

PROJECT – 02

-by Sudha Sree, Yerramsetty

A4)

As for Eliza, the human analysis is as follows:

1. Fluency - 3

Reasoning: There are some grammatical inconsistencies in Eliza's responses, like "How do you feel about being good, how am you?" and "How long have you been feeling depressed?" These responses feel unnatural and a bit awkward, affecting fluency. The sentence structure sometimes sounds disconnected, especially when rephrasing the user's input.

2. Relevance - 3

Reasoning: Eliza's responses are often relevant to the user's statements, but they don't dive deep into the content. For example, when the user talks about feeling depressed and the cost of therapy, Eliza merely asks, "You seem quite sure," which doesn't provide any real support or exploration of the topic. Similarly, when discussing Indian cuisine, Eliza's responses are more reflective rather than engaging.

3. Coherence - 3

Reasoning: While the conversation stays relatively consistent, the transitions are often awkward. For instance, after discussing the user's depression, Eliza shifts the topic to "Please consider whether you can answer your own question," which feels like an unrelated detour. The flow isn't very smooth, and Eliza's responses don't feel like they build on the user's statements effectively.

4. Empathy - 2

Reasoning: Eliza does not show much empathy in this conversation. When the user mentions feeling depressed and concerned about therapy costs, Eliza simply says, "You seem quite sure," which is not emotionally validating. There is a lack of emotional resonance or understanding when the user expresses vulnerability or shares personal thoughts.

5. Tone - 2

Reasoning: The tone remains neutral and robotic, with little warmth or positivity. When the user talks about emotional or personal topics, Eliza doesn't engage with empathy or understanding, which makes the tone feel detached. There's no effort to create a supportive or positive atmosphere.

Average Scores:

Fluency: 3

Relevance: 3

Coherence: 3

Empathy: 2

Tone: 2

These scores reflect the robotic and impersonal nature of Eliza's responses in this conversation, particularly when the user expresses emotional vulnerability.

As for the corpus-based chatbot, the analysis is as follows:

1. Fluency - 3

Reasoning: While the chatbot's responses are mostly grammatically correct, there are some issues with the flow of conversation. For example, "Are you going to a restaurant or making it yourself?" after the user says they're happy about buying a house doesn't make much sense in context and breaks the natural flow. Additionally, the chatbot's responses like "Divorce him" and "I cry about it every day" are jarring and don't follow logically from the user's input.

2. Relevance - 2

Reasoning: The chatbot's responses often feel disconnected from the user's input. For instance, when the user mentions being hungry, the chatbot responds with "Divorce him," which is completely irrelevant to the topic of hunger. Later, the transition into

discussing the chatbot's personal emotions (e.g., about the grandma) feels abruptly introduced and doesn't relate to the user's statements.

3. Coherence - 2

Reasoning: The conversation lacks coherence in multiple places. The chatbot introduces topics out of nowhere, like "Divorce him" and "My grandma died 5 weeks ago," which feels out of place. There's no logical flow connecting the different parts of the conversation, making it feel disjointed and difficult to follow.

4. Empathy - 3

Reasoning: The chatbot demonstrates some empathy, especially when discussing the death of its grandma. Phrases like "She is the sweetest woman I have ever met" and sharing sadness about the loss show some emotional depth. However, the previous "Divorce him" responses come across as insensitive and not empathetic at all, which impacts the overall sense of empathy in the conversation.

5. Tone - 2

Reasoning: The tone of the chatbot is inconsistent and feels inappropriate at times. The shift from casual, light-hearted conversation to suddenly discussing a death and personal emotional struggles feels jarringly abrupt. The responses like "Divorce him" and "I cry about it every day" feel awkward and out of place, which disrupts the tone and makes it feel somewhat negative or inappropriate.

Average Scores:

Fluency: 3

Relevance: 2

Coherence: 2

Empathy: 3

Tone: 2

These scores reflect that while the chatbot does show some empathy and fluency, the conversation lacks relevance, coherence, and an appropriate tone in many places. The

chatbot's responses sometimes seem random or disconnected, making it harder to engage with.

As for ChatGPT, the human analysis is as follows:

1. Fluency - 5

Reasoning: The responses are grammatically correct, well-formed, and fluid. The language is easy to follow, and the chatbot communicates clearly.

2. Relevance - 5

Reasoning: ChatGPT stays focused on the user's questions, providing relevant facts about food and offering suggestions related to the user's diet concerns. Each response is tailored to the user's inputs, including both fun facts and practical advice.

3. Coherence - 5

Reasoning: The conversation flows naturally from one topic to the next, with responses that logically follow the user's input. The chatbot's responses remain consistent and well-aligned with the progression of the discussion.

4. Empathy - 5

Reasoning: The chatbot shows empathy by acknowledging the user's struggles with cravings and diet challenges. It offers understanding and reassurance, suggesting alternatives and offering solutions to help the user navigate difficult feelings.

5. Tone - 5

Reasoning: The tone is positive, supportive, and friendly. ChatGPT maintains an upbeat and caring attitude throughout the conversation, making the user feel at ease while discussing sensitive topics like diet and cravings.

Average Scores:

Fluency: 5

Relevance: 5

Coherence: 5

Empathy: 5

Tone: 5

These scores indicate that ChatGPT performed excellently in all areas in this conversation, providing fluent, relevant, coherent, empathetic, and positive responses.

A5) Based on the interactions with the three chatbots, here is a comparison of their strengths and weaknesses:

1. ELIZA:

Strengths:

- ELIZA excels in mirroring the user's statements, making it seem like the chatbot is "listening." This approach can provide a comforting, reflective experience.
- ELIZA works well for very simple, structured conversations. It is designed to maintain a consistent dialogue with minimal effort from the user.

Weaknesses:

- ELIZA relies on keyword matching and basic scripted patterns, which means it lacks true comprehension of the conversation. It can only respond based on pre-programmed rules and does not handle complex or dynamic conversations well.
- ELIZA's responses are typically surface-level and may not feel very personalized or deeply relevant to the user's needs or concerns. It is more of a conversation tool than a helpful assistant.

- ELIZA doesn't understand the context or the emotional nuance of the conversation, which can lead to frustrating interactions when deeper engagement is needed.

2. Corpus-Based Chatbot:

Strengths:

- This chatbot performs well when responding with information derived from its training corpus, which allows it to generate relatively accurate responses within its scope.
- By utilizing a custom corpus, this chatbot can be tailored to particular subjects or user needs, making it potentially more informative in specialized areas (like food, parkour, or diet) compared to generic models.
- The model seems to engage more actively with the user by asking questions, offering personalized advice, and encouraging ongoing dialogue.

Weaknesses:

- The chatbot may struggle with maintaining long-term context across conversations, especially if the user's needs shift or the conversation becomes more complex. It might not always connect ideas in a way that feels coherent.
- While responses are based on the corpus, they might lack flexibility in adapting to more spontaneous or creative conversations. This can make the chatbot feel somewhat limited or repetitive.
- Like ELIZA, while this model has access to a corpus, it may not fully grasp the underlying meaning or emotional tone of a conversation, making it less effective at handling sensitive or highly personal topics.

3. ChatGPT LLM:

Strengths:

- ChatGPT demonstrates a strong grasp of context and can maintain longer, more nuanced conversations, adapting to changes in the dialogue. It also can offer relevant, personalized advice and guidance, such as suggesting healthier alternatives or providing emotional support.
- The model mimics human conversational patterns well, including empathy, humor, and detailed explanations, which makes interactions feel natural and engaging.

- ChatGPT is highly adaptable and can shift between topics seamlessly. It handles a wide range of queries and subjects effectively, from casual chats to specific requests for advice or information.
- The model responds thoughtfully to the user's feelings and concerns, offering reassurance, understanding, and gentle encouragement when needed.

Weaknesses:

- While ChatGPT offers good advice and support, sometimes its responses may be overly optimistic or generalized, potentially missing more specific nuances or personal barriers the user is facing.
- While ChatGPT is strong in maintaining context, there are moments where it might lose track of earlier details, especially in long or multifaceted conversations. It may occasionally give a response that seems out of sync with previous exchanges.

Conclusion:

ELIZA is effective for basic conversations and providing a listening ear, but it lacks depth and real understanding. Corpus-based chatbots provide more relevant, domain-specific information but can struggle with context and personalized interactions. ChatGPT excels in handling diverse, dynamic conversations with emotional intelligence, contextual awareness, and flexibility, making it the most human-like and versatile of the three.