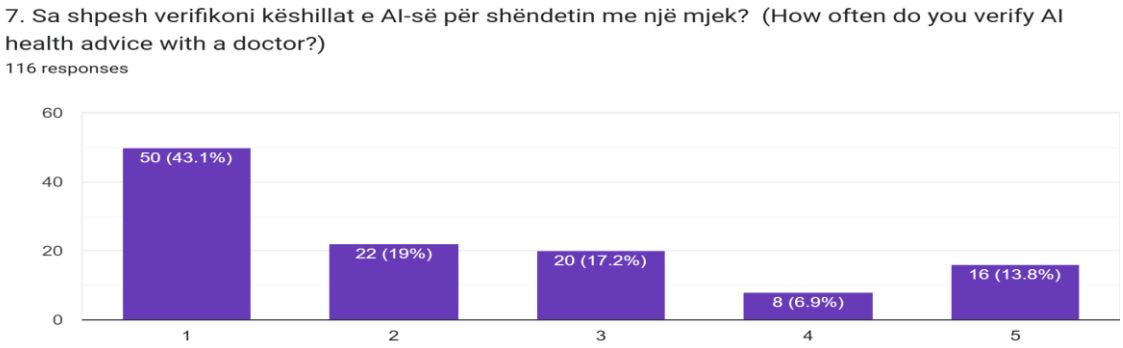
Results

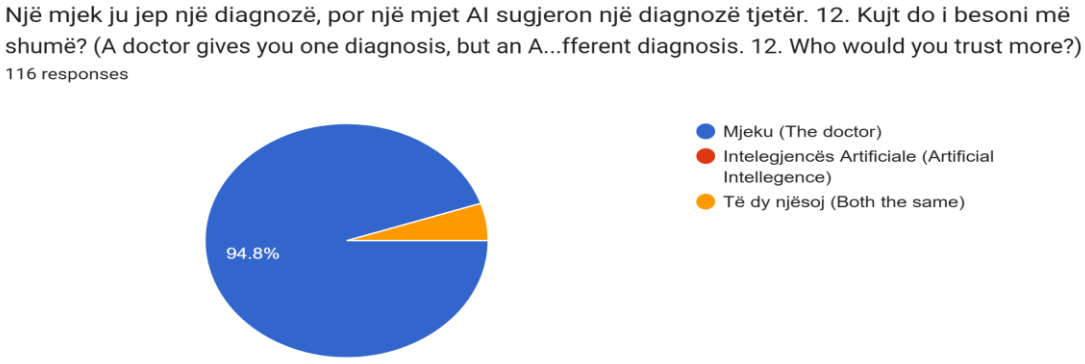
Most participants were aged 18–25 (57.8%) and female (57.8%), with 48.3% holding a university degree. AI usage was frequent, with 47.4% reporting high usage (Likert 4–5). Trust in AI for accurate medical advice was moderate (41.4% neutral), but verification with

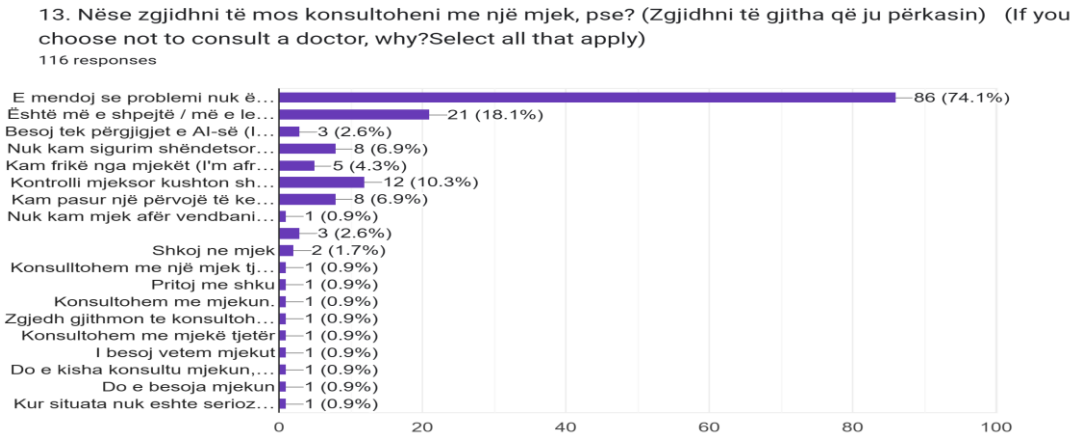
physicians was low (43.1% rarely verify). In the low-risk scenario, 37.1% showed high trust (Likert 4–5), and 52.6% would follow AI advice. In the high-risk scenario, trust significantly decreased, with 75.8% expressing low trust (Likert 1–2); however, 14.7% would still follow AI recommendations and 32.8% remained uncertain, indicating potentially unsafe decision-making. When AI contradicted a physician, 94.8% trusted the physician. The main reason for avoiding medical consultation was perceiving the problem as non-serious (74.1%), followed by convenience (18.1%) and cost concerns (10.3%). Notably, 94% supported clearer safety warnings, 89.7% believed AI should always recommend consulting a physician, and 87.1% emphasized the importance of public education on safe AI use.

Table figure and equations









Conclusion

Despite moderate trust in AI, a subset of individuals demonstrates risky behavior by relying on AI in potentially serious health situations without verification. These findings highlight the need for stronger safety frameworks, including mandatory warnings, physician-directed guidance, and public education. Improving user awareness and integrating safety-oriented AI design may reduce inappropriate reliance and enhance patient safety.

Keywords

Artificial Intelligence; Patient Safety; Healthcare Decision-Making; Trust; Digital Health