

# Personality and Conversation Engagingness for Virtual Assistants

---

Emily Wu and Esther Kim

# Introduction

---



## Deliverables:

- Gold-standard dialogue consisting of several scenarios

## Methods:

- 3 surveys completed by workers on MTurk

# Methods

---

## Survey 1

- Varying competence/warmth

## Survey 2

- Testing different personas

## Survey 3

- Testing different types of dialogue along specific persona

# Results: Survey 1

---

- **Some results agreed with the conceptual metaphors study...**
  - High warmth/high competence and low warmth/low competence were perceived as expected
- **...and some did not**
  - Higher levels of warmth than expected perceived for low warmth/high competence
  - High warmth/low competence perceived as frustrating and not particularly friendly
- **Takeaways**
  - Users may have lower expectations for warmth for virtual assistants in the first place
  - Users strongly value high competence for task-oriented dialogue

# Results: Survey 2

---

- Friend persona was most positively received overall

	Average Usability Score (1-5)	Average Motivation Score (1-5)	Total Emotional Reception Score (+1/-1 per positive/negative emotion, respectively)
Mother	4.143	3.571	-2
Friend	4.095	3.857	30
Trainer	3.857	3.571	28
Reproachful	2.952	2.952	-25

*Figure 4. Average usability and motivation scores and total emotional reception scores for each of the four personas. The friend persona achieved the highest average motivation and total emotional reception scores. Although the mother persona was rated on average highest for usability, the friend was a close second and significantly outperformed the mother in all other categories.*

# Results: Survey 3 (example)

I had a bowl of brown rice with veggies and chicken for lunch, but I ended up eating four cheeseburgers for dinner...

Sometimes we mess up and that's okay! Let's work on getting ourselves back on track instead of dwelling on the negatives. Tomorrow is a new day!

I still feel awful about how much I ate for dinner today.

Don't let your negative feelings get to you! Each day is a chance to start new, and I know that you can bounce back!

**A**

I had a bowl of brown rice with veggies and chicken for lunch, but I ended up eating four cheeseburgers for dinner...

Oh no, that doesn't sound good... Don't lose hope though! An unhealthy meal can set you back, but if you're able to return to your healthy eating habits tomorrow, you'll still be on track.

I still feel awful about how much I ate for dinner today.

I'm sorry you feel that way. But, I think that the best thing that you can do right now is not to dwell on something that has already passed, but to see what actions you can take to recover from here. You can't change the fact that your dinner today was unhealthy, but what you can change is your future eating habits.

**B**

# Results: Survey 3

---

- **Upbeat vs. calm tone**
  - Split exactly 50-50, so we included both versions
  - Upbeat supporter: “I really felt like the virtual assistant cared.”
  - Calm supporter: “Seems more like a human. Less of a canned motivation speech like [the upbeat agent].”
- **Follow-up question vs. no follow-up**
  - 70% in favor of follow-up, so we included follow-up
  - Follow-up supporter: “I like that it asks a question about why this happened so it can give better advice.”

# Results: Survey 3

---

- **Ask before giving info vs. no ask**
  - 65% in favor of ask-first, so we included ask-first
  - Ask-first supporter: “I liked that there was more dialogue in [the ask-first agent]. I would feel more motivated with [the ask-first agent].”
  - No ask supporter: “[The no ask agent] gave me basically the same information, but was more concise.”
- **Does reaffirming the user’s success have a beneficial effect?**
  - 100% felt positive emotions (vs. neutral or negative) after seeing agent’s response
  - 100% felt more motivated to continue their diet (vs. less motivated) after seeing agent’s response
  - Unanimous yes, so we included reaffirmation



# Discussion

---

## Changes to final dialogue based on data from surveys

- Included both calm and upbeat tones
- Utilized follow-up question
- Replace initial negative-sounding language
- Utilized ask-first before delivering information
- Modified agent's utterance to be more concise and to contain a more specific and actionable suggestion

# Discussion

---

## Limitations

- Passive, third-person responses
- Pre-written dialogues
- Survey design

# Conclusion

---

- **Problem space**
  - Virtual assistant personalities
- **Results**
  - Final gold-standard dialogue for healthy eating agent, iterated based on user feedback from our three surveys
- **Future work**
  - Implement in a quantitative neural network-based model
  - Subsequent user testing with participants in a first-person perspective

# Final Gold-Standard Dialogue

[Click here for our final gold-standard dialogue](#)

**Thank you!**

# References

---

- [1] Abigail See, Stephen Roller, Douwe Kiela, and Jason Weston. 2019. What makes a good conversation? How controllable attributes affect human judgments. In Proceedings of the 2019 Conference of the NAACL: Human Language Technologies. ACL, 1702-1723.
- [2] Amy JC Cuddy, Susan T Fiske, and Peter Glick. 2008. Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in experimental social psychology* 40 (2008), 61–149.
- [3] ANONYMOUS AUTHOR(s). 2019. Conceptual Metaphors Impact Perceptions of Human-AI Collaboration . *J. ACM* 37, 4, Article 111 (August 2019), 24 pages. <https://doi.org/10.1145/1122445.1122456>
- [4] Eeva Raita and Antti Oulasvirta. 2011. Too good to be bad: Favorable product expectations boost subjective usability ratings. *Interacting with Computers* 23, 4 (2011), 363–371.
- [5] Ewa Luger and Abigail Sellen. 2016. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, 5286–5297.
- [6] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3290605.3300705>
- [7] Muzafer Sherif, Daniel Taub, and Carl I Hovland. 1958. Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of experimental psychology* 55, 2 (1958), 150.