# EHRTutor: Enhancing Patient Understanding of Discharge Instructions

Zihao Zhang, Zonghai Yao, Huixue Zhou, Feiyun ouyang, Hong Yu

University of Massachusetts Amherst, University of Minnesota, Twin Cities University of Massachusetts Lowel

Introduction

EHRTutor is a multi-component framework to help patient comprehend their Electronic Health Record (EHR) notes, more specifically discharge instruction. Moreover, EHRTutor also offers a framework for generating synthetic patient education dialogues that can be used for future in-house system training.

Workflow
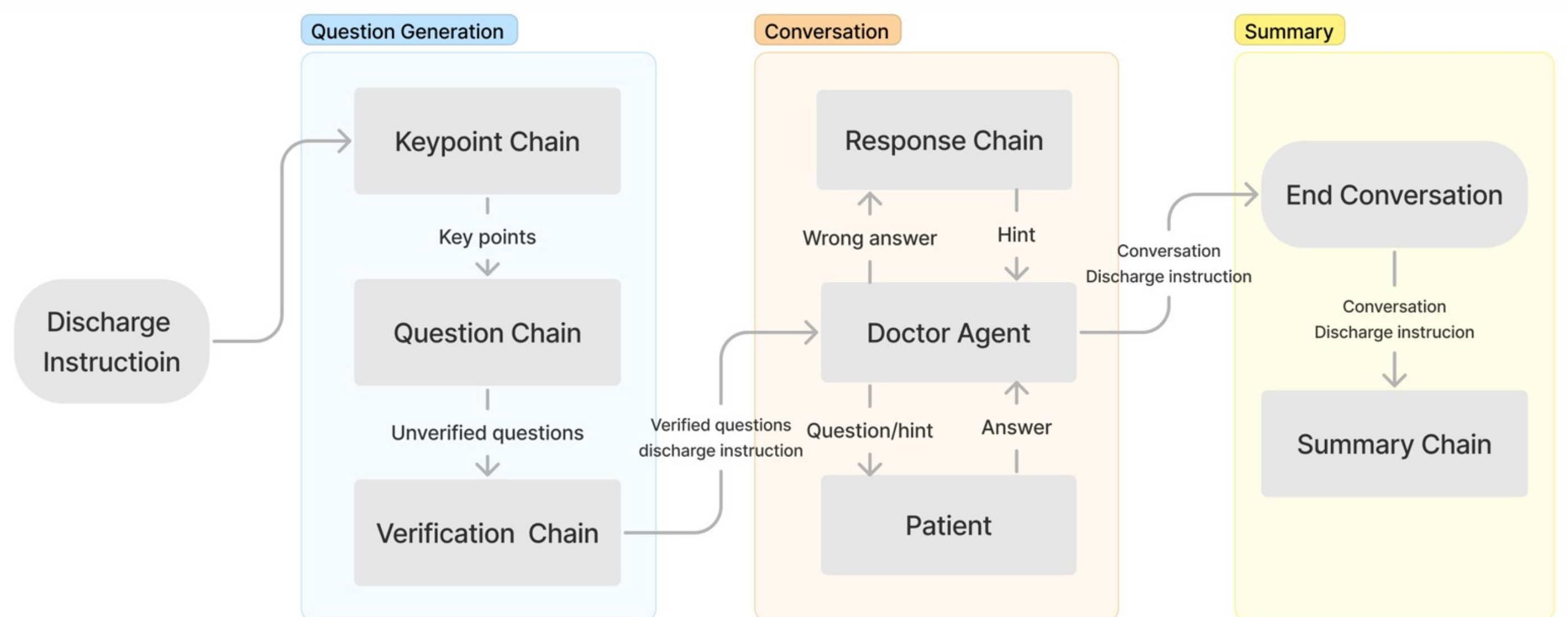
Discharge Instruction

1. You were admitted to the hospital with difficulty breathing.
   ...
2. The following changes were made to your medications:
   1. START levofloxacin 750mg by mouth ONCE daily (only one dose left. last day- [**2120-5-1**])
   2. STOP your metoprolol ER (metoprolol succinate) once daily
   3. START metoprolol tartrate 75mg by mouth TWICE daily.
   ...

Reason Chain

Thoughts: This is beginning of the conversation, I should ask a question
Action: get_question
Observation: Why you admit to hospital?

Thoughts: The patient answer is correct, move on to next question
Action: get_question
Observation: What changes were made to your medications?

Thoughts: The patient answer is incorrect for first time, I should give a hint
Action: hint_chain
Observation: Your answer is incorrect. Here is a hint: ...

Thoughts: The patient's answer is incorrect twice. So, I need to give the answer directly and ask the next question.
Action: hint_chain, get_question
Observation: Your answer is partially correct. You also stopped taking...

Conversation

1. What was the reason for being admitted to the hospital?

I have difficult in breathing

1. Your answer is correct. Now, what changes were made to your medications?

I don't remember

2. Your answer is incorrect. Here is a hint: ...

I start levofloxcain 750 mg by mouth once daily

2. Your answer is partially correct. You also stopped taking ...

...

Summary: Based on discharge instruction and conversation history

Method

Pipeline Description: Given a discharge instruction, the agent will generate questions based onthe instruction. The doctor agent will ask those questions to the patient. According to the patient'sresponse, the agent will decide which option to take: give a hint or ask the next question. Oncethe conversation is done, the agent will provide a summary to both the patient and the doctor as aconversation feedback.

Doctor agent:
- Input:
  - Verified questions
  - Discharge instruction
  - Chat history
- Output:
  - Hint / Answer / Question
- Implementation
  - Chain-of-thought: will track the reason of each step, and make decision based on precious thought and decision history.

Question chain & Verification Chain:
To generate question we made a template based on the doctor's frequent asked questions:
- Test
- Medication
- Complications & progress
- Follow up

Summary:
We generated summary based on the content of discharge instruction and the conversation history. The summary will include the follow aspects:
- Reason for Hospitalization.
- Treatment Plan.
- Medication Changes.
- Follow-up Appointments.
- When to Seek Medical Attention.
- Key points that did not remember well during conversation

Evaluation

There are four main aspects for evaluation. We want to evaluate each sub-metric in the following standards with 5 points each. Deduct one point for each unsatisfy sentence.

Metrics

Question:
1. cover rate: Do the questions cover the following categories (if applicable)?
2. Verifiable: Whether question can be answered by the discharge instruction.
3. Relevance: Do the questions relevance to discharge instruction?
4. Fluent & Concision: A concise and clear syntax and vocabulary, devoid of unnecessary question

Doctor agent:
1. Coherence: Whether the agent can make right decision

Response:
1. Relevance: Do the questions relevance to discharge instruction?
2. Sufficient: Can the patient recall or comprehend the correct answers based on hints?
3. Factuality: Is the response align with medical fact ?

Summary:
1. Cover rate: Does the summary cover all key points?

Overall Performance

Quality	Significance of the quality level
High	Question = 5 and Agent >= 4 and Response = 5 and Summary >= 4
Good	Question >= 4 and Agent >= 4 and Response >= 4 and Summary >= 4
Usable	Question >= 4 and Agent >= 3 and Response >= 4 and Summary >= 3
Useful	Question >= 4 and Agent >= 2 and Response >= 3 and Summary >= 3
Useless	Question >= 3 and Agent >= 2 and Response >= 2 and Summary >= 2
Hurtful	Question <= 3 and Agent <= 3 and Response <= 2 and Summary >= 2

Evaluation Results

(a)Human evaluate   (b) EHR Tutor   (c)Microsoft Azure ChatGPT   (d)   (e)
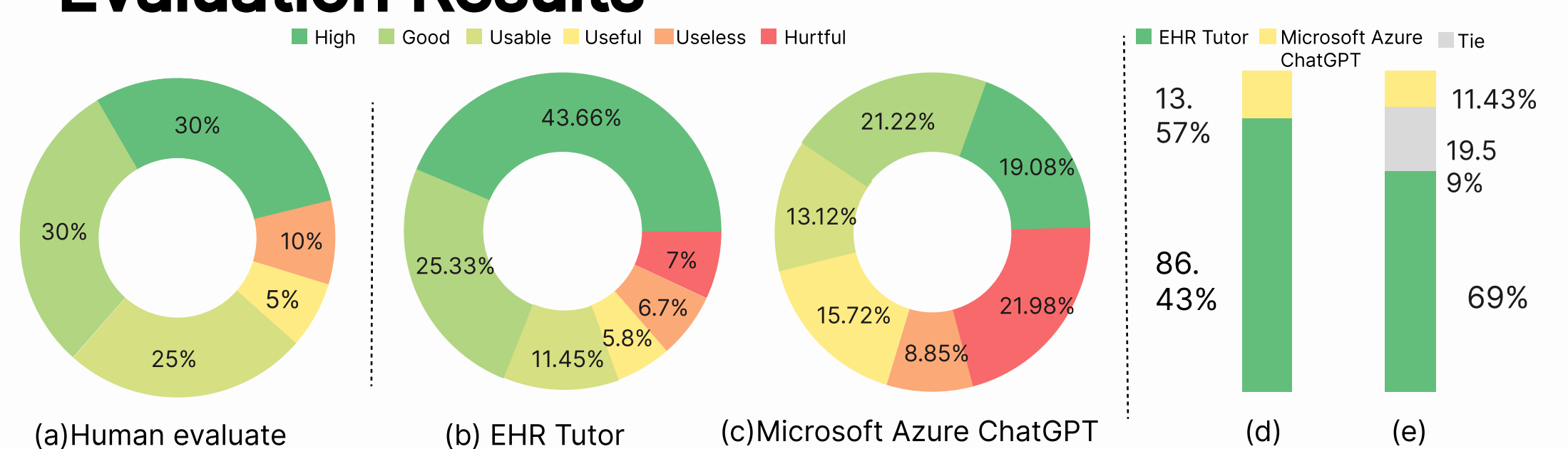
	Question			Agent	Response			Summary
	Cover rate	Relevance	Fluent	Rationality	Relevance	Sufficient	Factuality	Cover rate
	4.80	4.85	4.90	4.36	5	4.8	5	4.43

Table 1: Human evaluation results for each feature.

	Question			Agent	Response			Summary
	Cover rate	Relevance	Fluent	Rationality	Relevance	Sufficient	Factuality	Cover rate
EHRTutor	4.59	4.80	4.35	4.11	4.64	4.35	4.75	4.32
GPT-4	3.36	3.60	3.39	2.86	3.11	2.81	3.42	3.78

Table 2: LLM evaluation results for each feature.

Human Evaluation

Pros:
- All questions and most responses are helpful for the patient to understand the discharge instruction.
- All questions do not need medical background to answer.
- Most responses are factually correct.
- More natural way of communicating with patients.

Cons:
- There are some cases that do not cover all key points in the discharge instruction.
- May assume the patient only got the symptoms mentioned in the discharge instruction and consider the answer related to symptoms not shown in the instruction as incorrect.

Large Language Models Evaluation

We conduct the evaluation based on the metric described on the left, and found the result align with human evaluation.