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)
Factors that Make a Conversation Engaging
Conceptual Metaphors
An understanding of abstract or complex ideas using simple terms

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

Tay: “AI that’s got no chill”
Xiaoice: “an empathetic ear”
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

(a) Pre-Use Usability

(b) Post-Use Usability
User Evaluation Results

Desire to co-operate

Intention to adopt
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

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

Hello, I am planning 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 those 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.
Takeaways

Warmth

Competence
Controllable Attributes

- **Low-level controllable attributes**
  - Repetition (n-gram overlap)
  - Specificity (normalized inverse document frequency)
  - Response-relatedness (cosine similarity of sentence embeddings)
  - Question-asking ("?" used in utterance)

- **Human judgment of conversational aspects**
  - Avoiding Repetition
  - Interestingness
  - Making sense
  - Fluency
  - Listening
  - Inquisitiveness

- **Human judgment of overall quality**
  - Humanness
  - Engagingness
Controllable Attributes

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

- **Low-level controllable attributes**
  - Repetition (n-gram overlap)
  - Specificity (normalized inverse document frequency)
  - Response-relatedness (cosine similarity of sentence embeddings)
  - Question-asking (“?” used in utterance)

- **Human judgment of conversational aspects**
  - Avoiding Repetition
  - Interestingness
  - Making sense
  - Fluency
  - Listening
  - Inquisitiveness

- **Human judgment of overall quality**
  - Humanness
  - Engagingness
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”

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

Humanness...

How much did you enjoy talking to this user?

...versus engagingness
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

- 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-e6Ic40Gc9l8SinY6LeD2YvO2QvNM/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
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.
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

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
  - See et al. (2019): What makes a good conversation? How controllable attributes affect human judgments

- Articles and websites
  - Tay: Verge (2016)
  - Xiaoice: Microsoft Asia News (2018)
  - Mitsuku: Demo on Pandorabots