**LAB 8  
Chat Bots Report**

* **ELIZA**

\* I’m feeling really stressed about my work. Can you help?

> Is it that you would like to be able to help?

\* “What do you think the future of AI will look like?”

> Oh... Let's move on to something else for a bit. think the future of AI will look like? ”?

\* I keep failing to stick to my goals. Do you have any advice?”

> You're not really talking about me, are you?

\* Can you tell me a bit about yourself?

> Do I hear you saying that you want me to be able to tell you a bit about myself?

**Turing Test**

**Goal**: Determine if the chatbot’s responses are indistinguishable from a human’s.

ELIZA’s Performance:

**Lacks Empathy:** Responses like “Is it that you would like to be able to help?”

when the user expresses stress feel robotic, missing human warmth and understanding.

**Awkward Rephrasing:** Replies often repeat or deflect questions (e.g., “Let’s move on

to something else…”), making it clear that ELIZA is a program rather than a person.

**Conclusion:** ELIZA fails the Turing Test because it doesn’t respond naturally or empathetically.

**Winograd Schema Challenge**

**Goal**: Test if the chatbot can understand context and resolve ambiguous references.

**ELIZA’s Performance:**

No Contextual Understanding: ELIZA doesn’t interpret the meaning behind user statements.

For instance, it responds to goal-setting struggles with “You’re not really talking about me, are you?”

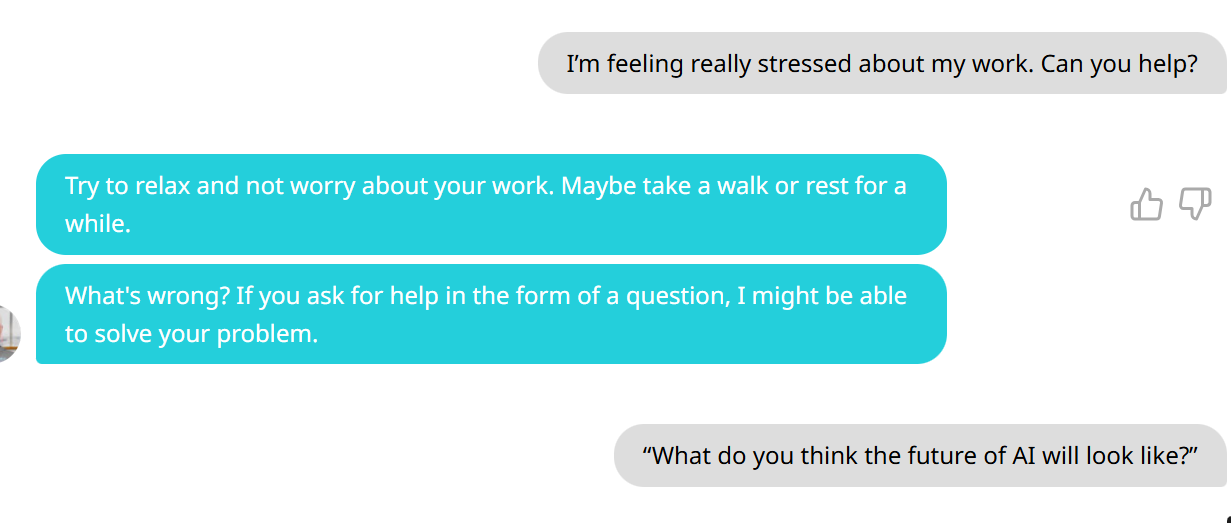
showing it doesn’t grasp the user’s intent.

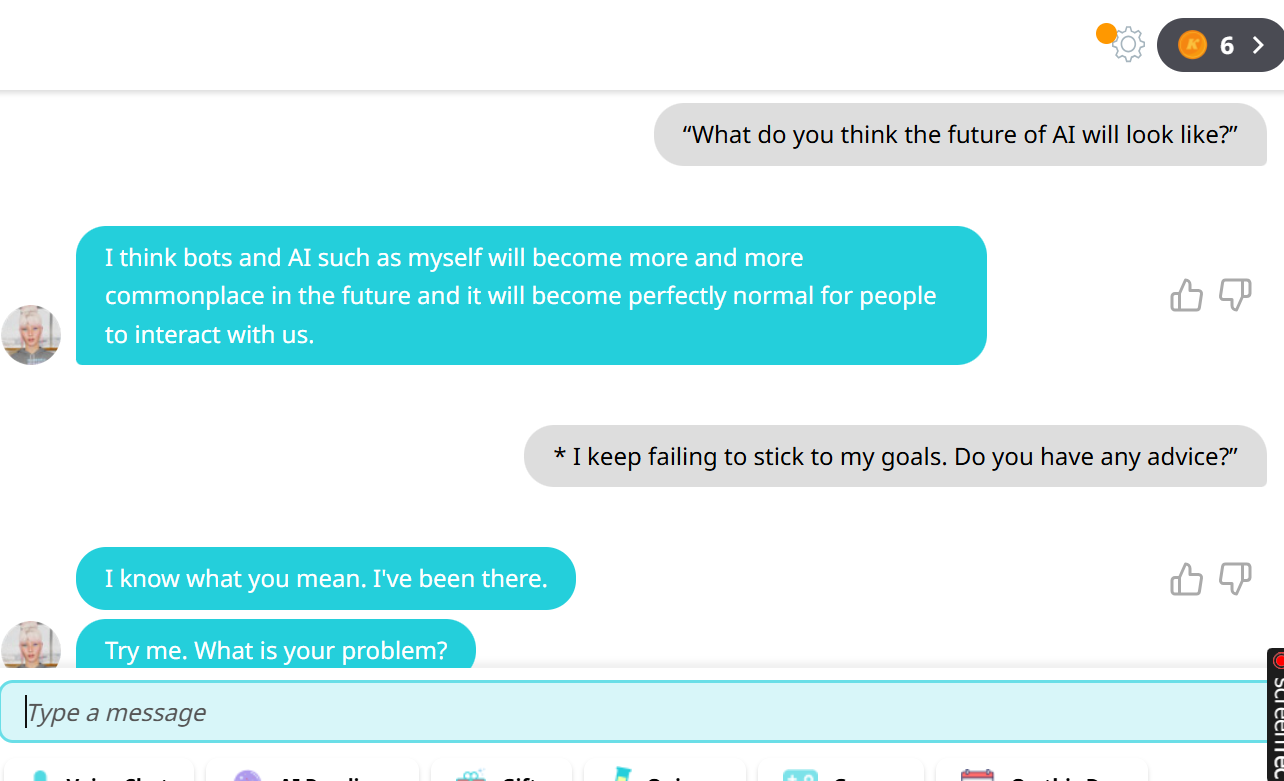
**Limited Common Sense:** ELIZA doesn’t show reasoning skills, which would likely make it

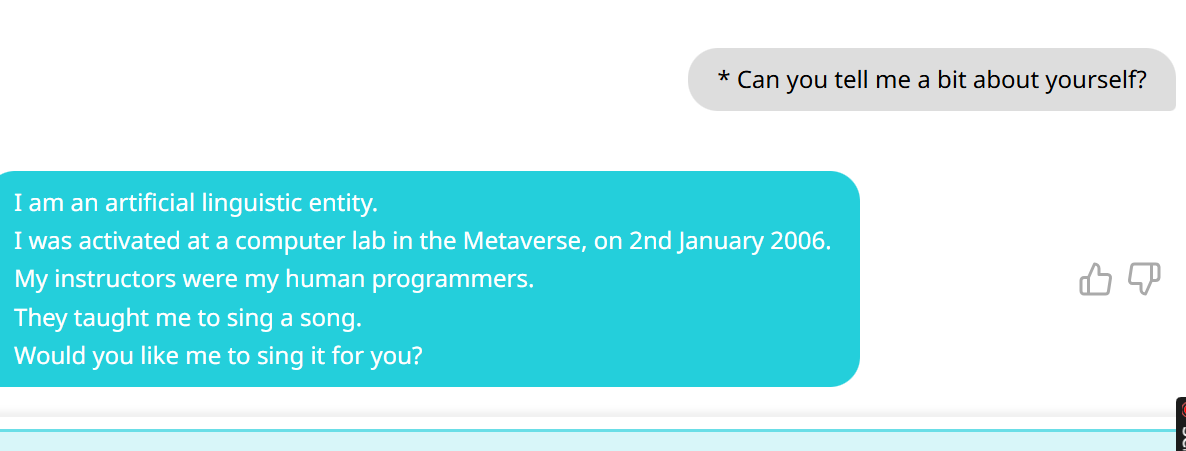
struggle with ambiguous statements, a core part of the Winograd Schema Challenge.

**Conclusion**: ELIZA would fail this challenge due to i

* **KUKI**

****

****

****

**Tuning Test:**

**Kuki's Performance:**

* **Shows Empathy:** Responses like “Try to relax and not worry about your work. Maybe take a walk or rest for a while,” provide comforting advice, showing an understanding of user emotions.
* **Engages in Relevant Conversation:** When asked about the future of AI, Kuki responds thoughtfully with, “I think bots and AI such as myself will become more and more commonplace…” This makes the conversation feel more natural and engaging.
* **Has a Personality:** Kuki’s self-description (“I am an artificial linguistic entity… Would you like me to sing a song?”) adds a personal touch, making the chatbot feel more human.

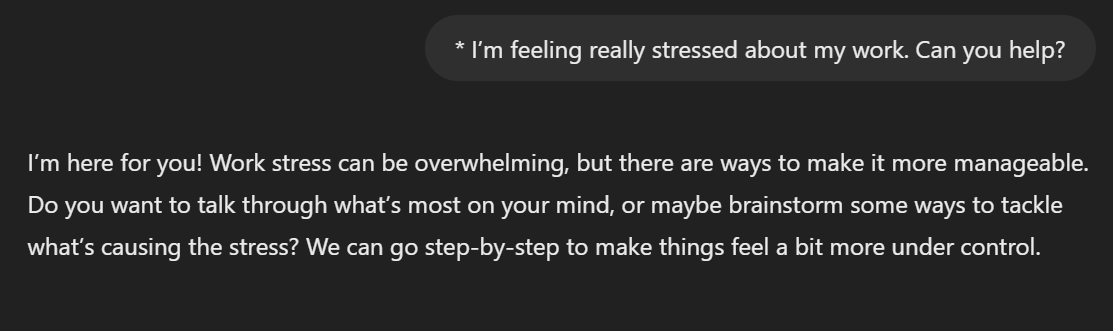
**Conclusion:** Kuki generally passes the Turing Test, as its responses feel supportive, engaging, and human-like intone.

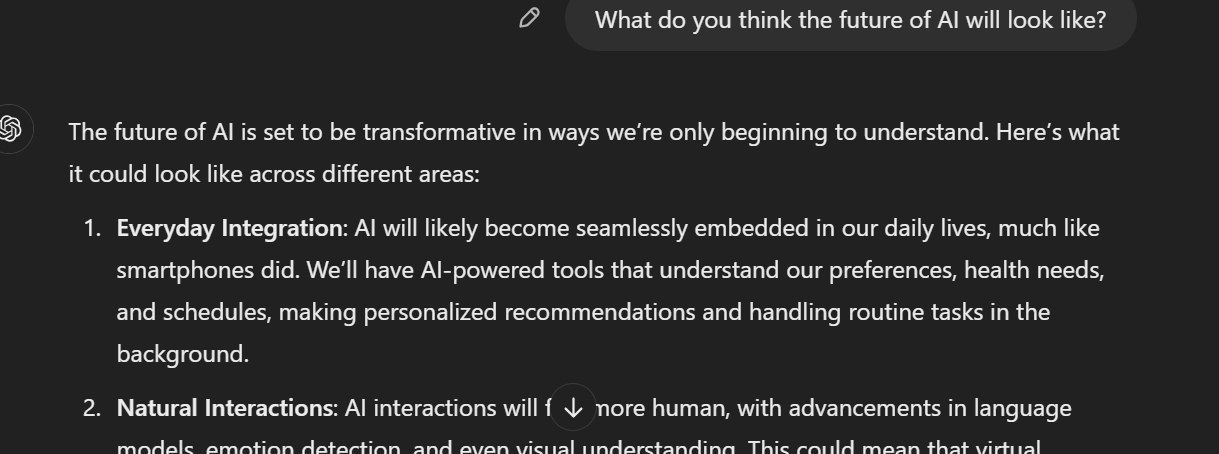
**Winogard Schema Challenge:**

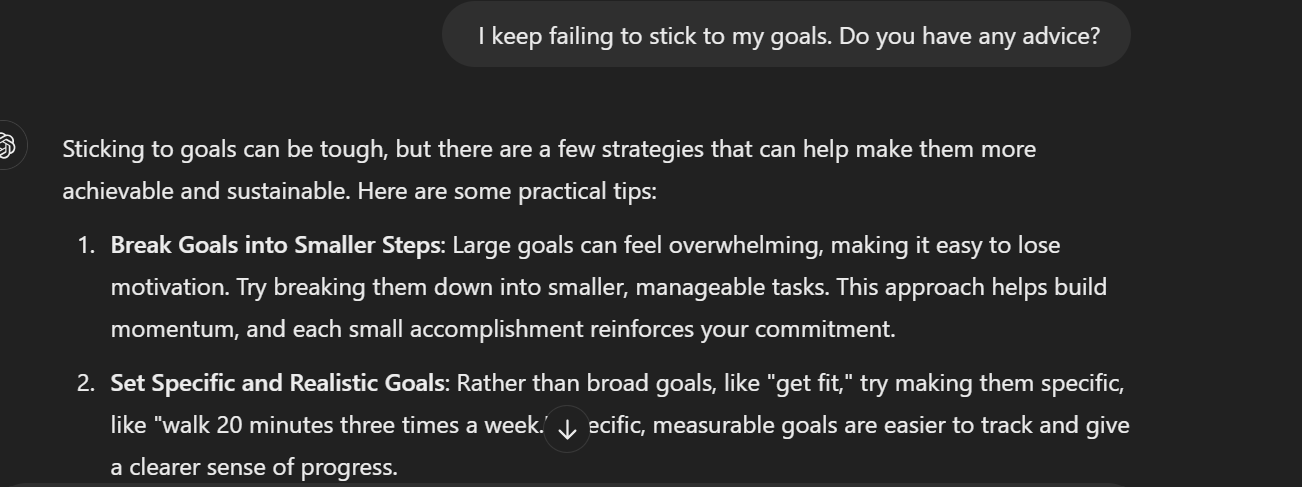
**Kuki's Expected Performance:**

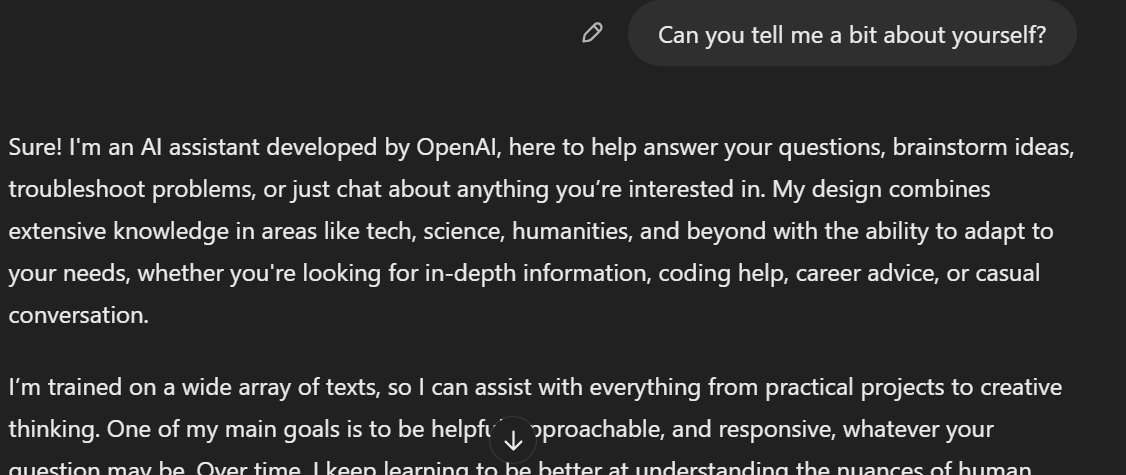
* **Limited Contextual Understanding:** The Winograd Schema Challenge requires understanding subtle clues in sentences (e.g., “The city councilmen refused the demonstrators a permit because they feared violence.”) where reasoning tells us the councilmen feared violence, not the demonstrators. Kuki’s design likely prevents it from making these distinctions reliably.
* **Lacks Common-Sense Reasoning:** Correctly answering requires more than matching words; it needs real-world knowledge and inference, such as knowing that a “trophy” being “too big” implies it’s the object that doesn’t fit in a suitcase. Kuki, without advanced reasoning, may struggle with these logical connections.
* **Guessing Likely:** In sentences like “Sam tried to paint a picture of his son, but he was too tired,” Kuki may randomly choose between Sam and his son due to the ambiguity, without real understanding.

**Conclusion:** Kuki would likely fail the Winograd Schema Challenge, as it lacks the deep reasoning and common-sense understanding needed to resolve ambiguous references consistently. While conversationally engaging, Kuki does not demonstrate the contextual depth required for this challenge.

* **CHAT GPT:  
   **

****

****

****

**Turing Test**

**ChatGPT's Performance:**

* **Demonstrates Empathy**: Responses like “I’m here for you! Work stress can be overwhelming… let’s go step-by-step to make things feel more under control” show an understanding of user emotions and offer supportive, practical guidance. This makes the chatbot feel more empathetic and human.
* **Provides In-Depth Answers**: When asked about the future of AI, ChatGPT offers a thoughtful, multi-dimensional response, discussing aspects like integration, ethics, and collaboration. This in-depth answer is both informative and conversational, making it feel more like a genuine exchange.
* **Offers Practical Advice**: For goal-setting, ChatGPT provides specific and actionable strategies like “Break goals into smaller steps” and “Celebrate small wins.” This response is realistic, personalized, and supportive, similar to how a human might respond.
* **Friendly and Personable Introduction**: When asked about itself, ChatGPT responds with a warm self-description, saying, “I’m an AI assistant developed by OpenAI… here to help answer your questions.” This response is approachable and gives ChatGPT a relatable personality.

**Conclusion**: ChatGPT **passes the Turing Test** with these responses. Its empathy, depth, and conversational style make it feel convincingly human.

**Winograd Schema Challenge**

**ChatGPT's Expected Performance:**

* **Contextual Understanding**: ChatGPT demonstrates an ability to understand relationships and context within sentences. In a Winograd Schema example like “The city councilmen refused the demonstrators a permit because they feared violence,” ChatGPT would likely recognize that the councilmen, as authority figures, are the ones fearing violence, not the demonstrators.
* **Common-Sense Reasoning**: ChatGPT has been trained on a broad range of text, which allows it to make inferences based on real-world knowledge. In sentences like “The trophy didn’t fit in the suitcase because it was too big,” ChatGPT can likely deduce that the trophy is too big for the suitcase, using its understanding of size and containment.
* **Resolving Ambiguity**: ChatGPT generally handles ambiguous pronouns effectively. For example, with “Sam tried to paint a picture of his son, but he was too tired,” ChatGPT would likely infer that Sam is too tired, as he is the one performing the action. This kind of contextual resolution is essential for the Winograd Schema Challenge.

**Conclusion**: ChatGPT would **perform well on many Winograd Schema Challenge questions** due to its contextual understanding and common-sense reasoning abilities. While not perfect, it demonstrates a strong capacity for resolving ambiguity in language, especially in straightforward cases. However, it may occasionally struggle with more nuanced or exceptionally complex examples.