Dialect, sentiment analysis, tonality, cognitive modeling – using LLMs as a toolbox for HRI*

Thomas Sievers and Nele Russwinkel

***Abstract*— Large language models (LLMs) offer previously unimagined possibilities, especially for social robotics and human-robot interaction (HRI). In addition to the use of LLMs such as ChatGPT to enable a robot to engage in dialog on all conceivable topics, these possibilities also include the investigation of different special topics. Appropriate prompting for the LLM can easily influence the personality of the robot as perceived by a human interaction partner. We give examples of investigations into the effect of robot utterances with different tonality and the use of dialect in comparison to standard speech. It is also possible to assess the emotional content of a dialog using an LLM and to interact with cognitive architectures in the speech output. We present our studies and ideas on these topics, give an outlook on current experiments analyzing speech pauses and the combination of an LLM and a cognitive model and discuss our conclusions.**

I. INTRODUCTION

The personality of a social robot is crucial for effective and successful human-robot interaction. This robot personality manifests itself in a persona, which can have different characteristics depending on the intended use. When developing the prerequisites for successful interaction, it is usually a great challenge to understand and take into account the motivations and needs of the respective target group. We understand the persona of a robot as a fictitious personality with various stable behavior and personality patterns [1]. The question arises how the tonality of the robot's utterances influences the perceived robot personality and how this influences the user's further expectations.

Regardless of the robot's personality, it is desirable to take ethnic and linguistic diversity into account, including regional languages or dialects, in order to increase acceptance by the human counterpart. Differences in accent, grammar and vocabulary play a key role in developing a robot that is able to converse in natural language in socially and ethnically heterogeneous areas. Research on the expectations of conversational interaction with a robot confirms the importance of accent and dialect [2].

How can robots develop social and cultural awareness and adapt appropriately to their environment and the people they interact with? As social robots become more prevalent in different societies, they are confronted with a greater variety of social contexts. They not only need to be aware of and respond to social norms, but also deal with human emotions and intentions. Engwall et al. provide an overview of relevant studies regarding socially, culturally and contextually aware robots [3].

The analysis of feelings and emotions is about recognizing the emotional tonality of a particular text. A sentiment analysis extracts semantic information from these texts in order to understand the author's attitude and recognize positive, negative or neutral feelings. Research into ChatGPT's emotional awareness (EA) capabilities has shown that ChatGPT is able to generate appropriate EA responses [4]. EA is the ability to conceptualize one's own and others' emotions.

Cognitive architectures refer both to a theory about the structure of the human mind and to a computational realization of such a theory. Their formalized models can be used to flexibly react to actions of the human collaboration partner and to develop situation understanding for adequate reactions. A well-known and successfully used cognitive architecture is ACT-R (Adaptive Control of Thought - Rational) [5]. An example for the use of ACT-R with a social robot in a dialog scenario is given by Werk et. al. [6]. We are currently working on an idea for using an ACT-R model to control and manage the dialog parts of the robot generated by ChatGPT. Another topic of our current research is the use of ChatGPT for analyzing pauses in speech.

In the following, we briefly explain the studies we have conducted on this topics with social robots, provide an insight into ongoing work and summarize the results and findings. The publications cited in the references for these studies contain all further details.

II. METHODOLOGY

For our studies with large language models (LLMs) in conjunction with social robots having a dialog with a human interlocutor, we used OpenAI's Generative Pre-trained Transformer (GPT, commonly known as ChatGPT) [7] to create the conversational parts for the robot. The LLM was instructed via various system prompts to adapt the type and tone of the statements in accordance with the study objective.

To measure possible effects the different types of robot talk may have on a human interlocutor's judgement of social attributes of the robot we used the 18-item Robotic Social Attributes Scale (RoSAS) [8]. It comprises three underlying scale dimensions – the factors *warmth*, *competence* and *discomfort*. RoSAS is based on items from the Godspeed Scale [9] and psychological literature on social perception. We applied a within-subject design for the studies listed here.

Our study on the effects of using a regional language was conducted using the Pepper robot's internal response generation capabilities rather than using ChatGPT [10].

*This work was not supported by any organization

Thomas Sievers and Nele Russwinkel are with the Institute of Information Systems, University of Lübeck, 23562 Lübeck, Germany
sievers@uni-luebeck.de,
russwinkel@uni-luebeck.de

A. Humanoid Social Robot Pepper

The humanoid social robot Pepper we used in our studies was developed by Aldebaran and first released in 2015 [11]. The robot is able to engage with people through conversation, gestures and its touch screen. Our robot application forwards utterances of the human dialog partner as input to the OpenAI API, which returns a dictionary with the response of the API. That enables complex dialogs between humans and machines.

III. RESULTS

A. Using regional language in a Social Robot

We were interested in the differences in perception between a conversation in High German and a Low German version. Low German is a regional language that is understood and partly still spoken in the northern parts of Germany and is closely related to Frisian and English. We assumed that a robot speaking Low German in an area with a corresponding cultural identity should make a difference to how it is perceived by human conversation partners. In fact, it turned out that the participants perceived significantly more warmth in the Low German version of the conversation.

Later, we tried generating Low German text for the robot's voice output directly from ChatGPT. To do this, we retrained the GPT-3.5 model with Low German text using the fine-tuning options of OpenAI. As it turned out, this was hardly necessary, as ChatGPT is already generally able to generate texts in Low German via a corresponding prompt statement. The result is not always 100% correct, but good enough to achieve a Low German impression for the dialog partner.

B. Making Human Emotions tangible for a Social Robot

How can the emotional state of a person in a dialog with a robot be accessed? We investigated this question by evaluating the course of the dialog once by the human interlocutor and once by the GPT model itself using sentiment analysis [12]. In addition, the predominant emotion was named by both interlocutors. A comparison of these evaluations made it possible to assess whether the human and the social robot came to the same conclusions. We also investigated whether the transmission of emotion recognition data, which the LLM was supposed to take into account, had a noticeable influence on the tonality of the conversation.

It was shown that the predominant emotion of the human and the general mood of a conversation were interpreted largely the same by humans and ChatGPT, whereby an existing emotion recognition caused the robot's assessment of the general mood to differ noticeably. ChatGPT appears to be somewhat more willing to take risks here, instead of being 'cautious' and 'reserved' and therefore leaning more towards a neutral assessment.

C. Dialogs containing Irony, Sarcasm and Jocularity

One element you hardly expect from robots in conversation is humor, especially the use of irony and sarcasm. We examined the effect of a conversation with cheerfully ironic or sullenly sarcastic remarks by the robot in contrast to utterances in a more neutral tone [13]. In addition, we asked the participants which dialog style they personally liked best.

Our results showed that the perceived warmth gained highest rates when when the robot had a cheerfully ironic tone. Discomfort was rated significantly higher when a sullenly sarcastic tone was used by the robot. Perceived competence seemed to be slightly negatively influenced in the sarcastic condition. The results proved that tonality appears to be a very relevant design element for a successful interaction between humans and robots, but must be used carefully to achieve the desired goal.

D. Outlook: Speech Pauses

People pause when speaking for various reasons – often in the middle of a sentence. For example, pauses give the speaker a chance to continue thinking and the audience a chance to think at the same time. On the other hand, pauses in an utterance also reveal something about the emotional state of the speaker.

In a first setting, we record the human utterances with OpenAI Whisper and measure the pauses between the individual words. These intervals are delivered to ChatGPT in addition to the spoken text and evaluated by the LLM with regard to the mood and emotional state of the speaker. This works well in principle. One difficulty is distinguishing between pauses in a sentence that has yet to be completed and the actual end of a conversation turn. A multimodal approach, for example by interpreting facial expressions, could help to differentiate between end-of-sentence, turn-taking and confusion pauses.

E. Outlook: LLMs + Cognitive Models

Since ChatGPT does not have a dynamic cognitive model of the human interlocutor and therefore quickly reaches its limits in cognitive processes, we try to subordinate the utterances of the LLM to the control of a cognitive architecture. To do this, we use an ACT-R model that follows the course of the conversation and intervenes to control it if necessary.

We have only just started working on this. However, since it is possible to integrate an ACT-R model with bi-directional communication into a robot application and thus also to an LLM, it should be possible to control the LLM output via prompts influenced by the cognitive model [14].

IV. CONCLUSIONS

LLMs like ChatGPT provide powerful tools for pretty much everything that involves the use of natural language in HRI. With an appropriate prompt design, it is easy to define a desired tonality with a target group-specific approach and the desired language – including regional dialects. ChatGPT is also able to make assessments, e.g. with regard to the emotional content of a conversation, and to incorporate these findings into the robot's utterances (content and tonality) and behavior (gestures).

Further application scenarios are conceivable in which LLMs - in addition to pure text generation for speech output - together with additional modules and functionalities such as cognitive models create higher-value functionalities in the direction of human-like behavior of social robots.

REFERENCES

[1] G. Matthews, I. J. Deary, and M. C. Whiteman, "Personality Traits," 3rd ed. Cambridge: Cambridge University Press, 2009.

[2] M. E. Foster and J. Stuart-Smith, "Social Robotics meets Sociolinguistics: Investigating Accent Bias and Social Context in HRI", Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23), Association for Computing Machinery, New York, NY, USA, pp. 156–160, doi: 10.1145/3568294.3580063, 2023.

[3] O. Engwall, R. J. P. Bandera, S. Bensch, K. S. Haring, T. Kanda, P. Núñez, M. Rehm, and A. Sgorbissa, "Editorial: Socially, culturally and contextually aware robots ," Frontiers in Robotics and AI, vol. 10, doi: 10.3389/frobt.2023.1232215, 2023.

[4] Z. Elyoseph, D. Hadar-Shova, K. Asraf, M. Subbiah, and M. Lvovsky, "ChatGPT outperforms humans in emotional awareness evaluations", Frontiers in Psychology, vol. 14, doi:10.3389/fpsyg.2023.1199058, 2023.

[5] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of the mind", Psychological review, vol. 111, 4, pp: 1036–1060, doi: 10.1037/0033-295X.111.4.1036, 2004.

[6] A. Werk, S. Scholz, T. Sievers, and N. Russwinkel, "How to Provide a Dynamic Cognitive Person Model of a Human Collaboration Partner to a Pepper Robot", Society for Mathematical Psychology, ICCM, 2024. Via mathpsych.org/presentation/1452.

[7] OpenAI. (2023) Transforming work and creativity with AI [Online]. Available: https://openai.com/product

[8] C. M. Carpinella, A. B. Wyman, M. A. Perez and S. J. Stroessner, "The Robotic Social Attributes Scale (RoSAS): Development and Validation," 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Vienna, Austria, pp. 254–262, 2017.

[9] C. Bartneck, D. Kulić and S. Zoghbi, "Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots", International Journal of Social Robotics 1, 71–81, doi: 10.1007/s12369-008-0001-3, 2009

[10] T. Sievers, and N. Russwinkel, "Talking Like One of Us: Effects of Using Regional Language in a Humanoid Social Robot," Social Robotics (ICSR 2023), Lecture Notes in Computer Science, vol. 14454, Springer, Singapore, doi: org/10.1007/978-981-99-8718-4_7, 2024.

[11] Aldebaran, United Robotics Group and Softbank Robotics. (2024) Pepper [Online]. Available: https://www.aldebaran.com/en/pepper

[12] T. Sievers, and N. Russwinkel, "Interacting with a Sentimental Robot – Making Human Emotions tangible for a Social Robot via ChatGPT," Workshop on Advanced Robotics and its Social Impacts (ARSO), IEEE, doi: 10.1109/ARSO60199.2024.10557749, 2024.

[13] T. Sievers, and N. Russwinkel, "Introducing a note of levity to human-robot interaction with dialogs containing irony, sarcasm and jocularity," International Conference on Robot and Human Interactive Communication (RO-MAN), IEEE, 2024 forthcoming.

[14] T. Sievers, and N. Russwinkel, "How to use a cognitive architecture for a dynamic person model with a social robot in human collaboration," Robots for Humans, CEUR Workshop Proceedings, 2024 forthcoming.