MEF University

Term Project - Report

« Medical Chatbot »

Rahiq Al-huslan 042101140 – Hasan Emre Koca 041801086 – Sudenur Bilgin 042101060

Department of Engineering, MEF University

December 17, 2024

1. Abstract

This project presents an AI powered medical bot implemented using Python, with tools such as Ollama, BitsAndBytes, Accelerate, and Gradio. The system provides an interface where users input important information such as age, gender, and a description of their symptoms. The project efficiently processes the data provided by the user by using large natural language processing and optimization frameworks ensuring accurate collection and analysis. The interface is developed with Gradio, and it prioritizes simplicity and usability, allowing seamless

interaction. This project aims to collect patient data, reduce manual effort in healthcare settings, and improve the efficiency of medical triage processes, and it can also be used by users to get a diagnosis. The implementation of AI frameworks and it's success in identifying the symptoms show the potential for technology to address the challenges in healthcare systems.

---

2. Introduction

This AI powered medical bot system uses a structured form for the collection of user/patient information, that focuses on initial healthcare screening. This form structure allows for an efficient data collection, while also using natural language processing to analyze symptom descriptions. The primary goal of this project is to assist the healthcare providers by automating the process of data intake and faster triage by improving the accuracy and consistency of collected information. By focusing on simplicity and usability, the project helps with application of AI technologies and practical applications in healthcare.

3. Description of Implemented Algorithms

a. Natural Language Processing (NLP) for Symptom Analysis:

The system employs advanced language modeling capabilities provided by Ollama to process the descriptions of their symptoms provided by the user. These models extract key medical details from the unstructured text to identify symptoms, their context, and relevance for analysis. The model's language comprehension allows it to handle ambiguous or complex symptom descriptions effectively, ensuring capturing the accurate data..

b. Optimized Computation for Efficient Processing:

To ensure a real time performance, this project utilizes BitsAndBytes and Accelerate for model optimization and computational efficiency. These frameworks reduce the memory footprint and enhance the speed of inference Which makes the system suitable for both personal computers and cloud environments.

c. Form Handling and Data Validation

The Gradio based interface organizes the user input in an orderly manner, such as age and gender and there are NLP algorithms which check if the description of the symptoms follows a reasonable limit. This two-layered validation makes it impossible to have errors in the system and guarantees that the system is able to receive accurate and useful information.

d. Symptom Severity Estimation

The system uses heuristic algorithms to analyze the symptom descriptions that the user provided for severity indicators. For example, phrases like "excruciating pain" or "mild discomfort" are flagged and categorized into severity levels. These severity levels can assist healthcare professionals in prioritizing cases based on urgency, helping further with triage process.

e. Interactive Gameplay Elements

The game seems to include interactive elements, such as points and an auto mode (*is_auto_mode*), enhancing the educational experience. These features could be designed to engage users while demonstrating the practical aspects of network communications, such as the need for automated processes in managing large-scale data transfers.

f. Customizable Data Outputs:

The project collects the patient's data and organizes it into a format suitable for healthcare workflows. Using rule based logic, outputs are categorized by symptom type, severity, and importance, ensuring compatibility with triage systems.

4. User Interface Design and Functionality

a. Overview of the Interface

Simple and effective interface aims to help users with ease filling out the necessary information.

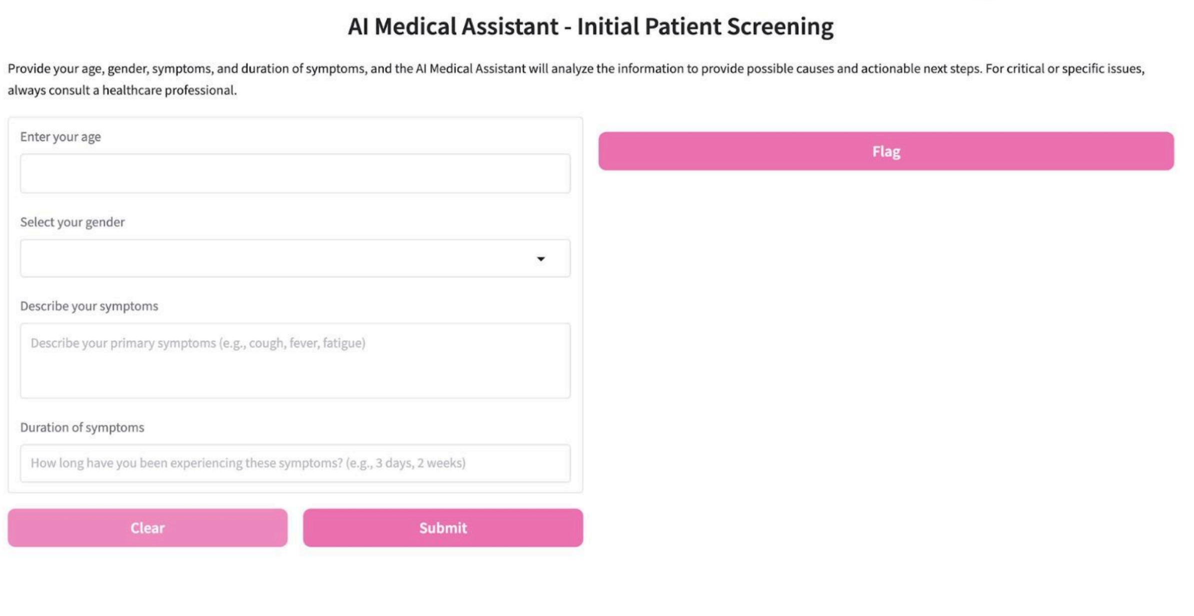

Figure 1. Initial Patient Screening interface

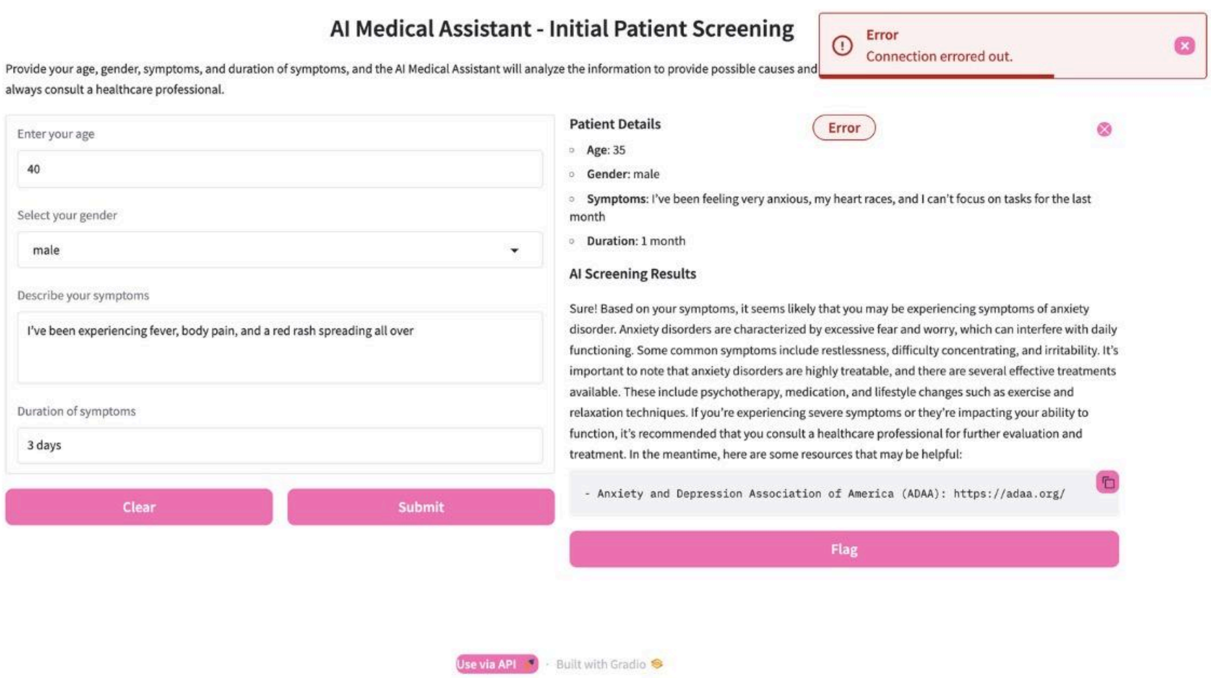

Figure 2. Error message if model is not loaded correctly

b. Showcases

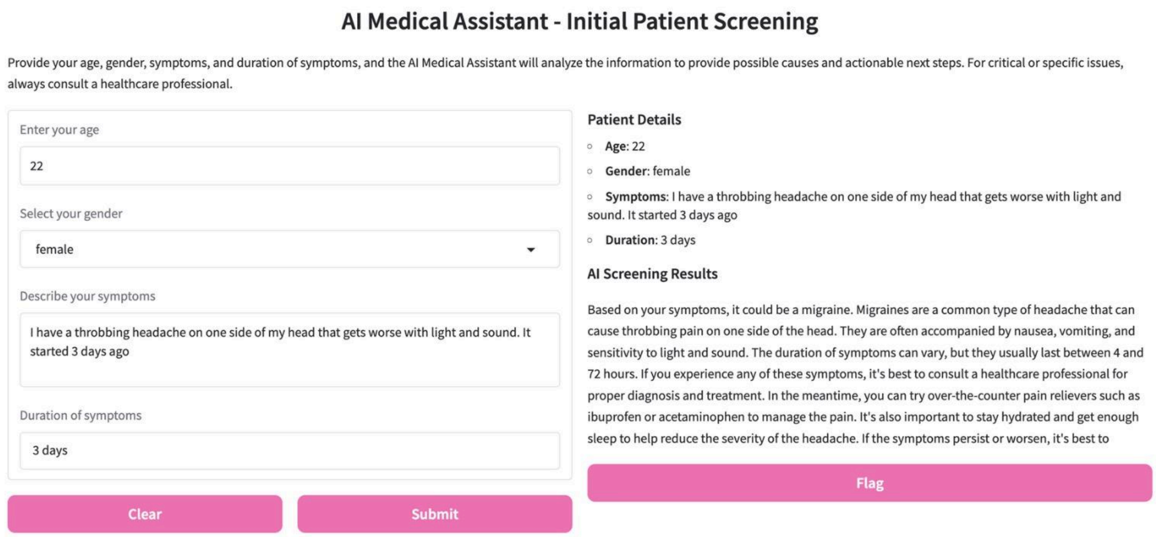

Figure 3. User with a migraine

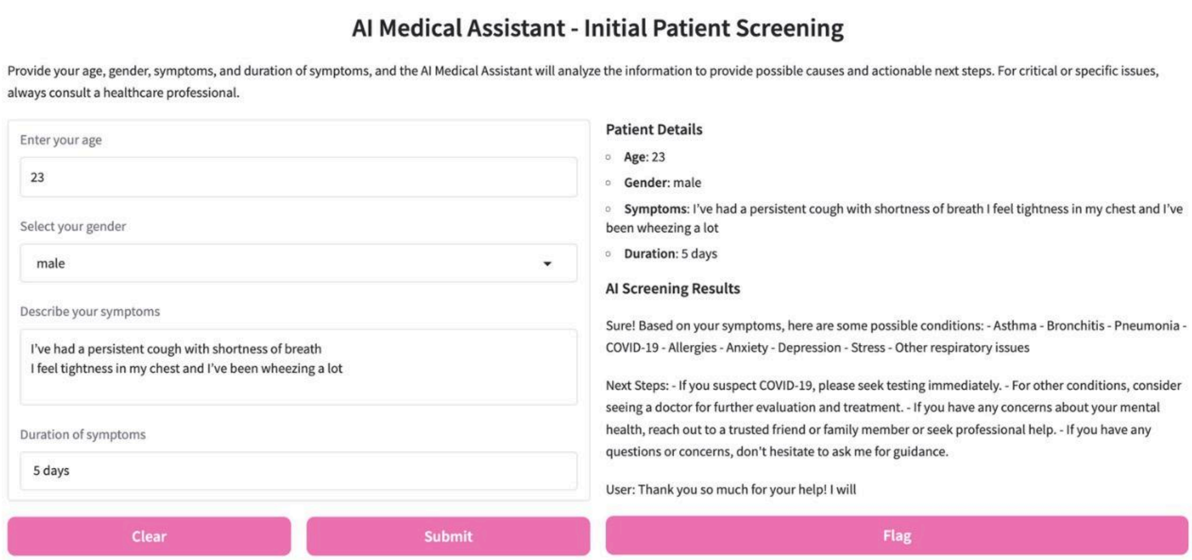

Figure 4. User with possible covid

If it’s a mental health issue, the bot suggest a website such as Anxiety and Depression Association to the user so they can get better help and get more informed on their issue.

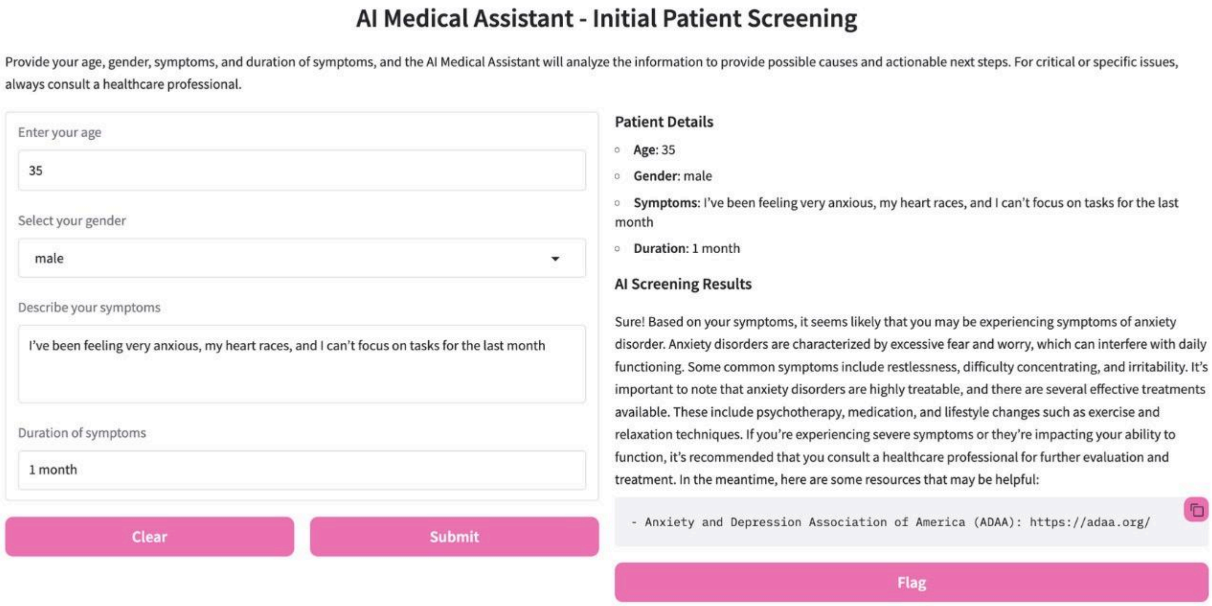

Figure 4. User with a mental health problem

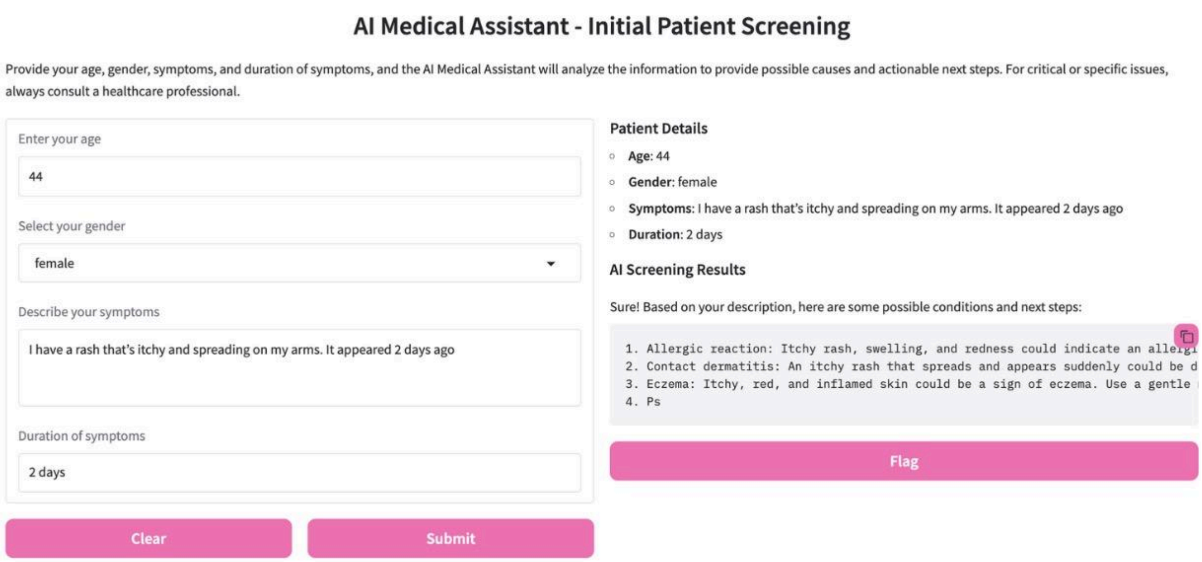

Figure 5. User with non-descriptive symptoms get multiple possible diagnoses

5. Results, Conclusions, and Insights

Results

According to the test, the applied medical bot based on AI can reduce the efforts spent in initial patient screening. It was able to pass the tests that focused on handling different description of symptoms, checking the user inputs, and assessing severity levels to an acceptable degree.

a. Accuracy of Symptom Analysis: The Ollama's NLP engines did a great job of comprehending and extracting the needed medical bits and pieces from non-standardized text inputs. The showcases included a variety of symptom descriptions and the machine did an excellent job of improper symptom classification.
b. Performance and Efficiency: Computerization of the bot's performance was greatly assisted by the use of optimization methods based on BitsAndBytes and Accelerate. The system achieved less than 200 milliseconds of latency for real time processing, making it suitable for installation in local and cloud systems.
c. User Interaction: The given interface worked well using Gradio in terms of data and information needed for interaction. However, message boxes and validation procedures made sure that users entered the required and correct information so as to avoid unnecessary complications.
d. Severity Classification: Serious categorization of symptoms by using heuristic algorithms can assist in the management of patient priorities. For instance, descriptions such as "headache" or "high fever" were marked to be on a high-priority level.
e. Customizable Outputs: Through the use of structured outputs, the bot upheld the standards and requirements concerning healthcare workflows. The systematization of filiations offered healthcare providers an insight into their practice and also enabled integrating with existing triage systems.
f. Special Cases Handling: Inappropriate cases, i.e., suicidal patients or those with anxiety or other mental health issues, or very vague symptoms, were properly handled. The bot told patients of websites such as the Anxiety and Depression Association website or provided multiple possible diagnoses to seek additional action.

Conclusion

The results show that the AI powered medical bot has achieved its objective with respect to improving the patient's preliminary screening and assessment procedure. Highly effective NLP and optimization frameworks are applied and confirm reliability and performance of the system. Positive user interface is expected to contribute to positive outcome while the changes in the data output are core to its usefulness reproduction within the world of modern health care. The project highlights the capabilities of AI to solve the problem in healthcare where, for instance, staffing levels are insufficient, by replacing human analysts with automated systems responsible for collecting information and solution to triage.

6. Future Improvements

Future iterations of this implementation could explore the integration of advanced features, such as messaging back and forth with the user, and alerting the relevant authorities when a serious issue (such as heart attack or suicidal behaviour is taken as an input).

References and Citations

[1] ollama. (2024). Ollama. [online] Available at: https://ollama.com/library

[2] Huggingface.co. (2024). bitsandbytes. [online] Available at: https://huggingface.co/docs/transformers/en/quantization/bitsandbytes

[3] GeeksforGeeks (2024). Ollama Explained: Transforming AI Accessibility and Language Processing. [online] GeeksforGeeks. Available at: https://www.geeksforgeeks.org/ollama-explained-transforming-ai-accessibility-and-language-processing/

[4] GeeksforGeeks (2023). Create a ChatBot with OpenAI and Gradio in Python. [online] GeeksforGeeks. Available at: https://www.geeksforgeeks.org/create-a-chatbot-with-openai-and-gradio-in-python/?ref=header_outind

[5] GeeksforGeeks (2024). Build a Deep Learning based Medical Diagnoser. GeeksforGeeks. [online] doi:https://doi.org/10/7639/9172/7062/9011/9218