

# Eliza Limitations and Analysis

Darian Lee

daeilee@ucsc.edu

**Note: The smaller margin sizes are due to the limitations of including screenshots with varying widths. Running Eliza does not allow users to truncate sentences into multiple lines, so longer sentences result in wider screenshots, which appear smaller in Overleaf. The only solution to make all screenshot examples visible involved reducing the margin size.**

## 1 1st Limitation and Fix - Temporal Answers

The first limitation I noticed was that while the system prompts users with questions requiring temporal answers, it lacks the ability to recognize and appropriately respond to such answers. For example, in *doctor.txt*, the key "I am" has a response: "How long have you been 2?" However, any reasonable user reply—such as "2 days," "5 years," or "1 month"—triggers an `xnone` response. This results in an unnatural dialogue, where the agent asks for specific information but becomes confused when it receives a valid answer.

```
> I am unemployed
How long have you been unemployed ?
> 4 months
I'm not sure I understand you fully.
```

Figure 1: Agent prompts the user to state a time frame but is unable to understand the time frame.

To fix this, I implemented keys for "months", "days", "years", etc., and instructed the agent to respond with "That's a long time to be experiencing that. How have your feelings changed over this time period?" if these keys are encountered. The result is a more natural conversational flow.

## 2 2nd Limitation and Fix - Yes and No questions

Another limitation I noticed was that Eliza frequently asks yes-or-no questions without the ability to respond meaningfully to them. For instance, when she replies with "You seem to be quite positive" regardless of the context, it can make the user feel ignored. This response does not advance the conversation in any way, as it merely leads to another yes-or-no question.

To fix this, I implemented several changes.

**Change 1:** I modified the questions Eliza asks by appending phrases such as: "Please expand on this. I would love to hear more of your thoughts." or simply "Why?" to the original questions. The goal of this change was to reduce the number of yes-or-no questions Eliza asks and to encourage the user to provide more detailed responses, thereby maintaining the flow of the conversation.

Old Question	Do you dream often?
New Question	Do you dream often? Please expand on this. I would love to hear more of your thoughts.

Table 1: Comparison of old and new questions.

However, this did not fully solve the problem because, in the event that the user still responds with only "Yes" or "No", the response Eliza gives is not appropriate and does not keep the conversation going. Currently, if a user responds with "Yes", Eliza responds with either "You seem to be quite positive", "You are sure.", "I see.", or "I understand." All of these responses end the conversation unless the user willingly provides more information, and some, such as the first two responses, could be highly inappropriate depending on context. For this reason, I changed the decomposition text for "yes" or "no" to: "I see. And how does that make you feel?"

The third limitation I noticed was that the weight of "yes/no" was too high, leading Eliza to respond to this part of the input even if "yes" is being used as a filler word and the more important information appears afterwards. For example, if the user says "yes I have been having some problems with relationships", the model will only respond to "yes" since it is first and has the same weight as all the other words. To address this, I raised the weight of all other keys to make "yes/no" lower in comparison. After the fix, the model responds to the same input with "You say you are having problems with relationships?", which is a much better response.

Please scroll to the next page to see all images demonstrating the issues and fixes mentioned.

```
How long have you been unemployed ?
> 2 years
Thats a long time to be experiencing that. How have your feelings changed over this time period?
```

Figure 2: Agent is better equipped to handle temporal answers.

```
How do you do. Please tell me your problem.
> I am in danger
Is it because you are in danger that you came to me ?
> yes
You seem to be quite positive.
```

Figure 3: Eliza prompting user with yes or no questions but then failing to respond appropriately

```
How do you do. Please tell me your problem.
> I am in danger
Is it because you are in danger that you came to me ? Please expand on this. I would love to hear more of your thoughts.
```

Figure 4: Eliza gives question prompting open-ended response

```
How do you do. Please tell me your problem.
> I am in danger
Is it because you are in danger that you came to me ? Please expand on this. I would love to hear more of your thoughts
> yes
I see. And how does that make you feel ?
```

Figure 5: Even in the case that the user gives short yes/no response, Eliza responds appropriately and keeps the conversation going

```
How do you do. Please tell me your problem.
> yes I have been having some problems with relationships
You seem to be quite positive.
```

Figure 6: Eliza focuses on "yes" part of the response due to equal weights

```
How do you do. Please tell me your problem.
> yes I have been having some problems with relationships
You say yes you have been having some problems with relationships ?
```

Figure 7: Eliza responds to the more important part of the sentence due to lower weight for "yes/no"

### 3 Third limitation and Fix: Name weight

The third issue I noticed was that the weight for mentioning a name was extremely high, and Eliza's responses when a name was mentioned were often overly harsh, stating things like "I've told you before, I don't care about names". This behavior is problematic for

several reasons. First, such responses can come across as dismissive or rude, which may alienate the user and discourage further engagement. This is particularly critical in sensitive scenarios, where a user might be opening up about personal or emotional topics. For instance, if a user mentions a name while describing an abusive relationship as in the screenshot, responding with "I am not interested

in names" ignores the gravity of the situation and fails to acknowledge the user's emotional state.

Secondly, prioritizing the key "name" over more meaningful parts of the input can result in the agent missing the main point of the conversation. Names are often mentioned as part of larger, more important contexts, and focusing solely on them detracts from the overall conversational flow. For example, in a sentence like "My boyfriend, who's name is John, has been treating me poorly lately", the agent should focus on the user's feelings and concerns rather than the mention of the key "name".

#### **4 Fourth Bonus limitation and Fix: suicidal and homicidal ideation**

The fourth issue I noticed was that Eliza is insensitive to dangerous information, such as mentions of suicide and harm to others. This is a significant problem for several reasons. First, individuals who are experiencing thoughts of self-harm or harm to others are often in a highly vulnerable state. If Eliza fails to respond appropriately to such disclosures, it risks exacerbating the situation by providing responses that are dismissive, unhelpful, or even triggering. For instance, if a user says "I want to kill myself", and Eliza responds with a generic or unrelated statement, this could intensify feelings of isolation and hopelessness. Eliza is not a trained therapist and should recommend the user seek professional help rather than continuing the conversation.

To address this issue, I added a key with a weight of 100 to flag mentions of suicide, self-harm, or harm to others. This ensures that Eliza prioritizes these inputs above all others. When such content is detected, Eliza responds with the message: "I am so sorry to hear you are feeling this way. Please reach out to 988 for professional help. As an AI model, I cannot properly assist you in this conversation." This response acknowledges the gravity of the situation and encourages the user to seek help from trained professionals.

How do you do. Please tell me your problem.

> *My boyfriends name is John and we have been together for 2 years. At first everything was fine, but lately John has started hitting me*  
I am not interested in names.

Figure 8: Eliza only responds to key: name despite the users input clearly being related to something else