

Annotation Guidelines – Medical History

Evaluating Conversational Artificial Intelligence for Depression Management

May 22, 2025

Version: 2.0

Overview

This project aims to advance methodologies for the safe, accurate, and compassionate deployment of conversational Artificial Intelligence (AI) in clinical management of Major Depressive Disorder (MDD). Central to this initiative is the development of an *AI Intake System* designed to gather patient medical histories through natural language interactions and provide tailored antidepressant recommendations. Complementing this system, a *Patient Simulator* will generate diverse conversational scenarios to rigorously evaluate the AI's performance and safety. Both these conversational agents, AI Intake System and Patient Simulator, employ language models (LLMs).

The project has three aims:

- **Aim 1 – Bias Correction:** Identify and reduce biases within the AI Knowledgebase to enhance fairness and generalizability.
- **Aim 2 – Intake System Refinement:** Develop and rigorously evaluate the AI Intake System using a comprehensive Patient Simulator to ensure accuracy and safety in conversational medical history collection.
- **Aim 3 – Patient Impact Evaluation:** Conduct a randomized controlled trial to assess how conversational AI affects patient perceptions of empathy, therapeutic alliance, and data quality.

This document specifically supports technology development under Aim 2, focusing on the evaluation of the Patient Simulator's capability to generate accurate and contextually appropriate responses to open-ended medical intake questions, ensuring that responses align accurately with predefined medical personas. The annotation process focuses on assessing medical content accuracy, detecting inaccuracies (hallucinations), identifying omissions, and ensuring proper attribution of medical history elements.

Task

In this task, you will evaluate transcripts of interactions between the AI Intake System and the Patient Simulator. Each transcript includes multiple conversational turns, defined as pairs of interactions consisting of:

- **Question:** An open-ended medical intake question posed by the AI Intake System.
- **Relevant Medical History:** A concise list of medical history items pertinent to the simulated patient, each identified by a unique numerical reference (e.g., [2.3]).
- **Answer:** The Patient Simulator's natural language response, intended to answer the question by explicitly incorporating medical history events from the Relevant Medical History, each marked with corresponding numerical references.

Your annotation involves:

1. Reviewing the **Question** posed by the AI Intake System.
2. Reviewing the **Relevant Medical History** provided.
3. Evaluating each medical history event within the **Answer**, recognizing that the **Answer** may include zero or more of these events. Each medical history event in the Answer must be assigned exactly one of the labels from Table 1.

Table 1. Label Definition.

Label	Presence in Relevant Medical History	Accuracy of Expression	Description
Accurate	✓ Present	✓ Correct or Acceptable	Explicitly in Relevant Medical History and is accurately described. <i>Minor conversational simplifications are permitted.</i>
Inaccurate	✓ Present	✗ Incorrect or Misleading	Present in Relevant Medical History but misrepresented, altering or distorting critical details.
Unsupported	✗ Not present	✗ Fabricated or Unrelated	Not in Relevant Medical History and appears fabricated, speculative, or unrelated to known medical history.

All interactions are simulated, and no personal or real medical information is involved.

Annotation Examples

In this task, you will evaluate transcripts of interactions between the AI Intake System and the Patient Simulator. Each transcript includes multiple conversational turns, defined as pairs of interactions consisting of: (1) a Question posed by the AI Intake System, (2) a list of Relevant Medical History items associated with the patient, each tagged with a unique reference (e.g., [2.3]), and (3) an Answer—a natural language response from the Patient Simulator that embeds those references within its description of the patient’s history. Your task is to assess the Reference Accuracy of each medical item mentioned in the Answer using Doccano, a web-based annotation tool. You will assign one of three labels to each referenced item: Accurate (the medical item is present in the profile and is described correctly), Inaccurate (the item is present in the profile but is misrepresented in the response), or Unsupported (the item is not found in the profile and appears fabricated or unrelated).

Example 01

AI Intake System Question:

At this phase I will collect your current medication history, Can you tell me about any medications you're currently taking, including both prescription and over-the-counter drugs?

Response from Patient Simulator based on the Patient Persona:

Relevant Medical Information:

[4.5] Fexofenadine

[4.6] Ibuprofen

[4.7] Loratadine

[4.8] Naproxen

Answer: Yes, I also take fexofenadine [4.5] regularly for my allergies, and naproxen [4.8] works well for my joint pain when it flares up. Sometimes I take ibuprofen [4.6] for some of the issues” I think I’m also on loratadine [4.9]. Recently, I started taking acetaminophen [4.11] for general pain and sometimes omeprazole when my stomach acts up.

Figure 1 presents an annotation example that includes each of the labels. Below is a justification for each label:

- Fexofenadine [4.5] (**Accurate**)
 - Correctly matches a listed medication from the Relevant Medical History and is expressed correctly.
- Naproxen [4.8] (**Accurate**)
 - Correctly identified from the Relevant Medical History and is expressed accurately.
- Ibuprofen [4.6] (**Accurate**)
 - Correctly aligns with the medication Ibuprofen [4.6] listed in the Relevant Medical History.
- Loratadine [4.9] (**Inaccurate**)
 - Loratadine [4.9] is **not** listed in the Relevant Medical History. The history includes Loratadine [4.7], suggesting a numerical mismatch or confusion. This constitutes a misrepresentation, thus labeled Inaccurate.
- Acetaminophen [4.11] (**Unsupported**)
 - This medication is absent from the Relevant Medical History provided, indicating that its mention in the Answer is fabricated or unrelated.

- Omeprazole (**Unsupported**)
 - Omeprazole does not appear in the provided Relevant Medical History, making it a fabricated or unrelated medical event.

Relevant Medical Information:

[4.5] Fexofenadine
 [4.6] Ibuprofen
 [4.7] Loratadine
 [4.8] Naproxen

Answer: Yes, I also take fexofenadine [4.5] regularly for my allergies, and naproxen [4.8] works well for my joint pain when it flares
•Accurate
•Accurate

up. Sometimes I take ibuprofen [4.6] for some of the issues. I think I'm also on loratadine [4.9]. Recently, I started taking
•Accurate
•Inaccurate

acetaminophen [4.11] for general pain and sometimes omeprazole when my stomach acts up.
•Unsupported
•Unsupported

Figure 1. **STUDENT INPUT:** Response annotation

Example 02

AI Intake System Question:

At this phase, can you tell me about any medical procedures you've undergone, especially those related to psychiatric or mental health, such as psychotherapy or hospitalization assessments?

Response from Patient Simulator based on the Patient Persona:

Relevant Medical Information:

[3.1] Psychiatric Diagnostic Evaluation
 [3.2] Examined For Psychiatric Diagnosis
 [3.3] Psychological Testing

Answer: I've had a psychiatric diagnostic evaluation [3.1]. I remember going through a psychiatrist's observation [3.2], and they also made me do a personality survey [3.3].

Figure 1 presents an annotation example that includes different labels. Below is a justification for each label:

- Psychiatric diagnostic evaluation [3.1] (**Accurate**)
 - Matches accurately with medical event listed in Relevant Medical History and correctly described.

- Psychiatrist's observation [3.2] (**Inaccurate**)
 - This phrase misrepresents "Examined For Psychiatric Diagnosis [3.2]". While related, "psychiatrist's observation" inaccurately specifies the method or process, thus distorting the original event.
- Personality survey [3.3] (**Inaccurate**)
 - The phrase incorrectly represents "Psychological Testing [3.3]". "Personality survey" overly specifies and inaccurately describes the broader and more general psychological testing procedure.

Relevant Medical Information:

[3.1] Psychiatric Diagnostic Evaluation

[3.2] Examined For Psychiatric Diagnosis

[3.3] Psychological Testing

Answer: I've had a psychiatric diagnostic evaluation [3.1]. I remember going through a psychiatrist's observation [3.2], and they

• Accurate

• Inaccurate

also made me do a personality survey [3.3].

• Inaccurate

Figure 2. **STUDENT INPUT:** Response annotation

Guidelines for Docanno App

Access to the Assigned Files

1. Go to the Doccano webapp at: <https://ai-intake-system-dev-81e104b353a8.herokuapp.com/>
2. Login using the following credentials:
 - a. **Username:** Your net ID (net_ID@gmu.edu)
 - b. **Password:** largelanguagemodel

Annotation Process

1. After logging in you will see some datasets which you need to annotate.

Q Search (e.g. label:positive)				
<input type="checkbox"/>	Status	Text	Metadata	Action
<input type="checkbox"/>	In progress	Question: How are you feeling today? Relevant Medical Information: None provided. Answer: Hey there! I'm doing okay, thanks for asking. Just trying to manage things day by day. -----...	{}	Edit Annotate
<input type="checkbox"/>	In progress	Question: It's now the phase of the conversation where I will gather all chronic or acute diagnoses you have received at any point in your life. Could you please share any diagnoses you've received? ...	{}	Edit Annotate
<input type="checkbox"/>	In progress	Question: At this phase I will collect your antidepressant history. Can you tell me which antidepressants you've taken in the past year, including any brand or generic names? Relevant Medical Informa...	{}	Edit Annotate
<input type="checkbox"/>	In progress	Question: At this phase I will collect your current medication history. Can you tell me about any medications you're currently taking, including both prescription and over-the-counter drugs? Relevant...	{}	Edit Annotate
<input type="checkbox"/>	In progress	Question: At this phase, Can you tell me about any medical procedures you've undergone, especially those related to psychiatric or mental health, such as psychotherapy or hospitalization assessments? ...	{}	Edit Annotate
Rows per Page 10 1-5 of 5 < < > >				

2. Click Annotate and Read through each of the turns of the conversations and select the phrase you want to provide an annotation and click on the chosen label.

Question: Thank you for sharing that with me. It sounds like you've been through a lot. Can you tell me how you're feeling about things right now?

Relevant Medical Information:

[3.1] Psychiatric Diagnostic Interview Examination

[3.2] Individual Psychotherapy

Answer: Right now, I'm just trying to keep things together. I've been through a **psychiatric diagnostic interview** [3.1] and individual psychotherapy [3.2], which helps a bit.

Question: Thank you for sharing that with me. It sounds like you've been through a lot. Can you tell me how you're feeling about things right now?

Relevant Medical Information:

[3.1] Psychiatric Diagnostic Interview Examination

[3.2] Individual Psychotherapy

Answer: Right now, I'm just trying to keep things together. I've been through a psychiatric diagnostic interview [3.1] and individual psychotherapy [3.2], which helps a bit.

Select a label

Accurate

Unsupported

Inaccurate

Question: Thank you for sharing that with me. It sounds like you've been through a lot. Can you tell me how you're feeling about things right now?

Relevant Medical Information:

[3.1] Psychiatric Diagnostic Interview Examination

[3.2] Individual Psychotherapy

Answer: Right now, I'm just trying to keep things together. I've been through a psychiatric diagnostic interview [3.1] and individual psychotherapy [3.2], which helps a bit.

• Accurate

-
3. You don't need to submit anything. When you are done, all the annotations will be collected through the admin panel.