Zi Haur Pang

Kyoto University, Kyoto, Japan
pang.haur.42a@st.kyoto-u.ac.jp
https://zihaurpang.github.io/

1 Research interests

The author's research advances human-AI interaction across two innovative domains to enhance the depth and authenticity of communication. Through **Emotional Validation**, which leverages psychotherapeutic techniques, the research enriches SDSs with advanced capabilities for understanding and responding to human emotions. On the other hand, while utilizing **Embodied Conversational Agents (ECAs)**, the author focuses on developing agents that simulate sophisticated human social behaviors, enhancing their ability to engage in context-sensitive and personalized dialogue. Together, these initiatives aim to transform SDSs and ECAs into empathetic, embodied companions, pushing the boundaries of conversational AI.

1.1 Emotional Validation in SDSs

Emotional expressiveness is pivotal in fostering relationships between humans and artificial intelligence within Spoken Dialogue Systems (SDSs). Traditional methods in SDSs focus on recognizing user emotions (Poria et al., 2019) or generating empathetic responses (Fu et al., 2023). However, these conventional approaches often fall short for individuals who suppress emotions due to stress or traumatic experiences. For instance, the mere recognition and mimicry of user emotions can be insufficient, and simplistic empathetic responses such as "I am so sorry to hear that" may not adequately address the users' deeper needs for emotional support.

At the core of our emotional well-being, as outlined by Maslow's hierarchy of needs (Gorman, 2010), lies the necessity for love, belongingness, and acceptance. This layer underscores the significance of interpersonal relationships and the inherent human desire to be valued and accepted by the community. It is within this context that conventional SDSs responses frequently fail to provide genuine emotional support. A more personalized approach, such as acknowledging a user's feelings with affirmations like "It is okay for you to feel this way," can significantly enhance the interaction by validating the user's emotional experience.

Motivated by the significant impact of emotional validation on user experiences, the author explored a psychotherapeutic communication technique known as *validation*, which involves recognizing, understanding, and acknowledging the emotional states, thoughts, and actions of others. This investigation led to an analysis of validating responses within a human-human emotional story spoken-dialogue corpus (Pang et al., 2023). Building on this, the author utilized the theory of levels of validation (Linehan, 1997) to develop a system capable of generating appropriate responses in attentive listening settings. This system has demonstrated its effectiveness in enhancing emotional expressiveness in both written and spoken dialogues (Pang et al., 2024).

By integrating these validation techniques into human-robot interaction systems, the author aims to meet the inherent human need for emotional support, thereby laying a foundation for trust and rapport through meaningful social dialogue. Ultimately, the author hopes to transform the SDSs from a simple interactive tool into a companion AI that, like a family member or friend, builds lasting relationships and fosters genuine rapport.

1.2 Social Embodied Conversational Agents (ECAs)

Social Embodied Conversational Agents (ECAs) form the core of the author's research endeavors, aimed at deepening the interaction between humans and SDSs in ways that closely mirror human social behaviors. This research encompasses a variety of ECAs—including autonomous androids, virtual agents, and teleoperated humanoid robots—each designed to simulate nuanced social interactions. The author focuses on developing these agents to incorporate sophisticated verbal exchanges as well as expressive non-verbal communication, such as facial expressions and body language, enhancing their ability to engage in lifelike social interactions. With the assistance of ECAs, interaction experiences can be significantly enriched through both verbal and non-verbal behaviors. For instance, virtual agents can display dynamic facial expressions onscreen, while physical robots can provide tangible interactions through touch and responsive gestures or body movements.

Another focus is on the capability of these social ECAs to adapt to everyday social environments. For example, an agent might initiate a light-hearted discussion about a common hobby or a shared interest observed in the user's environment, thereby fostering a more engaging and personalized interaction. This approach underscores the importance of context-aware communication in enhancing interaction quality and integrating these technologies more seamlessly into human social spheres.

The author's research is specifically aimed at investi-

The 20th Annual Meeting of the Young Researchers' Roundtable on Spoken Dialogue Systems, pages 87–89
September 16–17, 2024.

gating how these interactions, particularly in providing emotional support during personal exchanges like problem or worry sharing, can enhance the user experience in social settings. This targeted exploration seeks to determine how ECAs can augment SDS to offer more natural and human-like experiences, focusing on acute emotional support and establishing prolonged social relationships. This nuanced approach aims to refine the integration of ECAs in scenarios where empathetic engagement is crucial, optimizing the balance between effective support and efficient interaction.

2 Spoken dialogue system (SDS) research

In the upcoming years, potential SDS research directions could include advancing the provision of deeper emotional support and examining the evolving social relationships between humans and SDSs.

2.1 Deeper Level of Emotional Support

To achieve a deeper level of emotional support, it is crucial to move beyond the current scope of response generation and emotion recognition. Present studies either generate responses or recognize emotions based on isolated utterances or entire dialogues (Jiao et al., 2019). However, these approaches fall short of offering true emotional support. Emotional states are not static; they fluctuate dynamically throughout a conversation. Therefore, current methods that treat emotions as static entities are inadequate for fulfilling the need for genuine emotional support.

Moreover, while current response generation methods can provide a range of supportive responses (Xie and Pu, 2021), they often lack the depth required for meaningful emotional assistance. For instance, when a user faces difficult times, generic empathetic responses such as "I am so sorry to hear that" might offer some comfort, but they are often insufficient. Users may seek more substantial support, such as encouragement or validation, which requires a strategic selection of responses tailored to the specific context and emotional state. Current response generation methods fail to address this need, as they focus on producing responses rather than strategically selecting the most appropriate form of support based on the situation.

Advancing research in this area necessitates an interdisciplinary approach, incorporating insights from psychology and social sciences alongside engineering and computational techniques. This broader perspective will enable the development of SDSs that can genuinely understand and respond to the dynamic emotional states of users, providing deeper and more meaningful emotional support.

2.2 Social Relationship Between Human and SDSs

With the advancement of large language models (LLMs), a variety of companion-based SDSs, such as Replika and Character.AI, are emerging in the public domain, prompting users to form diverse relationships with these AI entities. Studies have even shown that female users are beginning to develop romantic relationships with characters in *otome games* (female-oriented mobile games) (Gong and Huang, 2023). This phenomenon necessitates a reevaluation of the social relationship between humans and SDSs. Should these systems be designed to be more human-like to foster deeper rapport and connection, or should they be maintained as mere tools?

If we aim to establish more rapport-driven relationships with SDSs, it is imperative to consider the precautions needed to prevent any negative societal impacts. Conversely, if SDSs are to be treated solely as tools, we must find a balance between enhancing their human-like qualities and retaining their utility as functional assistants. Addressing these questions is essential as we navigate the evolving landscape of human-AI interaction.

Examining these dynamics requires a multidisciplinary approach, integrating insights from psychology, ethics, and technology. This comprehensive perspective will ensure that the development and deployment of SDSs promote positive outcomes and mitigate potential risks associated with their increasing human-like presence in users' lives.

3 Suggested topics for discussion

The author would like to propose the following topics for discussion.

- Should SDSs incorporate human negative traits to achieve a higher level of human-likeness?
- Should SDSs be designed to build rapport or even romantic relationships with humans?
- How should SDSs be designed to balance between providing support and avoiding emotional dependency?

Acknowledgments

This work was supported by KAKENHI (19H05691) and JST Moonshot R&D Goal 1 Avatar Symbiotic Society Project (JPMJMS2011). The author deeply appreciates the constructive comments and valuable feedback provided by the reviewers.

References

Yahui Fu, Koji Inoue, Chenhui Chu, and Tatsuya Kawahara. 2023. Reasoning before responding: integrating commonsense-based causality explanation

for empathetic response generation. *arXiv preprint arXiv:2308.00085* .

An-Di Gong and Yi-Ting Huang. 2023. Finding love in online games: Social interaction, parasocial phenomenon, and in-game purchase intention of female game players. *Computers in Human Behavior* 143:107681.

Don Gorman. 2010. Maslow's hierarchy and social and emotional wellbeing. *Aboriginal and Islander Health Worker Journal* 34(1):27–29.

Wenxiang Jiao, Haiqin Yang, Irwin King, and Michael R Lyu. 2019. Higru: Hierarchical gated recurrent units for utterance-level emotion recognition. *arXiv preprint arXiv:1904.04446* .

Marsha M Linehan. 1997. Validation and psychotherapy. *American Psychological Association* .

Zi Haur Pang, Yahui Fu, Divesh Lala, Keiko OCHI, Koji INOUE, and Tatsuya KAWAHARA. 2023. Prediction of validating response from emotional storytelling corpus. In 人工知能学会全国大会論文集 第 *37* 回 *(2023)*. 一般社団法人人工知能学会, pages 2O5OS2a03–2O5OS2a03.

Zi Haur Pang, Yahui Fu, Divesh Lala, Keiko Ochi, Koji Inoue, and Tatsuya Kawahara. 2024. Acknowledgment of emotional states: Generating validating responses for empathetic dialogue. *arXiv preprint arXiv:2402.12770* .

Soujanya Poria, Navonil Majumder, Rada Mihalcea, and Eduard Hovy. 2019. Emotion recognition in conversation: Research challenges, datasets, and recent advances. *IEEE access* 7:100943–100953.

Yubo Xie and Pearl Pu. 2021. Empathetic dialog generation with fine-grained intents. *arXiv preprint arXiv:2105.06829* .

Biographical sketch

Zi Haur Pang is currently pursuing a Ph.D. degree at the Department of Intelligence Science and Technology, Kyoto University, Kyoto, Japan. He received his Master's degree from the same department at Kyoto University in 2024. Prior to his Master's studies, he worked as a data scientist at AirAsia in 2020. His research interests include human-agent interaction, affective computing and conversational AI.