DoctorHive

Project Proposal

Project Advisor:

Ms. Mariam Nasim

Group Members:

Abdul Moiz Shehzad	22L-7468
Muhammad Fahad Hussain	22L-7463
Muhammad Reyyan Saeed	22i-1964

National University Of Computer and Emerging Sciences
Department of Computer Science
Lahore, Pakistan

Abstract

There has been significant advancement in Artificial Intelligence in providing accessible medical information through the use of conversational chatbots and virtual agents. These systems are becoming more and more integrated into many platforms, offering users with instant responses related to health care queries and providing preliminary diagnostic suggestions. However, the reliability of these systems is a major concern. AI generated medical responses are often inaccurate and may present misleading conclusions that can create emotional distress for patients [1], [2]. For instance, searching simple symptom may lead to alarming diagnosis such as cancer, even when the likelihood of it is low. Unlike programmers who can validate and correct AI generated code, patients lack the expertise in medical domain to evaluate whether the information they receive is accurate, safe and is in the right direction or not.

The goal of this research is to create a multi agent AI system that simulates a team of cooperative medical specialists. In contrast to the conventional method of a single agent, DoctorHive gives several agents with different specialties, the independence to examine and react to patient data, as well as to consult and collaborate with specialized agents to reach a refined and better decision [3]. In conclusion, different agents collaborate to produce the final response. Each agent in their respective medical domains provide diagnosis and then produces a consensus based output highlighting the most probable diagnosis and recommending the next steps and procedure. By simulating expert led discussions for patient cases, DoctorHive aims to improve the safety of AI in healthcare, reduce the likelihood of false information, and increase the accuracy of information given to patients.[4][5].

1. Introduction

There is rapid growth in the use of AI in healthcare. Chatbots and Large Language Models provide quick access to medical diagnosis and advice [1]. While these tools are convenient, serious concern remains on their reliability. In most of the cases, AI generated responses can be inaccurate, overly extreme, or lack the complexity needed for clinical reasoning [2]. Patients who search for their symptoms online often face the worst case scenarios that cause unnecessary anxiety. Similarly, while LLM models such as GPT can provide medically relevant responses, they many times overlook important contextual details or generate misleading explanations [4].

The challenge is that non experts in healthcare lack the necessary domain knowledge to appraise an AI generated response for accuracy [5]. Users in programming fields can identify mistakes, evaluate AI outputs and flag issues for revision. In healthcare, however, patients have difficulty distinguishing plausible suggestions from dangerously incorrect ones. This calls for the need for systems that enable responses to be generated through the collaboration of multiple respondents.

The multi agent debate system being implemented in DoctorHive is addressing this problem. Through the help of the specialized AI agents in the areas of General Practice, Neurology, Cardiology, and Ophthalmology, each patient submission is treated separately. Such agents then discuss between themselves, point out gaps or inconsistencies and reach a global decision of a ranking of possible consequences [3]. The system simulates the behavior of a real medical team, and the results it generates are more balanced, reliable and transparent.

2. Goals and Objectives

The main objective of this project is to recreate a guided conversation among multiple experts agents to enhance the validity and reliability of AI produced medical answers. The objectives of the project are:

1. To create and deploy the system that will enable multiple specialized AI agents to independently scan the input of patients in their respective fields including, but not limited to, voluntary test reports and symptoms.
2. To create a debate mechanism that would enable these AI agents to dispute, improve or correct the answers of other agents [3].
3. To deliver a response to patients that is clear, concise and contains the most likely diagnosis and next step.
4. To facilitate the interpretability of clinical reasoning outputs to the user by presenting more than one specialist viewpoint, rather than one authoritative but possibly false answer [4].
5. To further understand how Retrieval Augmented Generation can be used by AI agents in the future to incorporate external medical knowledge bases, enhancing evidence based discussions and increasing overall quality and reliability of diagnosis [1].

3. Scope of the Project

DoctorHive will develop a proof of concept system demonstrating the virtues of multi agent cooperation in the generation of medical advice. The system will have the services of expert agents like a General Physician, Cardiologist, Neurologist, and Ophthalmologist, which will independently examine patient inputs, discuss their results, and give users the most likely diagnoses, and suggested next actions. The project, nevertheless, is not meant to serve as a substitute for professional advice. Advanced treatment planning, prescription writing, and interfacing with electronic health records are still beyond scope, as the objective is to create a stand alone prototype.

4. Initial Study and Work Done so Far

Preliminary research highlights that while conversational AI has made medical knowledge more accessible, it also poses significant risks due to misinformation and emotional misguidance [2], [5]. Kang [1] show that patients searching online symptoms often misinterpret results, leading to unnecessary anxiety and, in some cases, delayed medical consultation. LLMs have shown promise in medical applications, but their single response mechanism lacks safeguards against error propagation [4].

A recent research by Arora [3] into multi agent debate frameworks has demonstrated that when AI models are allowed to critique and refine each other's reasoning and responses, the overall reliability and quality of outputs improve significantly. DoctorHive builds upon this idea by applying it specifically to the medical domain, where the consequences of misleading responses are particularly serious. The system aims to combine the accessibility of AI chatbots with safety of multi perspective reasoning, offering a more trustworthy tool for patients seeking guidance. In line with this, we have also designed the initial pipeline of the application as a first conceptual version of how the system could be achieved.

References

[1] J. Kang, S. Zhang, and L. Chen, “A Survey of Large Language Models for Healthcare: Applications, Opportunities, and Challenges,” *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 5, pp. 776–792, Oct. 2023.

[2] S. Ji, T. Xu, H. Wang, and L. Shi, “Can ChatGPT Diagnose Medical Conditions? A Comparative Study on Diagnostic Accuracy with Physicians,” *IEEE Access*, vol. 11, pp. 65612–65623, 2023.

[3] K. Arora, A. Rani, and R. Kumar, “Mitigating Hallucinations in Generative AI Models Through Multi-Agent Debate,” in *Proc. IEEE Int. Conf. on Big Data (BigData)*, Dec. 2023, pp. 1452–1461.

[4] T. Xie, H. Li, and J. Zhao, “Reliability and Trustworthiness of AI-Generated Content: A Case Study with GPT Models,” in *Proc. IEEE Int. Conf. on Trust, Security and Privacy in Computing and Communications (TrustCom)*, Nov. 2022, pp. 890–897.

[5] M. Ali, F. Ahmed, and U. Farooq, “AI-Driven Medical Chatbots: Risks, Benefits, and the Role of Human Oversight,” *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 45–59, Jan. 2023.