6-14-2024

Generative AI In Medical Data Reliability Analysis: Limb At Risk Clinic Use Case

Liat Todris
Ariel university, liat.toderis@msmail.ariel.ac.il

Iris Reychav
Ariel University, irisre@ariel.ac.il

Roger McHaney
Kansas State University, mchaney@ksu.edu

Bernice Oberman
Sheba medical center, bernice.oberman@sheba.health.gov.il

Chen Speter
Sheba medical center, chen.speter@sheba.health.gov.il

GENERATIVE AI IN MEDICAL DATA RELIABILITY ANALYSIS: LIMB AT RISK CLINIC USE CASE

TREO Talk Paper

Liat Toderis
Ariel University
liat.toderis@msmail.ariel.ac.il

Iris Reychav
Ariel University
irisre@ariel.ac.il

Roger McHaney
Kansas State University
mchaney@ksu.edu

Bernice Oberman
Chaim Sheba Medical Center
bernice.oberman@sheba.health.gov.il

Ronen Loebstein
Chaim Sheba Medical Center & Tel-Aviv University
Ronen.Loebstein@sheba.health.gov.il

Chen Speter
Chaim Sheba Medical Center
chen.speter@sheba.health.gov.il

Abstract

This research evaluates OpenAI's ChatGPT-4 in medical innovation. Specifically, we focus on a use case related to diabetic foot care. It was conducted at the Sheba Medical Center's Limb at Risk Clinic in Israel. The study evaluates generative AI's ability to accurately answer predictive questions constructed by clinicians with outcomes compared to those generated with traditional statistical analysis. The project was organized within an architectural framework, and aligned with enterprise goals of making information more accessible without requiring services of a data expert. It utilizes the MDClone® platform as part of a broader effort transitioning to an institution-wide, data-driven approach to medical practice. Results indicate agreement between ChatGPT-4 used by medical clinicians and a statistician's analyses. Outcomes show AI can be effective in answering complex questions within a closed dataset. The study also suggests that non-medical end-users can successfully leverage the generative AI for reliable information in a controlled environment.

Keywords: Medical innovation, Predictive modeling, Data-driven approach, Generative AI.

1 Introduction

OpenAI's ChatGPT-4 offers incredible potential in many areas related to medical innovation. Among these are aiding with patient discharge notes, summarizing clinical trials, synthesizing research articles, informing medical ethics practices, and predicting personalized medicine applications. Additionally, AI tools can generate virtual patient simulations, critique resulting doctor patient communications, and review data sets (Eysenbach 2023; Hasanzad et al. 2023; Waisberg et al. 2023). In all instances, proper prompting within an organizational framework is crucial to prevent errors and missteps. The current research investigates a use case derived from custom data related to challenges in diabetic foot care, which can result in amputation or other undesirable outcomes. Sheba Medical Center's Foot Care Risk Center in Israel assessed use of AI technology to predict and prevent amputations and enhance related data-driven decision-making. Specifically, ChatGPT-4 was employed to answer questions related to treatment options and outcomes in the Limb at Risk Clinic within an existing organizational framework and compared to a traditional analysis approach to determine its accuracy.

1.1 Enterprise goals and approach

The current project was structured using organizational architecture principles with the ArchiMate® language and TOGAF standards to ensure a comprehensive and orderly methodological approach. From a broad perspective, the organization had a strategic goal of making information accessible to end users. This goal was operationalized at each level within the architecture ultimately yielding a chat application to satisfy business needs with the capability of addressing medical needs. This was done at The ADAMS Center at the Chaim Sheba Medical Center, Israel, with the MDClone® platform as part of the center's goal of transitioning to an institution-wide, data-driven approach to medicine.

2 Methods

The goal of evaluating ChatGPT-4's ability to correctly answer predictive questions regarding patients at Sheba Medical Center's Limb at Risk Clinic began with data acquisition. Data representing visits between October, 2021 and June, 2023, were retrieved anonymously from the Electronic Medical Record (EMR) system using the MDClone® platform (MDClone 2023) in a way that adheres to the principles outlined in the Declaration of Helsinki. This ensures ethical compliance with established guidelines for medical research. Independent variables such as age, gender, # of involved limbs, background diseases, diabetes, smoking, dialysis, Charlson score and mortality were collected. Treatment of patients looked at number of visits, vascular problem scores and # of catheterizations. Outcomes included patient mortality, limb status (amputation) and wound status. A set of 4 questions were defined for two professional parties. One was a statistician and the other, a technology professional with experience writing prompts for Chat GPT. Both performed their analyses on the data and then compared the results. These questions were investigated:

1. Who are the patients with a higher risk of dying within a year from the first visit?
2. Is there an association between the number of angiographies and the risk of dying within the first year?
3. What characterizes the patients who came to the clinic with worst admittance scores?
4. Is there an association between the number of clinic visits and the chance of major (above or below knee) Amputation free survival in the first year?

The statistician collated all relevant (personal, wound, limb) data to create a single record per person in the study. In case of several wounds the worst was used. Tables comparing differences in outcomes (e,g. death within 1 year, presence of gangrene, and amputation within 1 year) were constructed using Chi-squared tests for differences between categorical variables and t-tests (when Normally distributed) or a-parametric tests (when not Normally distributed) for continuous variables. A multivariate logistic regression model was constructed for each outcome using significant variables.

The Generative AI approach relied on prompts which the tool internalizes and then answers based on its access to the data. Creating meaningful prompts required trial and error. Careful thought, therefore, needs to go into the construction of these prompts to ensure relevant answers emerge.

3 Results and discussion

ChatGPT-4 and the statistician’s investigation of the data framed around the four guiding questions produced similar results. This indicates that non-medical end-users were able to ask questions in natural language and receive reliable answers using a closed dataset within a generative AI environment. This finding was based on two general assessments. First, we examined whether the AI would accurately respond to queries within a closed medical dataset without producing misleading information. Second, can individuals from a medical background, lacking expertise in computer science, effectively inquire about their data using natural language? The study’s outcomes both confirmed the AI's potential for medical professionals to independently engage with their data with reduced reliance on technical experts. However, the importance of carefully designed, natural language prompts are needed to ensure accurate and non-generalized responses. For example, caution must be exercised when asking certain questions such as those related to correlation between variables. Failure to do so may result in ambiguous answers necessitating further exploration and clarification. Therefore, question formulation is key and, in some instances, questions are better explored through dialog resembling a conversation. Table 1 provides the outcomes for each approach.

Question	Statistician Answer	ChatGPT Answer
1 Patients at higher risk of dying within a year	Older age, more clinic visits, and higher Charlson score	Age and Charlson score are moderately correlated with death
2 Association between angiographies and risk of dying within 1 year	Non-significant negative association	Non-significant negative association
3 Characteristics of patients who came to the clinic with worst admittance scores	Older, more angiographies, and less recoveries	Higher age, more clinic visits, higher Charlson score for worst admittance group
3 Association between clinic visits and major amputation-free survival	No statistical evidence of significant association	Clinic visits do not significantly influence limb status severity

Table 1. Outcome Comparison.

Overall, this study was successful. It indicates an end user without database or technical expertise has the ability to ask meaningful, natural language questions within a closed, generative AI environment, and receive reliable answers comparable to those generated with a traditional statistical analysis. Furthermore, this study suggests that teaching medical clinicians to formulate meaningful prompts is more feasible than instructing them to become experts in data cleaning and processing. We recommend that further research be conducted to substantiate our findings. Limitations of the current project relate to use of a single case in a particular medical facility. Therefore, further tests on additional medical cases from a variety of medical facilities should be examined. Likewise, those seeking answers should experiment with a series of complex questions and conversations within their AI environment. We also recommend that different doctors and medical personnel formulate the same question to see if differences in answers emerge.

References

Eysenbach, G. (2023). The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. JMIR Medical Education, 9(1), e46885.

Hasanzad, M., Patrinos, G., Sarhangi, N., Sarrami, B., & Larijani, B. (2023). Using ChatGPT to Predict the Future of Personalized Medicine.

Waisberg, E., Ong, J., Masalkhi, M., Kamran, S. A., Zaman, N., Sarker, P., ... & Tavakkoli, A. (2023). GPT-4: a new era of artificial intelligence in medicine. Irish Journal of Medical Science (1971-), 1-4.

MDClone. MDClone ADAMS Platform. The Heathcare Data Platform Website, https://www.mdclone.com/adams-platform (2023, accessed 18 August 2023)

Acknowledgement

Luz Erez, MDClone for helping with Chat GPT.