Selina Meyer

Regensburg University
Germany

selina.meyer@ur.de
https://selinameyer.github.io

1 Research interests

My PhD focuses on conversational agents for behaviour change, with a focus on the feasibility of applying Large Language Models (LLMs) such as GPT-4 in this context.

1.1 Prompt Engineering and Conversational Framework Design for LLMs for Wellbeing

I designed a conversational framework based on Motivational Interviewing theory, a client-centered therapy approach that emphasizes open questions and reflections to help clients find their own reasons and strategies for change Miller and Rollnick (2002); Clifford and Curtis (2016). The framework classifies user turns with codes relevant for behaviour change Miller et al. (2003) and selects a counsellor behaviour to be prompted to GPT-4 based on the user utterance type. I will evaluate to what extent GPT can be controlled in this context using the prompt engineering techniques employed in the conversational framework, and what current restraints of such models are in the context of interactions that rely on deeply social behaviours such as empathy and conveying an understanding of underlying expressed emotions and thoughts. In this context, I am also exploring how to mitigate potential harms of using such a technology in the context of behaviour change. To achieve this, it is important to define how LLM-output in the context of empathy and therapist-client interactions can be evaluated. Thus, I am interested in evaluation metrics for NLG in niche contexts where no ground truth is available, such as Sharma et al. (2020); Welivita and Pu (2020). In the same vein, I want to explore methods of harm mitigation, for instance caused by unhelpful advice or reinforcing negative behaviours if misused (i.e. supporting weight loss for anorexic users).

1.2 Effects of LLM-driven Motivational Chatbot on Behaviour Change Motivation

For the remainder of my PhD, I will mainly focus on the effects a conversational agent using the created conversational framework has on motivation and readiness to change behaviour. To do this, I will run two user studies, the first utilizing situated work task situations, where participants are requested to imagine they want to pursue a specific behaviour change before conversing with the chatbot. In this user study, we will measure whether the framework created leads to higher therapeutic alignment, user engagement, and perceived empathy and competence than a LLM-based chatbot that does not use the framework. The measures we will employ are based on similar research by He et al. (2022). The text data collected in this study will be analysed with regard to the quality of conversation and potentially harmful LLM-outputs. I will also explore, to what extent user behaviour influences the quality of the conversation. For instance, I hypothesize, that conversations with shorter user utterances might be less successful, as they give the chatbot less to work with.

In the second study, we will then test the conversational agent on people who are actually interested in changing their own behaviour. In this study, we will also measure effects of the chatbot on self-efficacy, readiness to change, and goal reflection. Participants will fill out all three measures both before and after the interaction with the chatbot. An increase in readiness to change, self-efficacy, or goal reflection will be a sign of the success of the intervention and the feasibility of using the chatbot to increase motivation for behaviour change. In future work, these evaluations could be complemented by a longer term study which investigates effects on behaviour change success.

2 Spoken dialogue system (SDS) research

I believe, that ChatGPT has caused a paradigm shift in the field of conversational AI research. Not only has it led to new opportunities of research, it also put chatbots on the map for the general population. This leads to a wider understanding of conversational AI and SDS in the general population. However, this also has the potential of leading to the privatization of conversational AI and SDS research, as it becomes harder and harder for researchers to compete with the financial prerequisites and manpower in industry. On the other hand, it could also mean increased collaboration between industry and academic research.

Ethical design is also a challenge that becomes increasingly important in times of LLMs. The curation of less biased datasets for training, the mitigation of the environmental impact of LLMs, and the containment of low-paid, unethical labour employed by industry creators of such models all call for solutions, which leave a rich gap for ethical research in the context of large models for con-

Proceedings of the 19th Annual Meeting of the Young Reseachers' Roundtable on Spoken Dialogue Systems, pages 42–43
September 11–12, 2023.

versational AI.

3 Suggested topics for discussion

Here, authors will suggest three topics for discussion in the discussion panels during the event. As an example, here are some of the discussion topics discussed in previous workshops:

- Evaluation of LLM-outputs when no ground truth/gold data is available
- Controllability of LLM-based text generation
- How can researchers compete with industry considering the difference in funding and manpower?

It is recommended to suggest topics, on which the author has knowledge, but also topics that they find interesting and relevant to the young community.

References

Dawn Clifford and Laura Curtis. 2016. *Motivational interviewing in nutrition and fitness*. Guilford Publications.

Linwei He, Erkan Basar, Reinout W Wiers, Marjolijn L Antheunis, and Emiel Krahmer. 2022. Can chatbots help to motivate smoking cessation? a study on the effectiveness of motivational interviewing on engagement and therapeutic alliance. *BMC Public Health* 22(1):726.

William R Miller, Theresa B Moyers, Denise Ernst, and Paul Amrhein. 2003. Manual for the motivational interviewing skill code (misc). *Unpublished manuscript. Albuquerque: Center on Alcoholism, Substance Abuse and Addictions, University of New Mexico* .

William R Miller and Stephen Rollnick. 2002. *Motivational Interviewing, Second Edition: Preparing People for Change*. Applications of Motivational Interviewing Series. Guilford Publications.

Ashish Sharma, Adam Miner, David Atkins, and Tim Althoff. 2020. A computational approach to understanding empathy expressed in text-based mental health support. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Online, pages 5263–5276. https://doi.org/10.18653/v1/2020.emnlp-main.425.

Anuradha Welivita and Pearl Pu. 2020. A taxonomy of empathetic response intents in human social conversations. In *Proceedings of the 28th International Conference on Computational Linguistics*. International Committee on Computational Linguistics, Barcelona, Spain (Online), pages 4886–4899. https://doi.org/10.18653/v1/2020.coling-main.429.

Biographical sketch

Selina Meyer is a third-year PhD student and research assistant at the Chair for Information Science of Regensburg University, Germany. With a background in the humanities, she is primarily interested in behavioural and user-centered aspects of computer science and its application in social sciences.