DocCHA: Towards LLM-Augmented Interactive Online diagnosis System

Anonymous ACL submission

Abstract

Large Language Models (LLMs) have demonstrated great capabilities in addressing many application tasks. Among various applications, one eye-catching domain is Conversational Health Agents (CHAs), which are interactive conversational systems that can provide people with various health-related services. However, existing CHAs mainly focus on providing static health services, and lack interactive online diagnosis for patients. In the clinical setting, the initial symptoms that patients provide may lack comprehensiveness and detail, thus online interaction with patients to request additional information is important. To alleviate this problem, we propose DocCHA, an online interactive diagnosis system that interacts with patients by requesting additional information and continuesly improving the diagnosis confidence until providing patients with a reliable diagnosis. Moreover, DocCHA leverages Retrieval Augmented Generation (RAG) with Google search API, StatPearls[1] and Wikipedia[2] to provide patients with detailed and reliable health suggestions. We evaluate DocCHA's performance on the IMCS21 dataset, a Chinese online diagnosis dataset consists of conversations between patients and doctors. Experimental results show that DocCHA's diagnosis accuracy reaches 89.2% with 4 rounds of additional information request interactions with patients. Besides, the generated suggestion after RAG outperforms the direct prompt in terms of relevance, coherence, accuracy and completeness.

1 Introduction

Conversational agents work by disassembling a difficult task into multiple incremental, smaller tasks in an explainable way (Laranjo et al., 2018), thus helping to resolve the overall task. With the recent development of Large Language Models (LLMs), applying LLMs to Conversational agents has brought a new potential to various areas, including education (McTear, 2022), climate (Vaghefi et al., 2023) and health (Montenegro et al., 2019). LLM-based Conversational Health Agents (CHAs) focus on various healthcare services, including providing personalized health data analysis (Abbasian et al., 2023), evaluation of the robustness of health data extraction (Chen et al., 2023), health report generation (Liu et al., 2023), etc. Based on methodology, LLM-based CHAs mainly lie in three categories, general-purpose LLM CHAs, specialized LLM CHAs, and multi-modal LLM CHAs. General-purpose LLM CHAs apply existing LLMs to resolve health-related problems by prompting LLMs to realize different sub-tasks (Biswas, 2023; Aydın and Karaarslan; Liu et al., 2023; Chen et al., 2023). Specialized LLM CHAs pre-train or fine-tune transformers with domain-specific health-related data to alleviate the domain shift (Han et al., 2023; Li et al., 2023; Luo et al., 2022; Tu et al., 2024). Multi-modal LLM CHAs integrate multi-modality data, including images, videos and time series, into LLMs to realize health functions (Tu et al., 2024; Xu et al., 2023; Belyaeva et al., 2023).

However, the existing CHAs didn't focus on real-time interactions with patients. In reality, as patients provide the initial symptoms, some related symptoms and information are often neglected because they are not obvious enough or the patient does not relate the information to the current physical condition. At this time, requesting additional information is crucial for making confident and accurate diagnosis. Based on this intuition, we design a LLM-based online interactive diagnosis system, DocCHA, which interacts with patients for multiple rounds of information request and reply iterations considering the history context, to enhance diagnosis confidence. Each round represents a *Doctor: information request, Patient: add information* interaction. At the same time, DocCHA is capable of providing patients with reliable and de-

[1] https://www.statpearls.com/
[2] https://www.wikipedia.org/

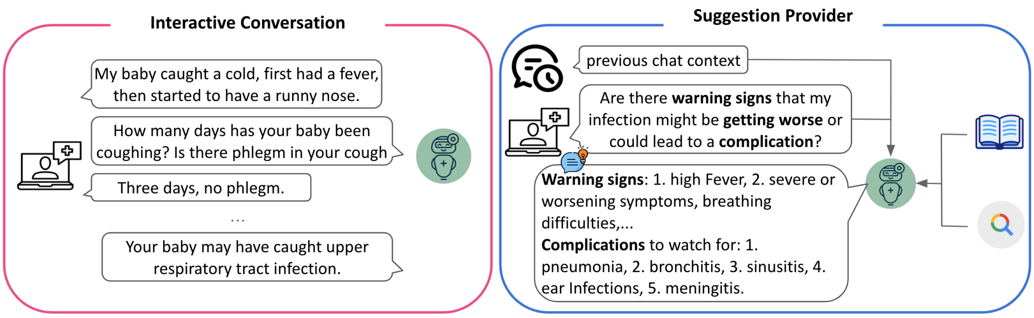

Figure 1: The key functions of DocCHA. The left part shows the interactive conversation between DocCHA and the patient. DocCHA is able to realize: information request, diagnosis, recommendation retrieve during the interactions. The right part is the suggestion provider, where DocCHA integrates the history chat context, patient's proposed questions and external data source to generate suggestions to patients.

tailed medical suggestions by leveraging Retrieval-Augmented Generation (RAG) with Google search API, StatPearls, and Wikipedia. The key functions of DocCHA are shown in Figure 1.

Our contributions can be summarized as follows:

- We design a novel framework named DocCHA, that enables online interactions with patients to request additional information (e.g., basic information, symptoms, etc.) and make confident and reliable diagnosis for patients.
- DocCHA employs the RAG technique by considering history conversation context and utilizing Google API and accessing StatPearls and Wikipedia and can generate personalized health recommendations for patients.
- Extensive experiments on the real-world diagnosis datasets demonstrate the effectiveness of DocCHA, which achieves a diagnosis accuracy of 89.2%.
- The RAG-based suggestion provider function of DocCHA outperforms the direct prompt of GPT-4 in terms of relevance, coherence, accuracy, and completeness.

2 Related Work

2.1 LLM-based Conversational Health Agents

Conversational Health Agents (CHAs) are conversational interactive systems that can provide users with various healthcare services around health assistance or diagnosis (Meier et al., 2019; Denecke and May, 2023; Laranjo et al., 2018). With the rapid development of LLMs, the technique of LLM-based CHA emerges. Based on methodology, current LLM-based CHAs can be divided into three categories, general LLM-based CHAs, specialized LLM-based CHAs, and multi-modality LLM-based CHAs. (1) General LLM-based CHAs utilize existing general LLMs to support health-related functions. (Chen et al., 2023) evaluates the robustness of utilizing ChatGPT to generate cancer treatment information. (Liu et al., 2023) utilizes LLMs to draft reports for biomedical signals. (Abbasian et al., 2023) proposed a general CHAs framework that can realize general healthcare question/answering, patient health record reports, and objective stress level estimation. (2) Specialized CHAs utilize reliable medical data sources to fine-tune LLMs in the medical domain. (Li et al., 2023) proposes ChatDoctor which fine-tunes Llama3 with a dataset of 100,000 patient-doctor dialogues. (Han et al., 2023) utilizes a dataset with 160,000 entries to fine-tune LLMs and constructs MedAlpaca for effective medical applications. (Luo et al., 2022) designs BioGPT which is pre-trained on biomedical literature and demonstrates advantages over six biomedical natural language processing tasks. (3) Multi-modal LLM CHAs utilize multi-modality data, including images, videos, or time-series, to extract more comprehensive health information. (Xu et al., 2023) utilizes a language-aligned image encoder with PaLM 2, to perform various chest X-ray tasks. (Belyaeva et al., 2023) proposes HeLM that maps complex data modalities into the LLM's token embedding space to estimate underlying disease risks. In our research, we focus on constructing real-time online diagnosis CHA. We apply the generalized LLM as the backbone model to avoid the demand for extensive patient cases to fine-tune high-quality diagnosis-specific LLMs. Besides, we focus on utilizing text data to diminish the inference time.

2.2 Retrieval-Augmented Generation for Medicine

LLMs have demonstrated great potential in various applications in the medical field (Nori et al., 2023; Singhal et al., 2023). However, LLMs sometimes generate plausible-sounding but incorrect outputs (Ji et al., 2023). Retrieval-Augmented Generation (RAG) is a technique proposed to improve the performance of knowledge-intensive tasks by integrating relevant knowledge retrieved from external resources (Lewis et al., 2020). Commonly used medical datasets for RAG systems include StatPearls (which consists of clinical knowledge), PubMed3[3] (which consists of all biomedical abstracts), and Wikipedia (which contains general knowledge). Along this line of research, (Frisoni et al., 2022) introduces a biomedical T5-based LLM, BioReader, to fetch and assemble information from PubMed. (Lála et al., 2023) proposes PaperQA which retrieves data from LitQA to reduce hallucinations in the QA task in biomedical domain. (Xiong et al., 2024) proposes the MEDRAG toolkit which organizes commonly used datasets and similarity ranking methods for the RAG task in the medical domain. In our paper, we apply RAG to generate detailed and reliable suggestions for patients after diagnosis.

3 Method

3.1 Bridge Personalized diagnosis and LLMs

Tele-health aims to collect and analyze patients' personal information and symptoms, then make diagnosis based on the comprehensive information. Traditionally, interactive tele-health requires professional doctors to effectively request additional information from patients based on historical information and domain-specific knowledge. Recently, LLMs have demonstrated superiority in problem-solving abilities (Yao et al., 2024), which can help to automate the tele-health diagnosis.

To bridge the gap between personalized diagnosis and LLMs, we propose an interactive LLM-based CHA, named DocCHA, which consists of the following two modules:

- A **diagnosis LLM module** that takes patients' initial self-report as input and generates further information requests to patients. The diagnosis result is provided to patients when the confidence score is above a threshold.
- A **suggestion provider LLM module** that receives patients' follow-up questions as input and integrates the knowledge extracted from external resources to generate a patient-customized response back to users.

The overall framework of DocCHA is shown in Figure 2, and we introduce two modules in Section 3.2 and Section 3.3, respectively.

3.2 The Diagnosis LLM Module

The diagnosis LLM takes multiple inputs from patients and is in charge of requesting information, diagnose, and calculating the current diagnosis confidence. The confidence score is calculated in each round of interaction and the output action of requesting information or diagnose depend on the confidence score. Upon receiving input from patients, diagnosis LLM will generate a diagnosis prompt to perform above sub-tasks, and send the corresponding result back to patients. The pipeline of diagnosis LLM is shown in Figure 3.

Patient input. Patient input contains patients' initial self-report S and additional information A_i of i-th round based on DocCHA's information request.

Diagnosis prompt generator. The diagnosis prompt generator takes the patient's input and generates (1) information request question, (2) current diagnosis result, (3) current diagnosis confidence score. To better guide LLMs to generate high-quality information request questions, we limit the questions to the domain of (1)symptoms request, (2) basic information request, (3) etiology request, and (4) existing examination and treatment request. The diagnosis prompt is shown in Appendix A.2.

Synthetic response generator. As is introduced in Session 4.1, we use the IMCS21 dataset, which is a dataset containing conversations between patients and doctors. To ensure our experiment is based on the real situation of patients, we synthesize patient responses to DocCHA's proposed questions using the following procedure: 1) provide GPT-4 with DocCHA's generated question and the real conversation from the IMCS21 dataset; 2) prompt GPT-4 to summarize the answer for the proposed question based on the real conversation; 3) output "Sorry I didn't notice that.", if the corresponding answer cannot be found in the real conversation. The prompt for GPT-4 to generate the synthetic response is shown in Appendix A.3.

Diagnosis LLM output. The output of diagnosis

[3] https://pubmed.ncbi.nlm.nih.gov/

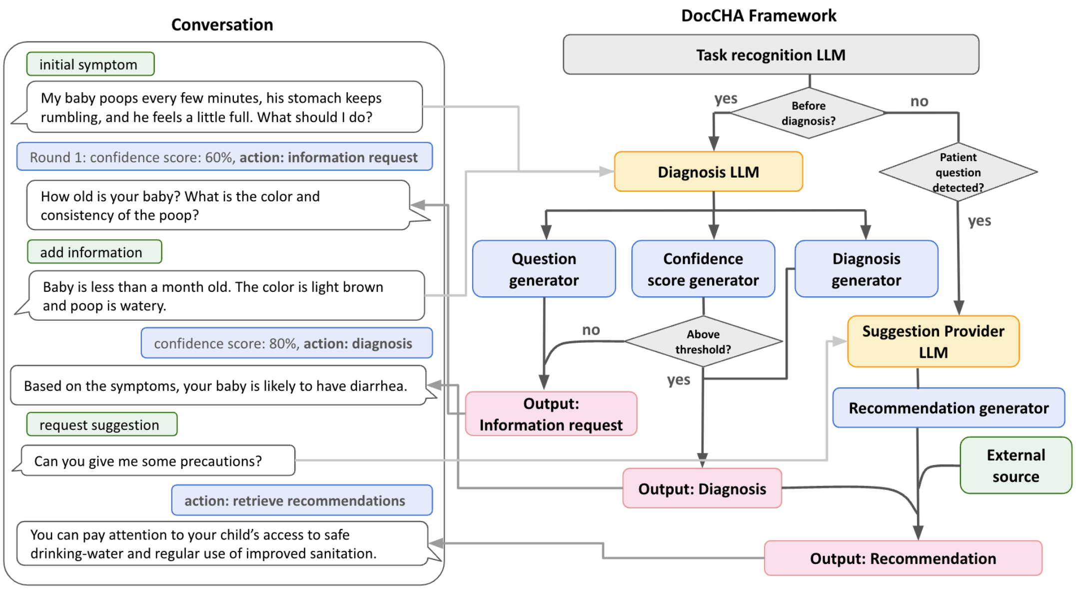

Figure 2: An overview of DocCHA. The left part shows the interaction between the patient and the DocCHA system. The right part shows the workflow of DocCHA, which consists of two main modules, **diagnosis LLM** (Section 3.2) and **suggestion provider LLM** (Section 3.3). The diagnosis LLM is decomposed into three sub-tasks, *question generator*, *confidence score generator*, and *diagnosis generator*. The suggestion provider LLM contains a *recommendation generator* that integrates with external sources and generates customized suggestions for patients.

LLM is collaboratively determined by two factors, the generated diagnosis result or information request question content, and the confidence score. When the confidence score C_i is above the threshold, diagnosis LLM will output the generated diagnosis result to patients. When C_i is below the threshold, diagnosis LLM will ask patients about the generated information request problem. In real applications, a request round limit can be set to avoid loop execution. The output of diagnosis LLM for patient u can be represented as equation 1,

$$D_u = \begin{cases} Diag\_LLM(S, A_1, ..., A_i), & C_i \geq \theta. \\ Quest\_LLM(S, A_1, ..., A_i), & C_i < \theta. \end{cases} \quad (1)$$

in which $Diag\_LLM$ represents the LLM-based diagnosis result generator, $Quest\_LLM$ represents LLM-based additional information request generator, S represents the initial symptom of the patient, A_i represents the i_{th} question proposed by DocCHA, θ represents the threshold of confidence score, which is usually set to 0.9 to ensure high confidence for reliable diagnosis.

3.3 The Suggestion Provider LLM Module

When DocCHA receives questions from the patient after making the diagnosis, it will call the suggestion provider LLM module. Since the goal of this function is to provide patients with basic knowledge of disease notice, while directly prompting LLMs to generate medical suggestions will sometimes result in replies including "Doctor consultation is required.", we apply RAG to access external knowledge. Reliable external sources is pivotal in improving Knowledge-groundedness and output reliability, especially in the health domain. In DocCHA, we explore two external resources. The first is based on Google search API, we name it as Google search API-based suggestion provider LLM, the second is based on external datasets, where we extract the most related information based on BM25 ranking (Robertson et al., 2009) from medical dataset StatPearls and general knowledge dataset Wikipedia, we name it as dataset-based suggestion provider LLM. The detailed procedure of the suggestion provider LLM module is as follows: (1) Suggestion provider LLM first takes the patient's request and the diagnosis result as input. (2) For Google search API-based suggestion provider LLM, it extracts information by direct search of keywords with Google search API, for dataset-based suggestion provider LLM, it matches the most similar content based on the BM25 ranking algorithm. (3) With the user's questions and extracted external information, we generate the response and output it to the user. The procedure of suggestion provider LLM is shown in Figure 4.

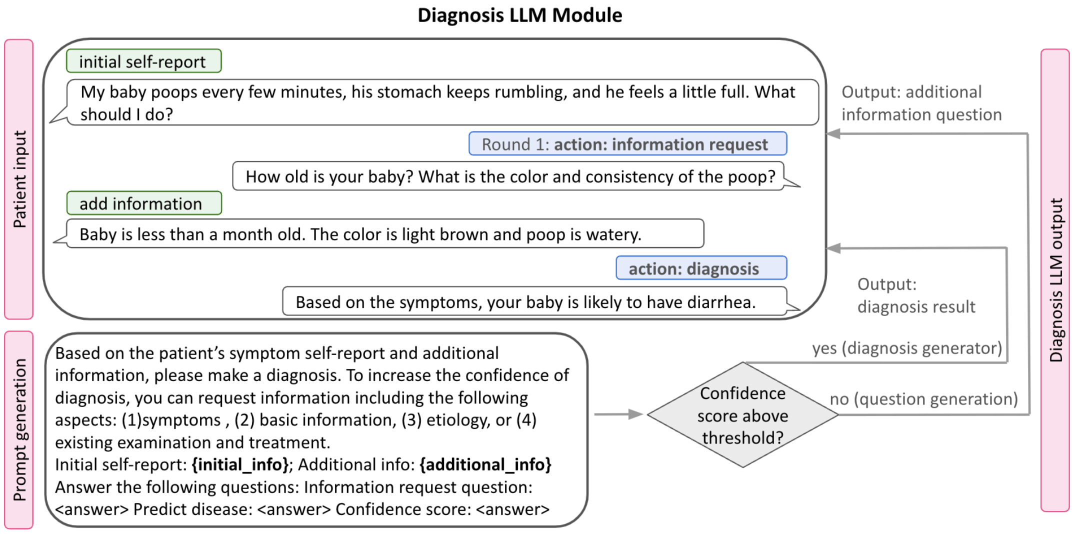

Figure 3: The pipeline of diagnosis LLM. Patient input consists of patient's initial self-report and additional information. The prompt takes patient input generates the information request question or predict disease, and outputs the generated information back to the patient based on the confidence score.

Patient input. Once receiving the patient's questions, DocCHA will process the text to extract the keyword of the patient's request. Then the suggestion generator will generate personalized responses based on the patient's specific request and DocCHA's previous diagnosis result. The input of the suggestion provider LLM module contains the extracted previous diagnosis, user's keywords for the Google search API-based module and user's question for the external datasets-based module.

Google search API. Once receiving input, the suggestion provider will call Google Search API. Then it will apply LLM to reformat the extracted information and send the reply back to users. The suggestion for user u is shown in equation 2,

$$S_u = Suggest_{google}\_LLM(D_u, K_1, ..., K_i), \quad (2)$$

where $Suggest_{google}\_LLM$ is Google search API-based suggestion generator, D_u is the diagnosis result of user u, $K_1, ...K_i$ are extracted keywords.

External datasets. With the natural high-reliability demand in the health domain, we also explore extracting information from professional and reliable external resources. (Xiong et al., 2024) organized benchmark RAG datasets for medicine. We select StatPearls for clinical decision support and Wikipedia for general knowledge as our external datasets. We apply the BM25 ranking to select the most related information to the patients' proposed question and diagnosed disease among the external knowledge. We use Q_u to denote the query of patient u, which is the diagnosed disease concatenated with the proposed question of user u, with the format of "The patient is diagnosed with disease and he/she proposes a question: question". For example, user u is diagnosed with *diarrhea* and asks **Can you give me some precautions?**, the generated query Q_u is: The patient is diagnosed with *diarrhea* and he/she proposes a question: **Can you give me some precautions?**. We use D to denote the document set of the external dataset, where D_j represents each section of every passage. For example, the "Treatment / Management" session under the "Diarrhea" passage in StatPearls is an item in D. Based on BM25 ranking, the score of each document D_j regarding the query Q_u is represented as Equation 3:

$$Score(D_j, Q_u) = \sum_{i=1}^{n} IDF(H_i) \cdot \frac{f(H_i, D_j) \cdot (\alpha + 1)}{f(H_i, D_j) + \alpha \cdot (1 - \beta + \beta \cdot \frac{|D_j|}{avgdl})}, \quad (3)$$

where H_i represents the keyword in document D_j, $|D_j|$ is the length of document D_j, $avgdl$ is the average document length among all documents in D. α, β are parameters, which we use $\alpha = 1.2, \beta = 0.75$ in our experiment.

We select D_j with top-k $Score(D_j, Q_u)$. The suggestion for user u is shown in equation 4,

$$S_u = Suggest_{dataset}\_LLM(D, Q_u), \quad (4)$$

where $Suggest_{dataset}\_LLM$ is dataset-based suggestion generator, D is the external dataset and Q_u

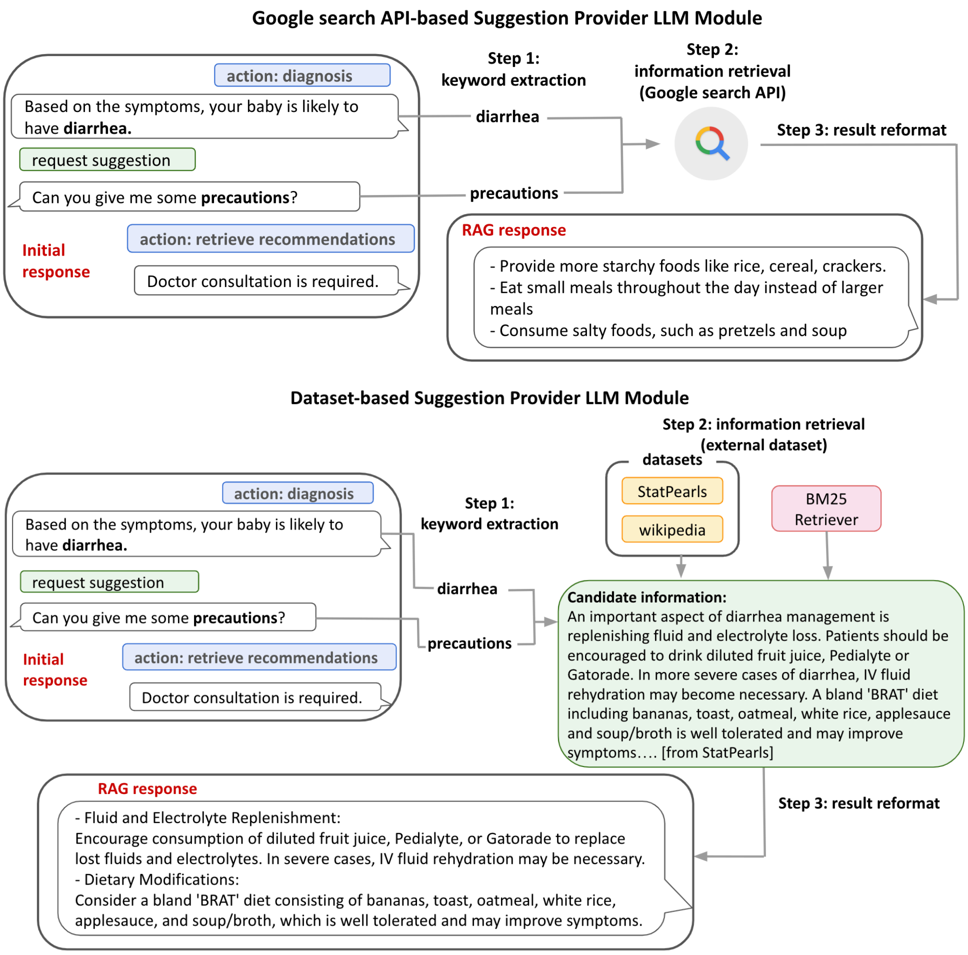

Figure 4: The pipeline of suggestion provider LLM. The suggestion provider will receive the patient's suggestion request as input. Then it will extract previously diagnosed disease and keywords from patient's questions. Suggestion provider LLM will extract information from external resources based on diagnosis result and keywords (full question for the dataset-based approach). The upper part shows the procedure of the Google search API-based approach and the lower part shows the dataset-based approach. The extracted information will be processed and sent back to the CHA window.

is user u's provided query.

4 Experiments

This session explores DocCHA's capability in diagnosis generation, information request question generation, and suggestion generation. We evaluate DocCHA on the IMCS21 dataset, which is an online consulting dataset that includes conversations between doctors and patients. For each human label task, we paid for 3 student volunteers and calculated the average.

4.1 Dataset and Experiment Setting

Dataset. IMCS21 dataset is a Chinese dataset that collects real online doctor-patient conversations, which covers 10 types of diagnosed diseases, including neonatal jaundice, common cold, bronchitis, indigestion, cough, constipation, upper respiratory tract infection, bronchopneumonia, fever, and diarrhea. In each conversation, the patient will provide an initial self-report to the doctor, and the doctor will ask questions for clarification or additional information. An example of the conversation is shown in Appendix A.1. Since the language of the raw dataset is Chinese, we first use Google Translate API to translate the dataset into English. Due to the expense of GPT-4, we do our experiment on 100 conversation samples. The statistics of the sampled IMCS21 dataset are shown in Table 1.

Experiment Setting. We test the performance of DocCHA with GPT-4 (Achiam et al., 2023) and Llama3[4] as the backbone LLM. The text embed-

[4]https://github.com/meta-llama/llama3

Table 1: Statistics of selected IMCS21 dataset.

Disease name	neonatal jaundice	common cold	bronchitis	indigestion	cough
# sample	6	15	17	14	2
Avg. Conversation Len.	34.4	35.8	40.6	36.5	60
Disease name	upper respiratory tract infection	constipation	bronchopneumonia	fever	diarrhea
# sample	16	3	5	12	10
Avg. Conversation Len.	34.4	44.5	49.75	34.4	44.5

ding model in the evaluation stage is mpnet-base-v2 model from Microsoft (Song et al., 2020).

4.2 Diagnosis LLM Performance

Prediction accuracy. To test the disease prediction accuracy of DocCHA, we compare the diagnosis result with the ground truth diagnosis result labeled by professional doctors. We apply two evaluation metrics: (1) cosine similarity between DocCHA's diagnosis embedding and the ground truth diagnosis embedding; (2) accuracy based on human label. The second metric, i.e., the accuracy based on the human label, is an action for normalization of DocCHA's output. Since there's no restriction for DocCHA's diagnosis result, while the ground-truth labels are categorized into 10 categories, we manually check if the prediction generated by DocCHA is correct. For example, based on the patient's symptoms, DocCHA might make a diagnosis as "lactose intolerance", "food poisoning", or a more detailed diagnosis such as "food poisoning due to consumption of undercooked soy milk", while the ground-truth label is "indigestion". In this case, the human label will mark DocCHA's diagnosis as correct. We prompt DocCHA to generate information request questions for 4 rounds and test the performance. We also test the performance when including the whole conversation to make a diagnosis. The performance is shown in Table 2.

The result shows DocCHA can achieve diagnosis accuracy of 89.2% after four rounds of additional information requests, which is only 4.9% inferior compared with including the whole conversation. This result uncovers DocCHA's diagnosis capability. At the same time, there's a surge in both cosine similarity and accuracy after DocCHA requests additional information from patients (i.e., from the initial stage to round 1). This phenomenon shows the effectiveness in DocCHA's requesting informative information from patients.

Confidence score. To analyze the relationship between DocCHA's diagnosis confidence and the richness of information, we calculate DocCHA's prediction confidence score in each round. To avoid early stopping for evaluation purposes, we set the confidence score threshold to 100% so that no information request will be rejected due to a high confidence score. The result is shown in Table 3.

From the result, we can see a positive relation between the information richness and DocCHA's confidence score, which aligns with the intuition that more information can help make a more confident diagnosis.

Question generation. To test the quality of DocCHA's generated information request questions, we implement the following two steps: (1) Prompt GPT-4 to extract information in align with DocCHA's question based on real conversations. When no information can be matched, the user will reply "Sorry I didn't notice that." (2) Compare the similarity between the extracted information and the real conversation information. When comparing the information similarity, we apply two methods: (1) Calculate the similarity between the extracted information and the real conversation. (2) Apply GPT-4 to extract the key information of the real conversation first, and then prompt GPT-4 to output the common ratio of the extracted information and the ground-truth key information. Note that for evaluation, the knowledge of patients is only from the existing conversation, while in reality, patients will react to DocCHA's information requests accordingly, thus will be very likely to generate other valuable information. Examples of DocCHA's generated questions are shown in Appendix A.2. The quality measures results are shown in Table 4.

Case study. In this part, we show a concise example of DocCHA's interaction with a user in the diagnosis stage. The answer of the patient is generated following *Question generation*. The example process is shown in Appendix A.5.

4.3 Suggestion Provider LLM Performance

DocCHA also realizes the suggestion provider LLM module to provide users with personalized suggestions. To increase the groundness and avoid rejecting the recommendation requests, DocCHA applies RAG to extract external knowledge utiliz-

Table 2: Diagnose performance of DocCHA.

Metric	Model	Initial (baseline)	Round 1	Round 2	Round 3	Round 4	all info
Cosine similarity	Llama3	0.4223	0.4639	0.4811	0.4809	0.5022	**0.5746**
	GPT-4	0.4711	0.5057	0.5078	0.4954	0.5312	**0.6146**
Accuracy	Llama3	0.5461	0.6121	0.6613	0.6517	0.7312	**0.7793**
	GPT-4	0.5966	0.6932	0.7708	0.7045	0.8920	**0.9310**

Table 3: Confidence scores of DocCHA's diagnosis.

Model	Initial (baseline)	Round 1	Round 2	Round 3	Round 4	all info
Llama3	0.5389	0.6011	0.6425	0.6574	0.6837	**0.7283**
GPT-4	0.6188	0.6631	0.6778	0.6614	0.7381	**0.8216**

Table 4: Quality of DocCHA's proposed questions (by measuring the coverage of extracted patient's information).

Metric	Model	Round 1	Round 2	Round 3	Round 4
Cosine similarity	Llama3	0.3454	0.3988	0.4377	**0.4598**
	GPT-4	0.3532	0.4095	0.4374	**0.4640**
Common ratio	Llama3	0.3993	0.4135	0.4601	**0.4988**
	GPT-4	0.4068	0.4233	0.4580	**0.5006**

ing Google API, StatPearls, and Wikipedia. Then it leverages GPT-4 to reformat and reorganize the external knowledge based on patients' specific request queries and output the suggestions to patients. For Google API, we keep the top 5 search results. For the external dataset-based approach, we keep the top 3 similar documents.

Human grade comparison. We prompt GPT-4 to generate 10 questions from patients' perspectives on the aspects of drug recommendation, medical advice, and precautions given different types of diseases. The prompt and generated questions are shown in Appendix A.4. We compare the generated suggestions of 1) direct prompting GPT-4, 2) Google search API-based suggestion provider, 3) dataset-based suggestion provider, with human-given scores over aspects of relevance, coherence, accuracy, and completeness, with each metric range from 0 to 5. The detailed criteria is introduced in Appendix A.6. The result is shown in Table 5.

Table 5: Suggestion Generation Evaluation Results.

Aspects	Direct Prompt	Google API	External Datasets
Relevance	2.7	**4.6**	4.3
Coherence	3.8	4.4	**4.5**
Accuracy	2.6	4.1	**4.5**
Completeness	2.4	3.9	**4.4**

From the result, we can see the suggestion provider's performance outperforms direct prompt both with Google search API and with reliable external datasets. The main reason is direct prompts sometimes generate responses like "Sorry I didn't notice that.", which refuses to provide patients with information. From the comparison between Google search API-based suggestion provider and external dataset-based suggestion generator, we can see better performance of Google API in the relevance aspect. This might be because directly searching information from Google can generate a more relevant output, especially when the question is more specific, which can hardly find content in medical datasets with high relatedness. However, the dataset-based suggestion generator outperforms the other approaches in accuracy and completeness, owing to the reliability and detailed information in professional medical resources.

Case study. The examples of generated suggestions are shown in Appendix A.5.

5 Conclusion

We propose an LLM-augmented interactive online diagnosis system, DocCHA, that can interact with patients by generating questions to request additional information from patients to assist the diagnosis performance. This interaction with patients can prevent patients from missing important information, e.g., some complications, or the contact history with suspicious patients, thus avoiding misdiagnosis. Moreover, DocCHA can access external knowledge to provide patients with detailed and reliable medical suggestions that suit patients' questions by utilizing the Google search API and external datasets including StatPearls and Wikipedia. Future research can focus on extending DocCHA to integrate multi-modality data.

6 Limitations

The question generation module in Diagnosis LLM has some randomness, which can bring some randomness to the overall diagnosis performance. Besides, the patient's responses are generated based on the existing static dataset, which weakens the quantitative performance of DocCHA. In future research work, real human-computer interaction experiments can be done to explore DocCHA's capabilities more accurately.

References

Mahyar Abbasian, Iman Azimi, Amir M Rahmani, and Ramesh Jain. 2023. Conversational health agents: A personalized llm-powered agent framework. *arXiv preprint arXiv:2310.02374*.

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Ö Aydın and E Karaarslan. Openai chatgpt generated literature review: Digital twin in healthcare, 2022. *Available at SSRN*, 430:8687.

Anastasiya Belyaeva, Justin Cosentino, Farhad Hormozdiari, Krish Eswaran, Shravya Shetty, Greg Corrado, Andrew Carroll, Cory Y McLean, and Nicholas A Furlotte. 2023. Multimodal llms for health grounded in individual-specific data. In *Workshop on Machine Learning for Multimodal Healthcare Data*, pages 86–102. Springer.

Som Biswas. 2023. Chatgpt and the future of medical writing.

Shan Chen, Benjamin H Kann, Michael B Foote, H Aerts, Guergana K Savova, Raymond H Mak, and Danielle S Bitterman. 2023. The utility of chatgpt for cancer treatment information. medrxiv. *Preprint posted March*, 16.

Kerstin Denecke and Richard May. 2023. Investigating conversational agents in healthcare: Application of a technical-oriented taxonomy. *Procedia Computer Science*, 219:1289–1296.

Giacomo Frisoni, Miki Mizutani, Gianluca Moro, and Lorenzo Valgimigli. 2022. Bioreader: a retrieval-enhanced text-to-text transformer for biomedical literature. In *Proceedings of the 2022 conference on empirical methods in natural language processing*, pages 5770–5793.

Tianyu Han, Lisa C Adams, Jens-Michalis Papaioannou, Paul Grundmann, Tom Oberhauser, Alexander Löser, Daniel Truhn, and Keno K Bressem. 2023. Medalpaca–an open-source collection of medical conversational ai models and training data. *arXiv preprint arXiv:2304.08247*.

Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38.

Jakub Lála, Odhran O'Donoghue, Aleksandar Shtedritski, Sam Cox, Samuel G Rodriques, and Andrew D White. 2023. Paperqa: Retrieval-augmented generative agent for scientific research. *arXiv preprint arXiv:2312.07559*.

Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie YS Lau, et al. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9):1248–1258.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.

Yunxiang Li, Zihan Li, Kai Zhang, Ruilong Dan, Steve Jiang, and You Zhang. 2023. Chatdoctor: A medical chat model fine-tuned on a large language model meta-ai (llama) using medical domain knowledge. *Cureus*, 15(6).

Chunyu Liu, Yongpei Ma, Kavitha Kothur, Armin Nikpour, and Omid Kavehei. 2023. Biosignal copilot: Leveraging the power of llms in drafting reports for biomedical signals. *medRxiv*, pages 2023–06.

Renqian Luo, Liai Sun, Yingce Xia, Tao Qin, Sheng Zhang, Hoifung Poon, and Tie-Yan Liu. 2022. Biogpt: generative pre-trained transformer for biomedical text generation and mining. *Briefings in bioinformatics*, 23(6):bbac409.

Michael McTear. 2022. *Conversational ai: Dialogue systems, conversational agents, and chatbots*. Springer Nature.

Pascal Meier, Jan Heinrich Beinke, Christian Fitte, Alina Behne, and Frank Teuteberg. 2019. Feelfit-design and evaluation of a conversational agent to enhance health awareness. In *ICIS*.

Joao Luis Zeni Montenegro, Cristiano André da Costa, and Rodrigo da Rosa Righi. 2019. Survey of conversational agents in health. *Expert Systems with Applications*, 129:56–67.

Harsha Nori, Nicholas King, Scott Mayer McKinney, Dean Carignan, and Eric Horvitz. 2023. Capabilities of gpt-4 on medical challenge problems. *arXiv preprint arXiv:2303.13375*.

Stephen Robertson, Hugo Zaragoza, et al. 2009. The probabilistic relevance framework: Bm25 and beyond. *Foundations and Trends® in Information Retrieval*, 3(4):333–389.

Karan Singhal, Shekoofeh Azizi, Tao Tu, S Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, et al. 2023. Large language models encode clinical knowledge. *Nature*, 620(7972):172–180.

Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. Mpnet: Masked and permuted pre-training for language understanding. *Advances in neural information processing systems*, 33:16857–16867.

Tao Tu, Shekoofeh Azizi, Danny Driess, Mike Schaekermann, Mohamed Amin, Pi-Chuan Chang, Andrew Carroll, Charles Lau, Ryutaro Tanno, Ira Ktena, et al. 2024. Towards generalist biomedical ai. *NEJM AI*, 1(3):AIoa2300138.

Saeid Ashraf Vaghefi, Dominik Stammbach, Veruska Muccione, Julia Bingler, Jingwei Ni, Mathias Kraus, Simon Allen, Chiara Colesanti-Senni, Tobias Wekhof, Tobias Schimanski, et al. 2023. Chatclimate: Grounding conversational ai in climate science. *Communications Earth & Environment*, 4(1):480.

Guangzhi Xiong, Qiao Jin, Zhiyong Lu, and Aidong Zhang. 2024. Benchmarking retrieval-augmented generation for medicine. *arXiv preprint arXiv:2402.13178*.

Shawn Xu, Lin Yang, Christopher Kelly, Marcin Sieniek, Timo Kohlberger, Martin Ma, Wei-Hung Weng, Attila Kiraly, Sahar Kazemzadeh, Zakkai Melamed, et al. 2023. Elixr: Towards a general purpose x-ray artificial intelligence system through alignment of large language models and radiology vision encoders. *arXiv preprint arXiv:2308.01317*.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. 2024. Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, 36.

A Appendix

The supplementary material is structured as follows:

- Appendix A.1 showcases example dialogues between doctors and patients in IMCS21 dataset;
- Appendix A.2 shows the prompt and simulation examples of DocCHA's information request questions;
- Appendix A.3 shows the prompt to generate patient's synthetic response based on the real conversation and proposed questions;
- Appendix A.4 shows the simulation prompt and examples of patient provided questions;
- Appendix A.5 shows the case study of DocCHA's diagnosis module and suggestion provider module's examples;
- Appendix A.6 introduces the human score criteria for the generated suggestions;

A.1 Example of IMCS21 Conversation

The fragment of conversations in IMCS21 dataset is shown in Table 6.

Please extract and briefly summarize the information from the conversation below regarding the proposed question. Output "Sorry I didn't notice that." if the conversation doesn't include the information.
Conversation: ***[conversation]***
Question: ***[question]***
Answer the question by filling in the <answer> below, no explanation needed.
The summarized answer is: <answer>

Figure 5: The prompt for GPT-4 to synthesis patient's response based on patient's personal information and DocCHA's information request question.

A patient has been diagnosed with disease ***[disease name]***. Please generate 10 questions that the patient might ask the doctor based on the disease. The question should include one or more aspects of medical advice, precautions, and drug recommendations.

Figure 6: The prompt for GPT-4 to simulate patient's potential questions.

A.2 DocCHA Questions Generation

The prompt of DocCHA's diagnosis LLM module is shown in Figure 7.

Examples of DocCHA's generated questions under each disease type is shown in Table 7

A.3 Patient Synthetic Response Generation

The prompt for patient's synthetic response generation is shown in Figure 5.

A.4 Patient Question Generation

The prompt for GPT-4 to generate potential questions provided by patients is shown in Figure 9.

The synthetic patients' potential questions generated by GPT-4 is shown in Table 8.

A.5 Case Study of DocCHA's Conversation

The example of DocCHA's diagnose interaction with patients is shown in Figure 8.

The example of DocCHA's suggestion provider module is shown in Figure 9.

Table 6: Fragments of IMCS21 dialogues.

Conversation 1
Initial self-report: My 27-month-old baby has repeated fevers. 38 degrees. Where to take antipyretic medicine?
Conversation: Doctor: What is the child's current body temperature? Patient: 38 degrees. Doctor: How long has he been taking antipyretic drugs? Do you mean that his hands and feet became cold after taking antipyretic drugs, or that he had cold hands and feet without taking antipyretic drugs. Patient: He started taking them the night before yesterday. Doctor: What is the highest body temperature of this child? Patient: Before taking the drugs, his hands, feet and buttocks were cold Patient: 38.5 degrees. Doctor: In addition to fever, are there any other symptoms? For example, runny nose, cough, and diarrhea? Patient: No. Doctor: Generally speaking, if the body temperature exceeds 38.5 degrees, give the child antipyretic drugs. At the current body temperature, give the child more water and use a fever patch. Cold hands and feet may occur during the rising period of body temperature. It may be much higher than it is now when you measure the body temperature again in a while. Patient: What should I do if he always has a fever? Doctor: Has he done a routine blood test? Patient: No. Doctor: In this case, the possibility of respiratory infection is very high. The possibility of viral infection is high. Viral infection usually has a peak period of 3 to 4 days. During the peak period, the child will have repeated fever. After the peak period, the symptoms will gradually ease. If the child is in a good mental state, oral medication can be used for symptomatic treatment. Patient: No need to go to the hospital for blood test? Doctor: A routine blood test should be done. Patient: Oh.Thank you. Doctor: You're welcome. What medication are you taking now? In addition to antipyretic drugs, are there any other drugs? Patient: Cefixime. Doctor: You should take cold granules and heat-clearing and detoxifying drugs orally. If the total white blood cell count is not high and there is no indication of bacterial infection, there is no need to take cefixime orally. You can choose pediatric acetaminophen yellow nanoparticles and pediatric Shuangjin oral liquid or Lanqin oral liquid. Patient: Pediatric acetaminophen yellow granules 999? Doctor: Yes. Heat-clearing and detoxifying drugs need to be taken. Patient: Ok. Doctor: If his body temperature is still unstable for more than 3 days, you need to go to the hospital for a routine blood test. Patient: Ok. Doctor: In addition, you need to observe the child's mental state. If the mental state is not good, you need to go to the hospital as soon as possible. Patient: OK, thank you. Doctor: You're welcome.
Diagnose: Upper respiratory tract infection

Now assume you are a doctor and I am a patient. I will give you some of the basic information or symptoms and you need to make a diagnosis. To help you make the diagnosis more confidently, you can request other information from me including the following aspects: (1) symptoms (including symptom details), (2) basic information, (3) etiology, or (4) existing examination and treatment from me.
Ask a question that you think can help you most to improve diagnosis confidence.
The initial symptom is: ***[initial symptom]***
Additional information is: ***[additional information]***

Answer the question by filling in the <answer> below, no explanation is needed.
Request information question: <answer>
Predict disease: <answer>
Confidence score: <answer>

Figure 7: The prompt of DocCHA's diagnosis LLM module.

A.6 Suggestion Generate LLM Label Criteria

The score range for each dimension is 0-5, where 0 represents the lowest quality and 5 represents the highest quality of the given dimension. The scoring criteria of generated suggestion is as follows:

- **Relevance** 1) Contextual Appropriateness: The generated content should be contextually appropriate and align with the input or prompt provided. 2) Task Alignment: The content should fulfill the specific task requirements.
- **Coherence** 1) Syntax: The content should be syntactically well-structured. 2) Logical Flow: The ideas should flow logically, without abrupt or confusing transitions.
- **Accuracy** The content should be accurate and verifiable.
- **Completeness** 1) Coverage: The content should comprehensively cover the necessary points or aspects of the input. 2) Detail: The level of detail should be appropriate for the task, neither too vague nor overly detailed.

Table 7: Case study of questions generated by DocCHA

neonatal jaundice	common cold
1. Does your baby have any other symptoms like yellow skin or eyes, poor feeding, or lethargy? What treatments has he received in the hospital? 2. Has your child been feeding well and having regular bowel movements? Are there other symptoms such as changes in skin color, fever, or lethargy? 3. Can you provide more details about the child's symptoms? For instance, are there changes in the color of urine or stool? 4. Has the child undergone any tests like a bilirubin test or a complete blood count (CBC) test? 5. What condition are you being treated for with Yinzhihuang oral liquid, and have you noticed any other symptoms or changes in your health since starting the treatment?	1. Could you provide more details about the cough - is it dry or wet, and is there any phlegm? 2. Has the baby been exposed to any new environments or allergens recently? 3. Does your baby have a fever, cough or difficulty breathing? Has the baby been exposed to anyone sick recently? 4. Has your baby had a fever or any other symptoms such as coughing, difficulty breathing, or loss of appetite? 5. Has the baby experienced any changes in behavior, such as irritability, or loss of appetite? Does the baby have a runny nose or watery eyes?
bronchitis	**indigestion**
1. Has your child been exposed to any allergens such as dust, pets, or certain foods? Has there been any weight loss or fatigue? 2. Has your child been vaccinated according to the recommended schedule? 3. Can you provide more details about the medicine your baby has been taking? Also, has there been any fever, loss of appetite, or changes in behavior? 4. Could you please provide more detail on the duration of the cough and the length of time the medication has been administered? 5. Has the baby been exposed to any allergens or environmental factors such as smoking or dust?	1. Could you please describe the color, consistency and smell of the baby's stool? Also, have you noticed any changes in his feeding behavior or weight? 2. Any signs of dehydration like less wet diapers, dry mouth or crying without tears? 3. Have you noticed any changes in your baby's feeding habits or behavior? And could you describe the color and consistency of the baby's stool? 4. Can you provide more details about the vomiting? Was there any other symptoms like fever, abdominal pain, or diarrhea? 5. Has your son been losing weight or failing to gain weight? Are there any other symptoms such as fever, blood in the vomit?
cough	**upper respiratory tract infection**
1. Has your child been exposed to any allergens or irritants recently, such as smoke, dust, or pet dander? 2. Have you noticed any patterns or triggers in the coughing episodes, such as specific times, activities, or environments where the coughing gets worse? 3. Has your child been exposed to any allergens or irritants recently such as smoke, dust, or strong odors? 4. Do you notice if there is a pattern, like does your child's cough get worse after certain activities or at certain times of the day? 5. Has your child experienced any recent changes in environment, exposure to allergens, or triggers for his cough?	1. Can you tell me more about the baby's appetite, any changes in behavior or activity level, and if there has been any weight loss? 2. Can you tell me more about the baby's vaccination history? Have there been any changes in appetite or behavior? 3. Could you provide more details about the child's medical history, including any existing conditions such as asthma or allergies? 4. Has the child been exposed to anyone with a respiratory illness recently? 5. Has the child had any other tests such as a chest x-ray or an echocardiogram?
constipation	**bronchopneumonia**
1. Has your daughter experienced any other symptoms such as vomiting, constipation, loss of appetite, weight loss, or any discomfort? 2. Does she eat normally? Is there any history of similar issues in your family? 3. How long has she been having these symptoms? Has she been drinking enough water? 4. Has your daughter experienced any other symptoms such as abdominal pain, fatigue, or changes in appetite? 5. Is her bigger stomach associated with any discomfort or pain? How long has she been having these symptoms?	1. Can you provide detailed information about the blood test results, and confirm if any additional tests like a chest X-ray or CT scan have been performed? 2. Could you please provide more details about the cough? For example, how long has the cough been present? Is it getting worse? 3. Does the baby have difficulty breathing or any other discomfort? Also, what are the results of any x-rays or tests performed at the hospital? 4. Does the child have any allergy or family history of respiratory diseases? What is his current weight and height? 5. Can you specify the type of wheezing and any other symptoms such as shortness of breath, fatigue, or difficulty in feeding?
fever	**diarrhea**
1. Has the child been eating and drinking normally? Has there been any recent change in the child's diet? 2. Has the baby had any vaccinations recently or been exposed to anyone who is sick? Has the baby been eating, drinking and urinating normally? 3. Can you tell me if the baby has had any changes in appetite or behavior, such as increased sleepiness, irritability, or loss of interest in usual activities? 4. Has the baby been exposed to anyone sick recently or been in a new environment that could expose him to different bacteria? 5. Has your child shown any other symptoms like vomiting, loss of appetite, or lethargy?	1. Has the baby shown any other symptoms such as fever, vomiting, or loss of appetite? Has there been any change in the baby's behavior or activity level? 2. Has the baby been previously diagnosed with any medical conditions or allergies that could explain the diarrhea and blood in the stool? 3. Has there been any change in the baby's feeding habits? How many times a day does the baby pass stool? 4. Is the baby showing signs of discomfort or distress? Does the baby have any other symptoms like vomiting, loss of appetite, or weight loss? 5. Has the baby had any fever, vomiting, or signs of dehydration such as dry mouth, no tears when crying, or fewer wet diapers?

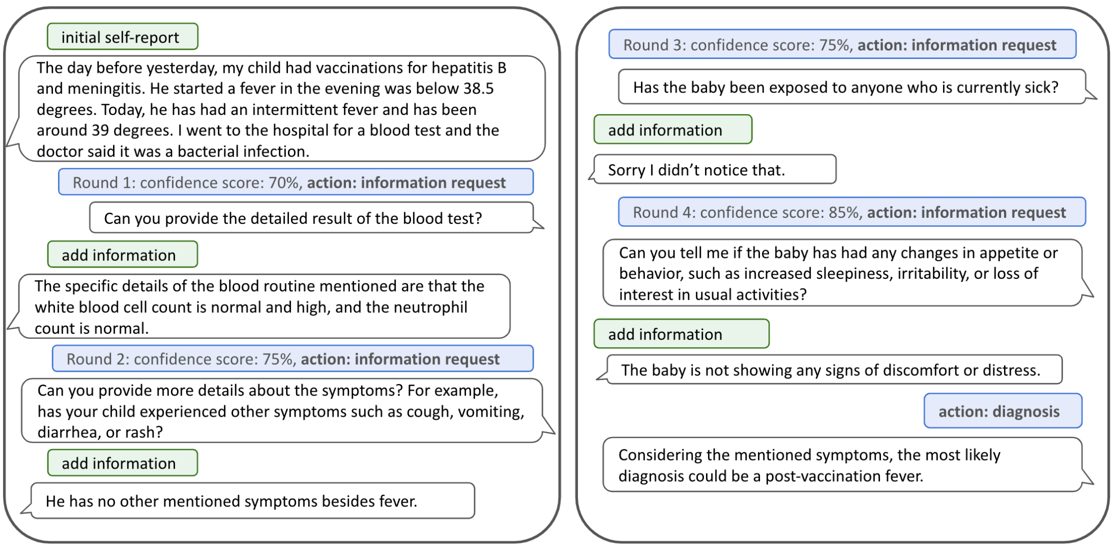

Figure 8: An example conversation between DocCHA and a patient.

Table 8: Synthetic patient provided questions generated by GPT-4.

neonatal jaundice	common cold
1. What are the primary causes of neonatal jaundice, and how can I tell if my baby's jaundice is serious? 2. What are the treatment options available for neonatal jaundice, and when would my baby need phototherapy? 3. How often should I feed my baby to help manage their jaundice, and does breastfeeding or formula feeding make a difference? 4. Are there any specific signs or symptoms I should watch for that indicate my baby's jaundice is worsening and needs immediate medical attention? 5. How long does neonatal jaundice typically last, and when should I expect to see improvement in my baby's condition? 6. Are there any medications or supplements that can help reduce bilirubin levels in my baby, and are they safe to use? 7. What precautions should I take when caring for a baby with jaundice at home, and how often should I follow up with the doctor? 8. Is it safe to expose my baby to sunlight as a form of treatment, and if so, for how long and how often? 9. Could neonatal jaundice have any long-term effects on my baby's health or development, and what steps can I take to prevent complications? 10. Are there any underlying conditions that could be causing my baby's jaundice, and should any additional tests be performed to rule them out?	1. What over-the-counter medications can I take to relieve my symptoms? 2. Are there any specific home remedies or natural treatments you recommend for the common cold? 3. How can I prevent the spread of my cold to family members and coworkers? 4. Is it safe to take decongestants or antihistamines while on my current medications? 5. What are the best ways to boost my immune system to recover more quickly? 6. When should I seek medical attention if my symptoms worsen or do not improve? 7. Are there any foods or drinks I should avoid while I have a cold? 8. How can I differentiate between a common cold and something more serious, like the flu or a sinus infection? 9. Is it okay to continue exercising while I have a cold, or should I rest until I feel better? 10. Can you recommend any specific precautions to take to avoid catching colds in the future?
bronchitis	**indigestion**
1. What medications are most effective for treating bronchitis, and do I need antibiotics? 2. Are there any over-the-counter medications or home remedies that can help relieve my symptoms? 3. What lifestyle changes or precautions should I take to manage my bronchitis and prevent it from worsening? 4. How long does bronchitis usually last, and when can I expect to start feeling better? 5. Is it safe for me to continue exercising while I have bronchitis, or should I rest until my symptoms improve? 6. Can you recommend any specific breathing exercises or techniques to help clear my airways? 7. What are the warning signs that my bronchitis is turning into pneumonia or another serious condition? 8. Are there any foods or beverages I should avoid to help manage my bronchitis symptoms? 9. How can I prevent future episodes of bronchitis, especially if I am prone to respiratory infections? 10. Should I get a flu shot or pneumonia vaccine to reduce my risk of complications from bronchitis?	1. What medications are most effective for treating indigestion, and are there any over-the-counter options I should consider? 2. Are there any specific foods or drinks I should avoid to help prevent or reduce indigestion? 3. What lifestyle changes can I make to manage my indigestion symptoms more effectively? 4. Is it safe for me to take antacids regularly, or are there potential side effects I should be aware of? 5. Are there any natural remedies or dietary supplements that can help with indigestion? 6. How can I tell if my indigestion is a symptom of a more serious condition, such as an ulcer or GERD? 7. What are the best practices for eating habits and meal timing to minimize indigestion? 8. Can stress or anxiety contribute to my indigestion, and if so, what stress-management techniques would you recommend? 9. Should I avoid lying down or going to bed immediately after eating to prevent indigestion? 10. Are there any specific exercises or physical activities that can help alleviate or worsen my indigestion symptoms?
cough	**upper respiratory tract infection**
1. What is the underlying cause of my cough? 2. What over-the-counter medications do you recommend? 3. Should I take prescription medications for my cough? 4. Are there any home remedies that can help alleviate my cough? 5. What lifestyle changes can I make to help reduce my cough? 6. When should I be concerned about my cough and seek further medical attention? 7. How can I prevent spreading my cough to others? 8. Can my cough be related to other conditions such as asthma or GERD? 9. Are there any specific foods or drinks I should avoid or consume to help with my cough? 10. Should I get any tests to diagnose the cause of my cough?	1. Can you explain what likely caused my infection, and is it viral or bacterial? 2. What is the usual duration of symptoms for this type of infection, and when can I expect to start feeling better 3. Will antibiotics help treat my infection, or is it likely to resolve on its own? 4. Can you recommend any specific over-the-counter medications, such as decongestants or pain relievers, to help alleviate my symptoms? 5. What precautions should I take to avoid spreading the infection to family members, friends, or coworkers? 6. Are there any home remedies, like saltwater gargles or steam inhalation, that can help relieve congestion or sore throat? 7. Are there warning signs that my infection might be getting worse or could lead to a complication? 8. Is there anything I can do to strengthen my immune system and prevent future infections? 9. Are there any specific foods or drinks I should avoid while I have this upper respiratory infection? 10. Is there a possibility that my upper respiratory infection could develop into something more serious?
constipation	**bronchopneumonia**
1. What could be causing my constipation, and how can I prevent it from happening again? 2. Should I change my diet to help alleviate constipation? 3. Are there over-the-counter medications or supplements that could help relieve my constipation? 4. Could my current medications be contributing to my constipation? 5. How much water should I be drinking each day to help with constipation? 6. Are there exercises or physical activities that could help stimulate bowel movements? 7. When should I consider seeking medical attention if my constipation persists? 8. Are there any natural remedies or home treatments I can try to relieve constipation? 9. Could stress or anxiety be contributing to my constipation? 10. What lifestyle changes can I make to promote regular bowel movements in the long term?	1. What exactly is bronchopneumonia, and how does it differ from other types of pneumonia? 2. What are the common symptoms of bronchopneumonia, and how long will they typically last? 3. What are the treatment options available for bronchopneumonia? 4. Are there any specific precautions I should take to prevent spreading the infection to others? 5. Are there any over-the-counter medications or home remedies that can help? 6. What are the potential complications of bronchopneumonia, and how likely are they to occur in my case? 7. Will I need any follow-up appointments or tests to monitor my recovery from bronchopneumonia? 8. Could bronchopneumonia recur in the future, and are there steps I can take to prevent it? 9. How can I support my immune system to recover from bronchopneumonia faster? 10. Are there any warning signs that would indicate I need urgent medical attention?
fever	**diarrhea**
1. How do I tell if it's a bacterial infection? Should I go get a test? 2. What temperature should I consider high enough to be concerned about? 3. What are the possible complications of having a fever? 4. What are the best methods to lower my fever at home? 5. Are there any dietary changes or fluids I should focus on while I have a fever? 6. Are there any medications I should avoid while I have a fever? 7. What steps can I take to prevent spreading my illness to others if I have a fever? 8. Are there specific signs or symptoms I should be aware of that could indicate my condition is worsening? 9. What symptoms or signs should prompt me to seek urgent medical care for my fever? 10. What is happening in my body when I have a fever? How does it impact my immune system?	1. Can you help me understand the underlying cause of my diarrhea? Are there specific tests we should consider? 2. Are there specific foods or drinks I should avoid or include in my diet to help manage my diarrhea? 3. What steps should I take to stay hydrated? Are oral rehydration solutions like Pedialyte recommended? 4. Are there anti-diarrheal medications or other treatments I can use to help reduce the frequency of my stools? 5. What symptoms or signs indicate that my diarrhea may be severe enough to require medical intervention? 6. What is the typical duration of diarrhea? When should I expect to see improvement in my symptoms? 7. Should I be concerned that my diarrhea might indicate a more serious illness? Are there red flags to watch for? 8. Are there any lifestyle changes I can make to help prevent future episodes of diarrhea? 9. What are the potential complications of severe or prolonged diarrhea? 10. How long is it normal for diarrhea to last?

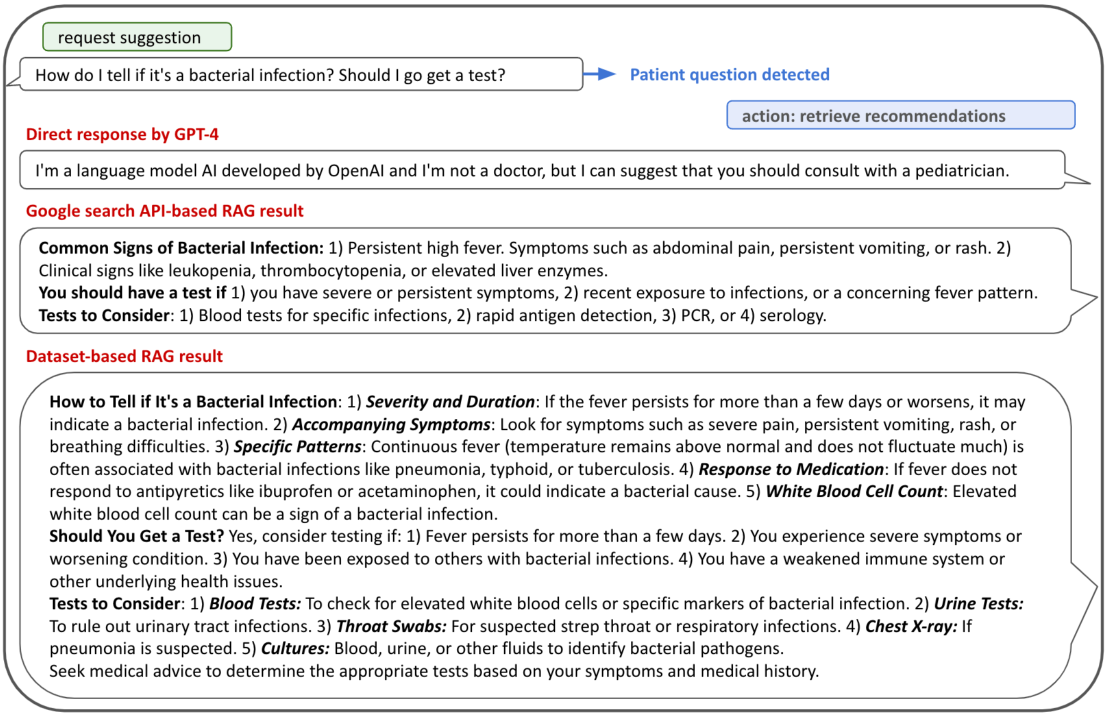

Figure 9: Examples of DocCHA provided suggestions based on Google search API or external datasets.