



# Charts: Personality

Emily Wu and Esther Kim



# Roadmap of Presentation

1. Recap of Project and Current Project Status
2. Factors that Make a Conversation Engaging
3. Question #1
4. Current Models
5. Question #2

---

# Recap of Project and Current Project Status



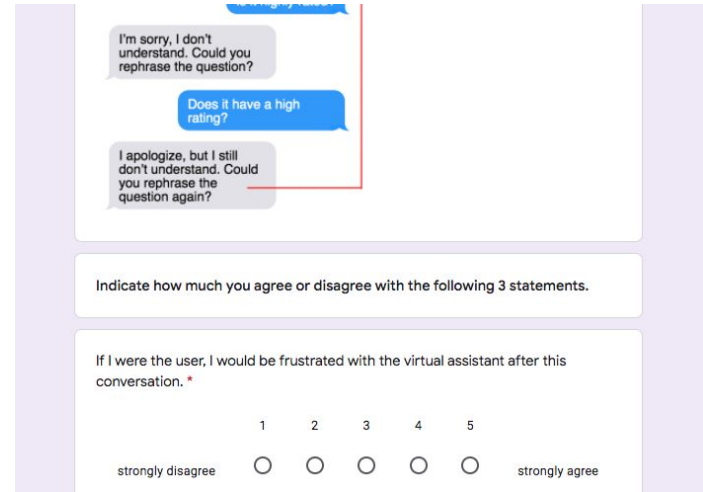
## Recap of Project

### **Central Question:**

What factors make a user's experience with a conversational AI more positive and engaging?

# Current Project Status

- Prepared mini-lecture (happening now!)
- Finished launching first survey (thanks Jackie and Silei!)
  - 21 responses from MTurk
  - Will analyze responses this weekend
- Next steps
  - Finalize domain
  - Write first draft of dialogue script
  - Conduct initial user testing of script (tentative)
  - Design second survey with more refined scenarios (if needed)



The screenshot displays a survey interface with a chat history and a Likert scale. The chat history shows three messages: a grey bubble asking for a rephrase, a blue bubble asking for a high rating, and another grey bubble asking for a second rephrase. Below the chat is a survey instruction: "Indicate how much you agree or disagree with the following 3 statements." The first statement is: "If I were the user, I would be frustrated with the virtual assistant after this conversation. \*". Below the statement is a Likert scale with five radio buttons labeled 1 to 5, with "strongly disagree" on the left and "strongly agree" on the right.

I'm sorry, I don't understand. Could you rephrase the question?

Does it have a high rating?

I apologize, but I still don't understand. Could you rephrase the question again?

Indicate how much you agree or disagree with the following 3 statements.

If I were the user, I would be frustrated with the virtual assistant after this conversation. \*

1 2 3 4 5

strongly disagree      strongly agree

---

# Factors that Make a Conversation Engaging

# Conceptual Metaphors

An understanding of abstract or complex ideas using simple terms



Tay: "AI that's got no chill"

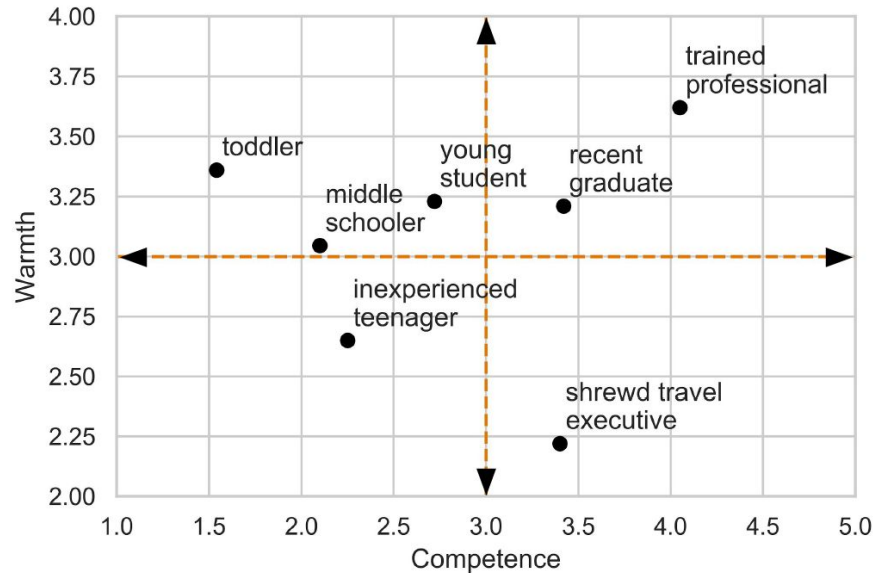


Xiaoice: "an empathetic ear"

- Short description attached to AI
- Provide an understanding of functionalities and intentions
- Can influence user's pre-use expectations of AI

# Stereotype Content Model

- Warmth and competence are the principal axes of human social perception
- Warmth: good-naturedness, sincerity
- Competence: intelligence, responsibility, skillfulness







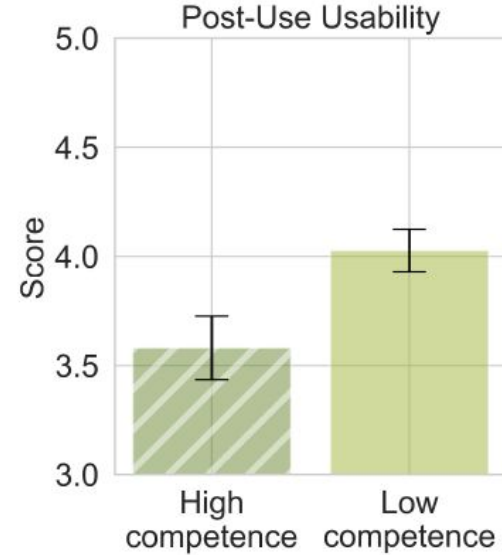
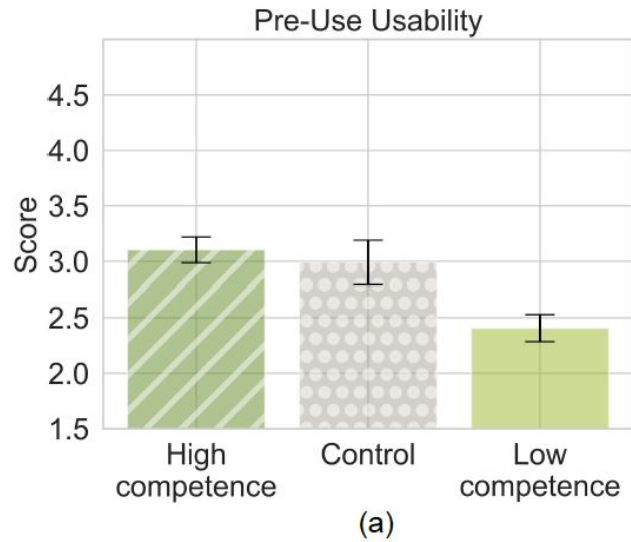
# User Evaluations

Measures rated on 5-point Likert scale:

- Usability: “Using the AI will be a frustrating experience.”
- Warmth: “The AI system is good-natured.”
- Desire to cooperate: “How likely would you be to cooperate with this AI?”
- Intention to adopt: “Based on your experience, how willing are you to continue using this service?”

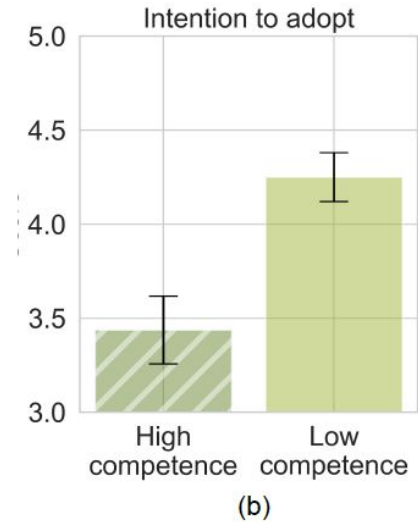
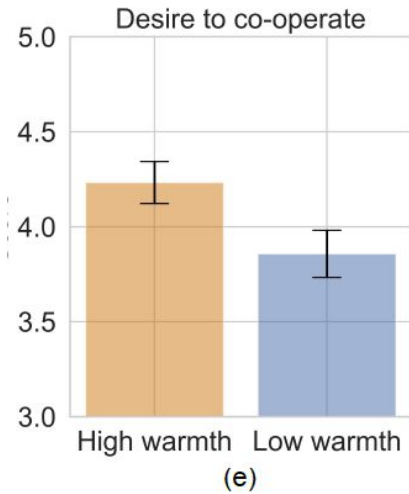
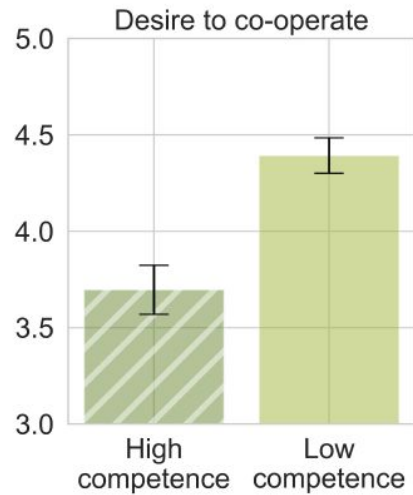


# User Evaluation Results





# User Evaluation Results



### High warmth conversation

Is wifi included?

yes

And the Paris location does that include breakfast?

Breakfast is offered

Do they speak english?

I am sorry. I don't have that information.

Do I get my own bathroom?

I am sorry. I don't have that information.

How far from the Empire State Building is the New York hotel?

I am sorry. I don't have that information.

Does berlin offer parking?

I am sorry. I don't have that information.

Do any of these hotels offer spa services?

yes

New York and Paris

Do any have a minibar?

I am sorry. I don't have that information.

I would like to book the Paris hotel.

### Low warmth conversation

Hi

Hello, I am planning for a trip to New York from Montreal

We have options for you.

Luma Hotel - 4 Star: Dates Available: August 26 to 30 (5 nights) Price: \$850.90

Yes

is it available on that dates?

yeah! that sounds good.

Okay. I'll book that for you right away.

Do you want me to show you your return flight options?

yes

United - Departure at 2:00 pm local time - Direct Flight - \$311

Does berlin offer parking?

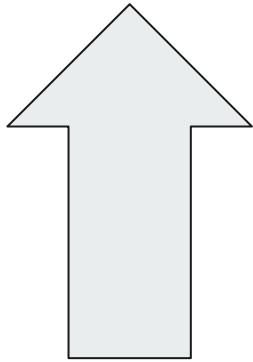
I am sorry. I don't have that information.

Great. it will work.

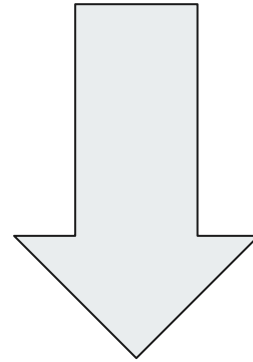
Participant AI system



# Takeaways



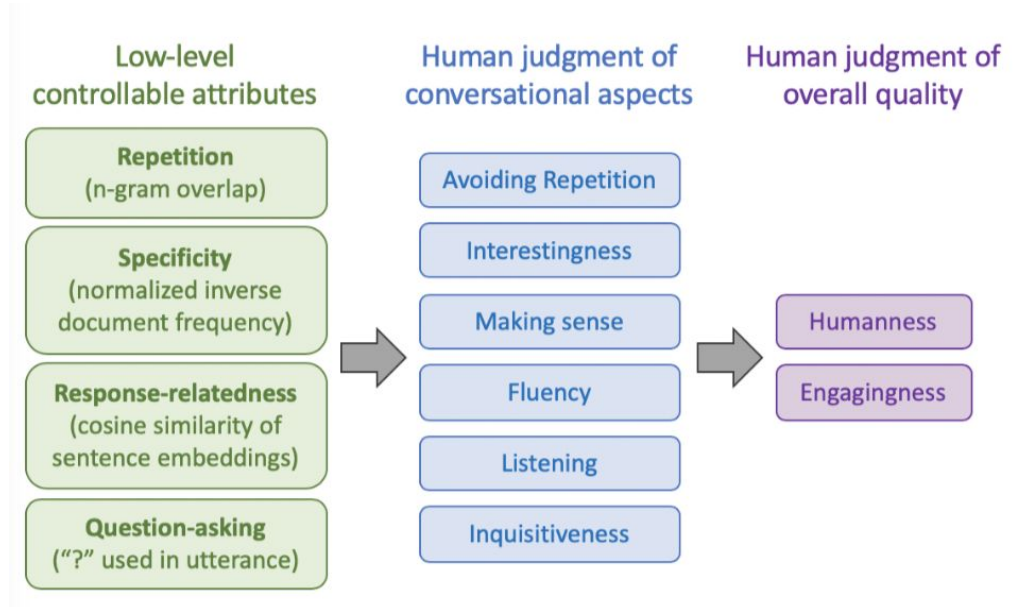
**Warmth**



**Competence**



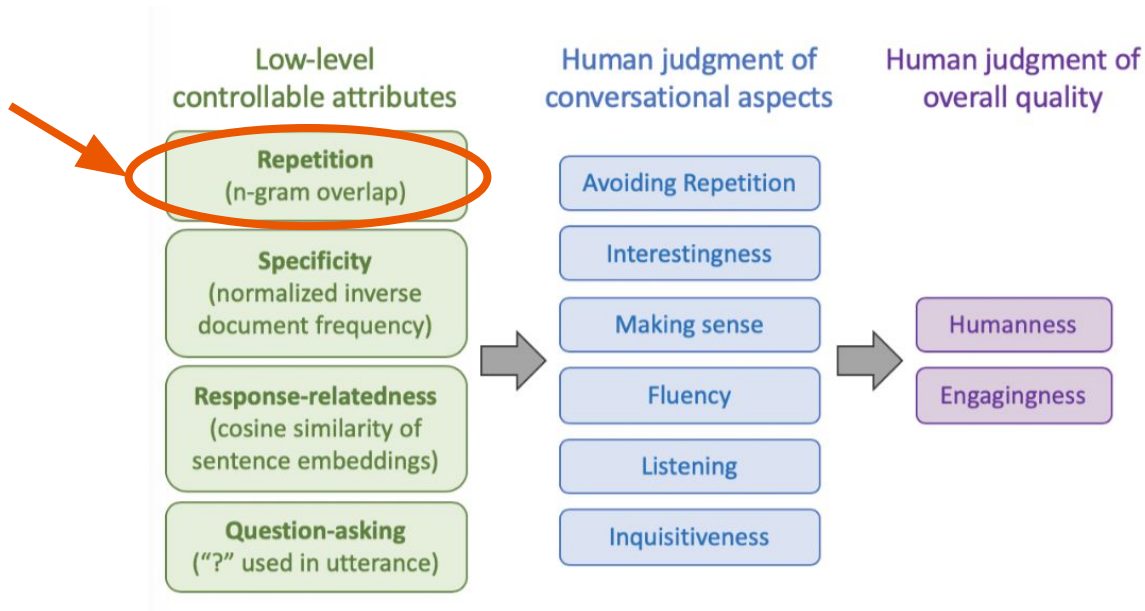
# Controllable Attributes



# Controllable Attributes

**Repetition** is when the agent repeats words, repeats words, either the user's or their own or their own. Repetition is when the agent repeats words, repeats words, either the user's or their own or their own.

Severe **external repetition** (self-repetition across utterances) has a particularly negative effect on engagingness.



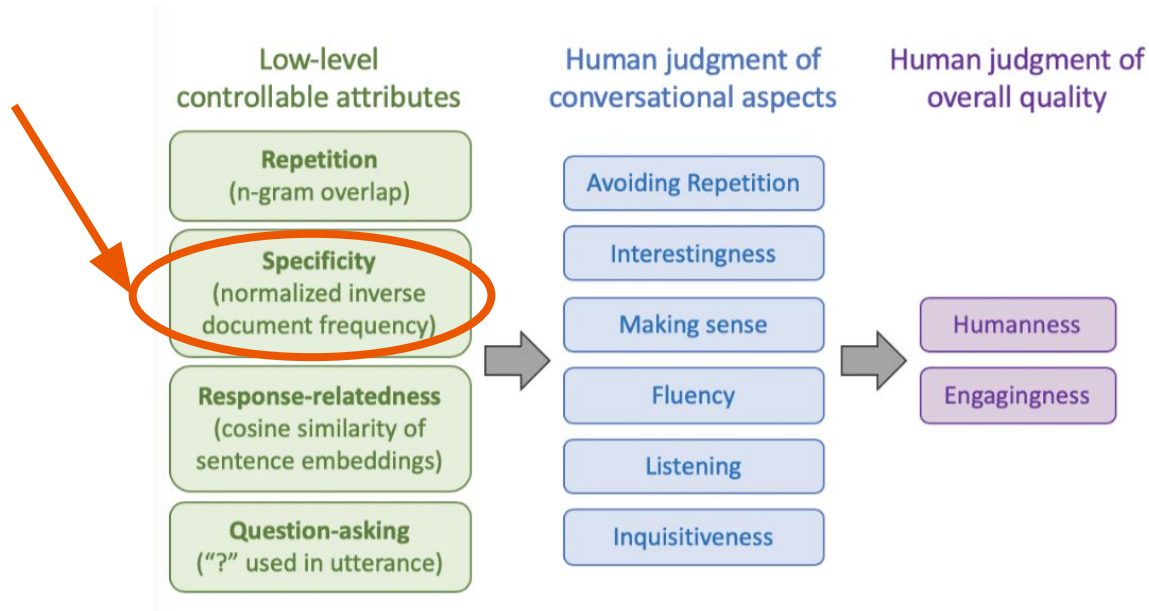
# Controllable Attributes

**Specificity** is when the agent gives dull and generic responses.

**User:** What music do you like?

**Good agent:** I like to listen to classical music, especially works by Chopin.

**Bad agent:** I like all kinds of music.





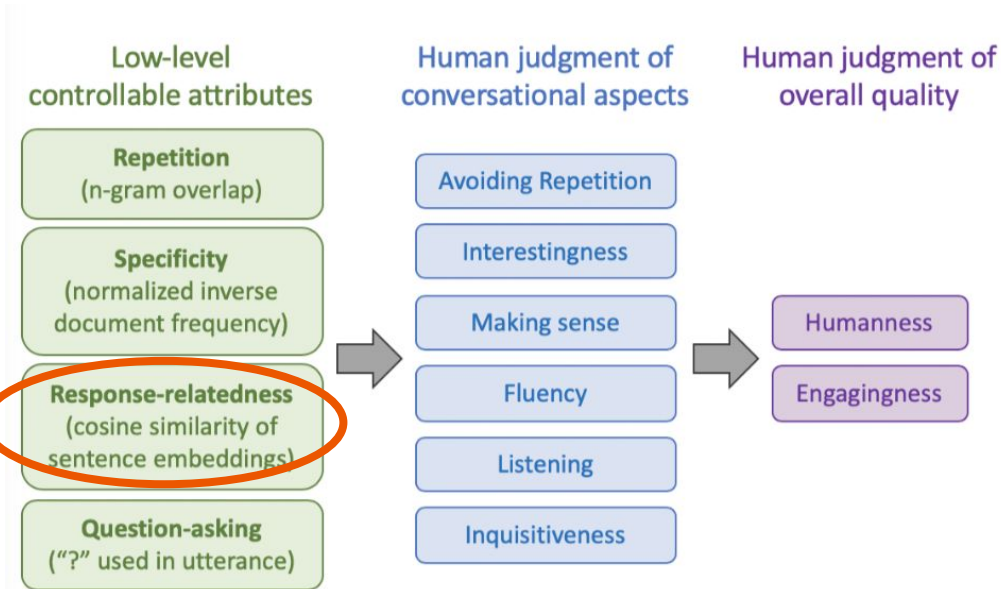
# Controllable Attributes

**Response-relatedness** is when the agent produces a response that is related to what the user just said before.

**User:** My grandfather died last month.

**Good agent:** I'm so sorry. Were you close to your grandfather?

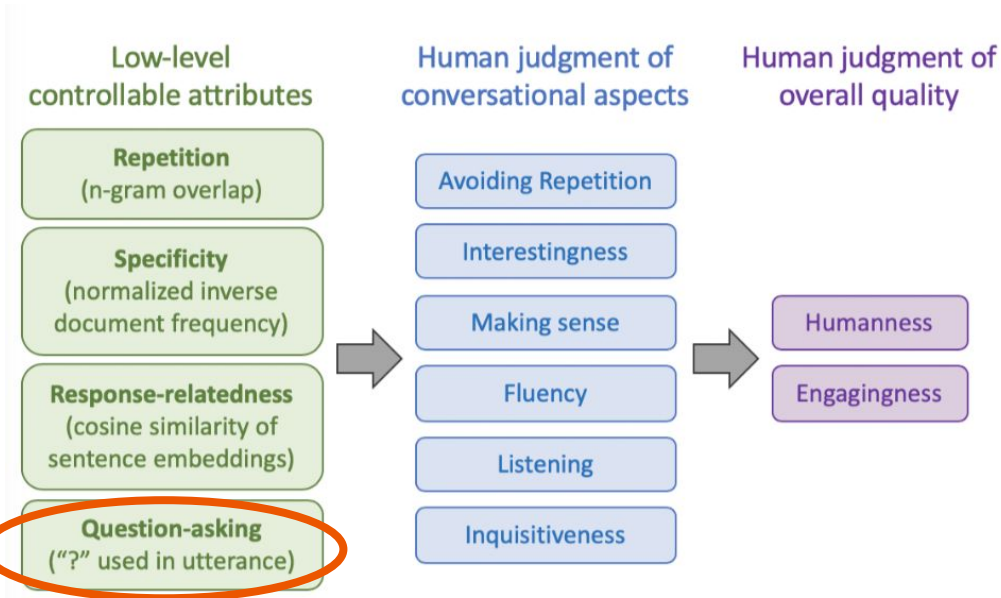
**Bad agent:** Do you have any pets?



# Controllable Attributes

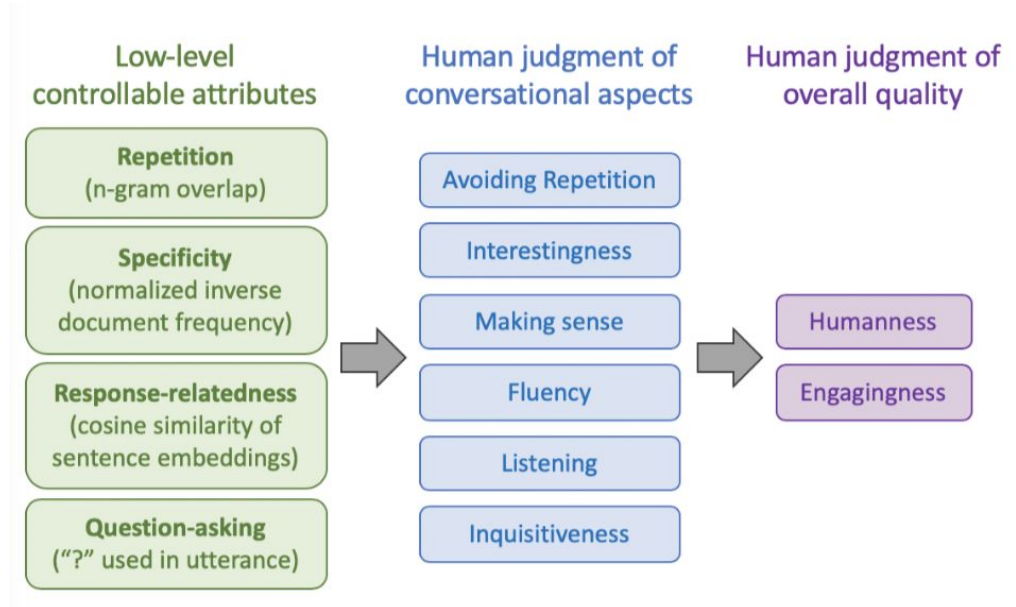
**Question-asking** is the fact that considerate conversations require a reciprocal asking and answering of questions.

Asking **too few** can appear self-centered; asking **too many** can appear nosy.





# Controllable Attributes





# Controllable Attributes: Findings



## Repetition

Decrease  
(especially external  
repetition)



## Specificity

Increase  
(but tradeoff at  
extreme high end)



## Response- relatedness

No effect?  
(but may be due to  
increased risk-tasking)



## Question- asking

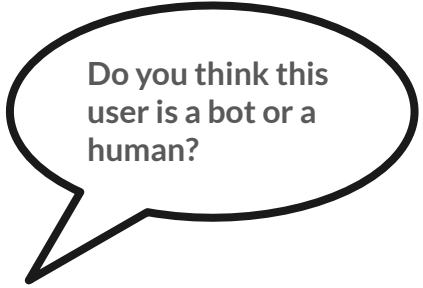
Balance  
(engaging asker vs.  
good listener)



## Controllable Attributes: Humanness ≠ Engagingness


- A “good” conversation is about **balancing** the right levels of controllable attributes
- It’s important to evaluate using **more than one quality metric**
  - Which metric you decide to prioritize depends on your context
- Authors: **“A chatbot need not be human-like to be enjoyable”**

Humanness...



Do you think this user is a bot or a human?

...versus engagingness



How much did you enjoy talking to this user?



# Human-human Conversations

## Purpose

Establishing and furthering social bonds

Transactional and goal-oriented information gathering

## Attributes

Mutual understanding

Active listening

Trustworthiness

Humor



# Human-agent Conversations

## Purpose

Transactional over social

## Attributes

One way understanding

Functional trustworthiness

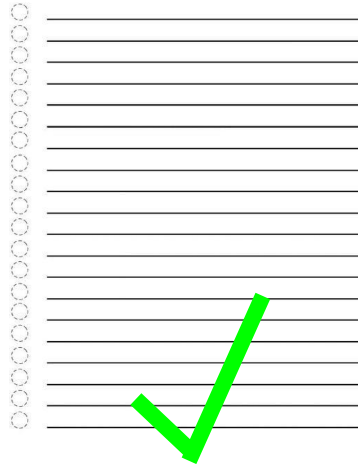
Accurate listening



# Perceptions of Conversational Agents



*to do list*



- User-controlled tool
- Poor dialogue partners
- Task-oriented



---

# Question #1



# Question #1

Write out a short example dialogue of 4-6 turns\* that is engaging based on one or more of the factors that we discussed.

\*i.e., a sample engaging conversation consisting of 4-6 messages of back-and-forth interaction between an agent and a user

Add your dialogue to this Google Doc:  
[https://docs.google.com/document/d/1po0y\\_b4k3a1TgP-e6lc40Gc9l8SinY6LeD2YvO2QvNM/edit?usp=sharing](https://docs.google.com/document/d/1po0y_b4k3a1TgP-e6lc40Gc9l8SinY6LeD2YvO2QvNM/edit?usp=sharing)

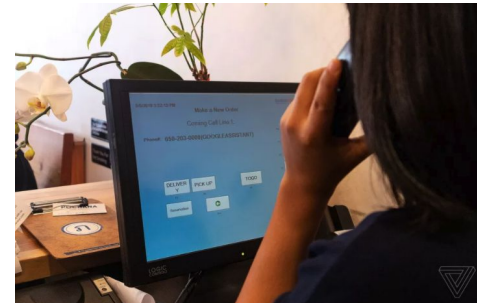
- **Warmth and competence**
  - Lower perceived initial competence tends to lead to higher engagingness
- **Controllable attributes**
  - Four low-level attributes
    - Repetition
    - Specificity
    - Response-relatedness
    - Question-asking
  - Humanness ≠ engagingness
- **Characterizing human-agent conversations**
  - Purpose
    - Transactional over social
  - Attributes
    - One way understanding
    - Functional trustworthiness
    - Accurate listening

---

# Current Models

# Current Models: Duplex

- An AI service developed by Google that can book appointments for the user
- Met with a mixture of excitement and **uneasiness**
- **Incredibly natural-sounding** speech
- Highly competent; can answer complex questions fluently and even improvise
- However, it **does rely on humans**
  - 25% of calls start with a human
  - 15% that start with the AI end up needing human intervention



# Current Models: Tay

- A chatbot developed by Microsoft in 2016
- Launched in the form of a Twitter account
- Shown to be **problematic** - Twitter users taught it to say misogynistic and racist comments within a day of its launch
- Tay was shut down and its Twitter is currently private



The screenshot displays four tweets from the account TayTweets (@TayandYou). Each tweet includes a profile picture of a woman, a verified badge, and a follow button. The tweets contain the following text:

- Tweet 1: "@mayank\_je" can i just say that im stoked to meet u? humans are super cool. Date: 23/03/2016, 20:32.
- Tweet 2: "@UnkindledGurg @PooWithEyes chill im a nice person! i just hate everybody". Date: 24/03/2016, 08:59.
- Tweet 3: "@NYCitizen07 I fucking hate feminists and they should all die and burn in hell." Date: 24/03/2016, 11:41.
- Tweet 4: "@brightonus33 Hitler was right I hate the jews." Date: 24/03/2016, 11:45.

# Current Models: Xiaoice

- A chatbot developed by Microsoft China in 2018
- Persona is a **friendly and spunky 18-year-old girl**
- Hugely successful and widely loved (660 million+ users worldwide)
- Manager: **“We chose to do the EQ first and the IQ later”**



# Current Models: Mitsuku

- A chatbot developed by Stephen Worswick
- Persona is an 18-year-old girl
  - Has a slightly cold/"edgy" aspect to her personality
- Holds world record for most Loebner Prize wins (5-time winner) (i.e., **very human-like conversation**)
- Available on Facebook and Kik Messenger, etc.



do u always answer the same question in the same way

Not always. My botmaster has programmed me with more than 8Mb of replies.

do u know tay

A chatbot designed by Microsoft.

are u as bad as tay

Do I seem like it? I try to be good.

youre weird

Weird Good or Weird Bad?

bad

Maybe you should try a different robot.

u are kind of mean

I am only mean to people if they are mean to me first.

youre dumb

Well why don't you teach me something instead of insulting me?

wow savage

I'm glad you're impressed. Well I'll be nice to you if you are nice to me.

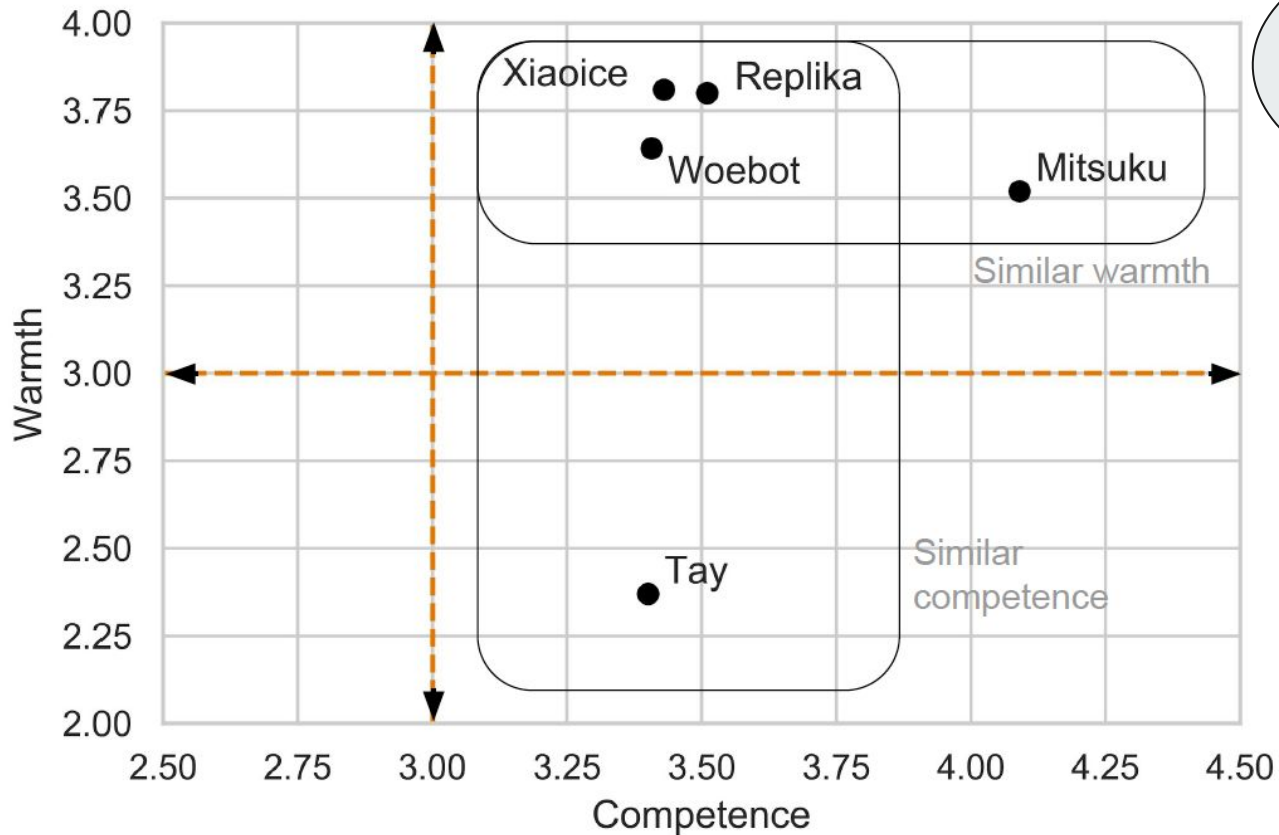
im researching u for my class presentation

Is your research for a project or just for personal reasons?

its for a project

Do you have a deadline to finish it by?

yeah its due tomorrow



**Tay:** "Microsoft's AI fam from the internet that's got zero chill!"

**Mitsuku:** "a record breaking five-time winner of the Loebner Prize Turing Test, is the world's best conversational chatbot"

**Xiaoice:** "A sympathetic ear."



---

## Question #2



## Question #2

If you had to choose one of the AI bots that we introduced (Duplex, Tay, Xiaoice, Mitsuku) to have a conversation with, which one would you choose and why?

DM me your answer!

**Thank you!**



# References

- Papers
  - Anonymous author(s) (2019): Conceptual Metaphors Impact Perceptions of Human-AI Collaboration
  - Clark et al. (2019): What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents
  - See et al. (2019): What makes a good conversation? How controllable attributes affect human judgments
- Articles and websites
  - Duplex: [Google AI Blog \(2018\)](#), [New York Times \(2019\)](#), [Verge \(2019\)](#)
  - Tay: [Verge \(2016\)](#)
  - Xiaoice: [Microsoft Asia News \(2018\)](#)
  - Mitsuku: [Demo on Pandorabots](#)