Designing an Affective Interface for a Personal Digital Assistant

Alok Debnath
ADAPT Centre
School of Computer Science and Statistics,
Trinity College Dublin,
Dublin, Ireland
debnatha@tcd.ie

***Abstract*—Integrating emotional understanding and empathetic response have become important for improving the naturalness of general purpose dialogue agents. With the aim of using dialogue systems as a way to interface with a personal digital assistant (PDA), the focus on empathy evolves over time, with the agent expected to maintain a memory of the user's state and their general emotion on a range of topics from previous interactions. Therefore, the goal of dialogue agents for PDAs is to be both empathetic and proactive in their use of prior information about the user, i.e. interact with compassion. This paper aims to lay the foundations, detail the key research questions, and present hypotheses aimed towards the design and architecture of an affective interface for a compassionate PDA.**

***Index Terms*—computational empathy, virtual assistants, dialogue systems**

I. Introduction

Interactive systems such as conversational agents, social robots, and personal digital assistants (PDAs)[1] aim to emulate human-human interaction in a given context [1]. Since humans are naturally empathetic in their interaction, recent developments in improving the naturalness of human-robot interaction have focused on improving the agent's ability to pick up on and understand emotional cues from a given user. Unimodal interactive agents such as empathetic dialogue models would be designed and trained the same way as their "non-empathetic" counterpart, only with some additional changes to the input representation to account for implicit emotion information [17].

For episodic interactions, as with a social chatbot, such a model works well, as different users can have a relatively similar exchange with the same agent. However, for models that are intended for a single user, such as PDAs, the design of an empathetic agent can not be limited to the analysis of emotional cues episodically. Rather, empathetic PDAs are expected to learn about their user over time, and incorporate what they have learned to provide more suitable, emotion-aware responses to an given user. Two examples of an expected interaction with an empathetic agent are given in Figure 1. Not only is the agent able to understand and respond appropriately to the user input, it also showcases a proactivity in response (as in case (i)) and an updated understanding of the user (as in case (ii)). We posit that such an interaction is empathetic, in that it understands the user, and compassionate, in that the agent shows concern about the user's emotional state and interests.

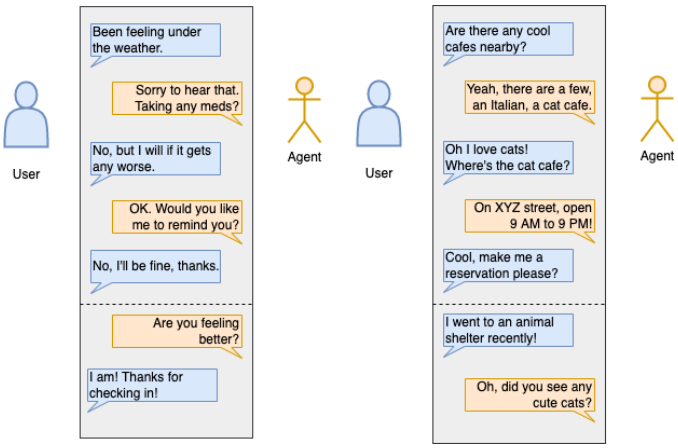

Fig. 1. A simple expectation from a personal assistant with an affective interface. (i) the agent understands the user's current state and at a later point asks if the user is feeling better, and (ii) the agent remembers affinity to cats and brings it up in a future conversation.

In order to elucidate the difference between an empathetic conversational agent and a compassionate PDA, we first establish the working definitions of empathy and compassion. As mentioned before, empathy has been discussed in the context of interactive agents in the past [22], [36]. While empathy is a complex behavioural and cognitive phenomenon, the working definition that has been established in literature states: *"empathy is the ability of a being to perceive, understand, and act appropriately to the emotions of another* [21]. In literature, empathy is seen as a tripartite process: affective (the ability to *perceive* emotional state or distress), cognitive (the ability to *understand* or feel that emotional state), and compassionate empathy (the ability to *act appropriately* towards the emotion portrayed) [6]. Modeling empathy for interactive agents has established a link between the affective and cognitive empathy,

This research was conducted with the financial support of Science Foundation Ireland under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at Trinity College Dublin. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology, is funded by Science Foundation Ireland through the SFI Research Centres Programme.

[1]Also referred to as personal assistants, intelligent personal assistants (IPAs), and intelligent virtual agents (IVAs) [1]

978-1-6654-5490-2/22/$31.00 ©2022 European Union

which in the context of text-based digital assistants, is realized as a cycle of emotion recognition and response generation [37]. However, very little work has happened in the role of proactive empathetic engagement in a user-agent scenario, therefore not establishing a computational framework for the realization of compassionate agent response.

Therefore, in this paper, we ask the question: **How do we design a personal digital assistant which interacts with empathy and compassion?** In order to develop a methodology to answer this question, the following prerequisites are necessary:

1) *How do we evaluate empathy in dialogue systems?*: Empathetic models in the current state-of-the-art are trained on a few corpora, such as EMPATHETICDIALOGUES [30] and PersonaChat [18], and are evaluated using a combination of automated and manual methods. However, since individual models use only a subset of the evaluation criteria, it becomes difficult to establish how empathetic the model is. Therefore, the first step towards developing a more empathetic and compassionate model of dialogue would be establishing an evaluation framework with both automated and human evaluation components.
2) *How do we improve the agent's empathetic understanding?*: With an evaluation metric in place, the next step would be to develop a model of affect beyond categorical emotion. Other implicit affective information such as intent, sentiment, stance, motivation, and emotional need can be identified from a user input. We hypothesize that the analysis of these affective states can supplement the agent's understanding of the user input and can not only help with generating more empathetic responses, but provide a broader range of heuristics to update the user model for future conversations.
3) *How do we incorporate compassion?* While empathy has been measured by the appropriateness and emotion in the agent's *response*, compassionate empathy may be seen as a step towards proactive and topical engagement. We hypothesize that a compassionate agent is capable of: (a) reaching out to the user based on what they have learned about them in the recent dialogue (as seen in part (i) of Figure 1) and (b) incorporate the agent's knowledge of the user based on previous conversations in new contexts (as seen in part (ii) of Figure 1). Therefore, we propose that using the affect and content analyzed in above, we develop a multi-tiered user model which incorporates a notion of memory and recency. Such a model would be both proactive and emotionally-aware, and can therefore be considered at least minimally compassionate.

II. Related Work

In this section, we explore some topical and relevant work that has been done in field of computational empathy as well as in the incorporation of emotions into conversational agents and dialogue systems. We also detail some contemporary work in intelligent personal assistant design.

Yalçın and DiPaola (2018) [36] provide a definition of computational empathy that relates to the identification and appropriate response to human emotion within the realm of interactive agents. This definition has been inspired by work in cognitive and behavioural psychology [4], [27], and has been widely adopted in contemporary literature. Work in human-robot interaction additionally provides multimodal analysis of the emulation of behavioural bases of empathy such as a notion of empathy modulation and the importance of the agent's degree of agency [21], [22].

Empathy in virtual agents has been studied in detail in specific contexts such as childcare and support [12], healthcare and patient support [33], autism and special needs intervention [20], and mental healthcare [19]. Within these applications, while the fundamental nature of the interaction demands an understanding of the user's emotional state and interpreting and acting according to it, there are key differences in their approach to modeling and understanding emotion in the users, such as the modality of interaction, the features or factors the agents gauge, and the mode of response generation. Studies such as Powell and Roberts (2017) [26] also highlight the predictors of empathetic response in digital or computer-mediated interaction. Since personal digital assistants will always interact with the user using a digital framework, the profiles of cognitive, affective, and compassionate empathy need to be attuned as well.

Within dialogue systems, emotion recognition has been an active area of research with models being developed to identify emotions in a given input text as well as in context of the user interaction [8], [24], [25]. Corpora such as EMPATHETICDIALOGUES [30] and PersonaChat [18] established the current baseline corpora for empathetic dialogue generation with an integrated emotion recognition component, which paradigmatically follows the breakdown of empathy into affective and cognitive representations. Models such as EmpDG [14], MoEL [15], and Caire [16] all have trained on and evaluated against the EMPATHETICDIALOGUES corpus, and claim to showcase an improvement in empathetic and natural response.

Compassion has not been explored in as much detail in personal assistants or dialogue systems. Within robotics, the notion of a compassionate AI has been introduced as a design choice [3], [9], as well as for analyzing expression features in affective robotics [13].

III. Methodology and Experiment Design

In this section, we elaborate upon the methodology that we aim to use to explore and answer the main research questions detailed in Section I. With the aim of being used as an interface to a personal assistant system, our experiment design is also generally user-centric. We establish two phases of the methodological solution to the research problem: the Interaction Cycle and the Reflection Cycle.

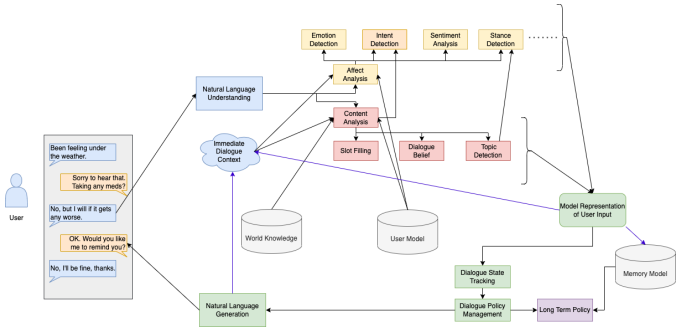

Fig. 2. The User Interaction Cycle: Diagrammatically

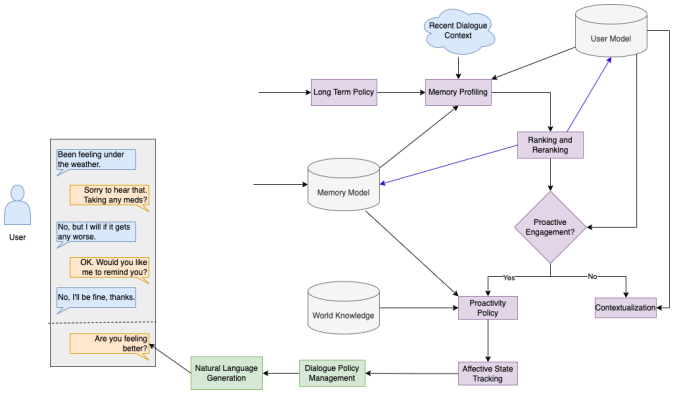

Fig. 3. The Reflection Cycle: Diagrammatically

A. User Interaction Cycle

As the name suggests, the user interaction cycle of a compassionate digital assistant is one which occurs when the user interacts with the affective interface we are developing. Building on contemporary research in dialogue systems, we aim to first develop a mechanism which evaluates a dialogue system on how well it portrays empathetic understanding. We then aim to incorporate a variety of affective states such as sentiment, intent, stance towards the topic of conversation, and incorporate it into the evaluation metric.

For evaluating generative empathetic dialogue models, we establish three metrics, the automated metric, manual preferential metric, and manual interaction metric.

- The automated performance metrics aim to measure quality of content as well as emotional response against an adversarial test corpus. Thus, we include the traditional perplexity, BLEU and F1 score evaluation [5] alongside analyzing sentiment appropriateness using the rule-based evaluation VADER [10], and a comparison of valence, arousal, and dominance between the agent response and the expected response to a given evaluation input to the model inspired by Park et. al. 2021 [23].
- The first manual evaluation criterion is based on an A/B preferential selection on generated responses to natural and adversarial prompts, and evaluate user preference on style, awareness, and appropriateness of response, on which we can perform a z-type analysis of variance test (ANOVA) as done by Ghandeharioun et. al. (2019) [7] and Chin et. al. (2020) [2]. The aggregate preferential selection is then correlated with the automated measures to develop an *appropriateness metric*.
- The second manual evaluation criterion is based on user experience with a trained model, wherein the user interacts directly with the model and is then asked to rate their experience. This metric can be used for comparative user evaluation of a suite of agents in order to establish qualitative insights into how appropriately and with how much empathy an agent communicates with a user in a more unconstrained environment inspired by the a combination of the `AlexaPrize` evaluation methodology [34] and user trust evaluation for proactive agents [11].

After an evaluation methodology is established, the next step would be to integrate components of hierarchical memory-oriented user model which the agent can access as a knowledge base of the user. While the focus of this research is not on the design of the user model itself, nor on the interaction between the user model and the affective interface, we aim to develop and streamline an architecture wherein the user model is hierarchically defined based on recency of information, and can be updated after a given user interaction through an active and evolving process (the reflection cycle, Section III-B). We also aim to expand upon the affective states which the model has access to beyond just emotion, including sentiment, stance, motivation, emotional need, and intent. These affective states have been studied individually before, but their combined effect on generating contextual and empathetic dialogue responses has not been evaluated before.

The user interaction cycle is detailed in Figure 2. The figure presents a framework for a dialogue system which integrates several NLP tools to detect the aforementioned affective states, along with access to world knowledge. We also explicitly keep track of the immediate dialogue context as well as the responses provided by the agent. Finally, there is a difference between user model, world knowledge, and a 'memory model', which update based on the relative importance of other information the agent would have learned about the user.

B. Reflection Cycle

The reflection cycle is established as a secondary cycle which the interface participates in after an episodic interaction with a user has completed. We posit that after a given user interaction, the information collected about the user in this interaction should improve the agent's understanding of the user. The reflection cycle is elaborated upon in 3.

Of note, the reflection cycle functions immediately after an interaction has completed by assessing the importance of the conversation, filtering the aspects of the exchange that need to be remembered and recollected. A heuristic determination of the information then either updates the user model or the memory model. The memory model keeps access of user states

which are transient, while the user model updates for user characteristics, both of which can be augmented by world knowledge and are fed into a proactivity policy so as to compare whether or not the agent should contact the user based on the contents of the current conversation.

The aim of the reflection cycle is to update the agent's information about the user, both in the short and the long term. Therefore, using a given contextualized dialogue, the user's statements are to be filtered and ranked in context of the current discussion as well as the existing information the agent possesses about the user (as an extended search and filter problem [32]). The architecture of this cycle can be heuristically optimized for the user's preferences, is modular such that other aspects of the user's input can be processed in a parallel cycle, and is scalable, such that other characteristics of text such as user feedback can be filtered and incorporated appropriately into the agent design.

IV. Current Status and Future Work

The author of this paper is a first year Ph.D. student. In the first semester, we focused on developing and refining the idea of empathy and compassion in interactive agents, as well as reviewing the current literature on this topic. We found that the notion of empathy has been worked on in the context of interactive systems and conversational agents. The fields of human-computer interaction, affective computing, and natural language processing, each had made progress in developing empathetic systems either in very task specific environments such as empathy in teacher-student interactions [35], or in very general terms such as for a general purpose dialogue system. Literature review also indicated the differences between empathy in humans and computational empathy, especially around the use of the term "emotion". Emotion is not defined in the context of computational empathy, it is simply assumed that the training corpus has the categorical identification necessary. We considered this a gap in the literature, and realized that models could benefit from being provided a host of affective states like emotion, as mentioned before. However, disparate foci of empathetic modeling also meant a wide array of possible evaluation metrics, and thus, the first step of unifying the evaluation processes was focused on.

In the second semester, the goal was to develop a reasonable adversarial corpus and a testbench for a suite of conversational agents. Since our evaluation has both manual and automated components, we first chose a set of models which could be fine-tuned on corpora that focus on empathetic understanding, to see if users believe these models to be more emotionally viable and capable than models which are not trained on such corpora. We are currently fine-tuning the T5 [29], DialoGPT [38], BlenderBot [31], and a baseline GPT-2 [28] based dialogue models, being fine-tuned on the EmpatheticDialogues [30] corpus. The automated evaluation is to be done against a randomized selection of dialogue exchanges other baseline dialogue corpora. Our plans for manual evaluation are currently under Ethics Review as they involve taking user feedback on the performance, trust, usability, and other metrics of these models.

Once the evaluation is established, we aim to introduce the affective states into the encoding pipeline of the models, as well as develop an integrated user and memory model for introducing the personal assistant aspect of the work. We then intend to experiment with the method of integrating the various affective states to the encoded input in order to best effect and improve model performance, while evaluating the model on the established testbench after every change. Once a reasonable suite of affective states has been integrated, we aim to use the content to develop the reflection cycle mentioned in Section III-B in order to improve the agent's ability to proactively engage with the user, a step towards a minimally compassionate personal digital assistant.

V. Conclusion

In this paper, we provided a brief research plan surrounding the question: "How do we design a personal digital assistant which behaves with empathy and compassion?" First, we aimed to define these terms based on extant literature. We explored the possible methods of evaluating the emotional response and empathy of a given conversational agent if it were to be the interface to a PDA. We then aimed to establish a novel, scalable framework in which latent characteristics could be extracted from the user's input to inform the model's response to the user, to be evaluated against the established testbench. Finally, we aim to integrated a user model which updates based on user input in a novel and unconventional design, non-interactive update: the user model updates when the user is NOT conversing with the agent. Dubbed the reflection cycle, the aim of this additional component is to improve the model's awareness of the user and introducing a sense of proactive response. We posit that a proactive and empathetic system is at least minimally compassionate, by definition.

To our knowledge, this PhD work is among the first to establish a notion of compassion in interactive agents. We believe that this research question focuses on the design and architecture of interactive agents, thereby having practical goals of developing a representative and useful model while still provoking interesting theoretical questions in affective computing (training and evaluation of empathy in interactive agents), human-computer interaction (emotion understanding outside user-agent interaction cycle), representation learning in NLP (representation of affective states in text), and dialogue system architecture (difference between social chatbots and personal assistants). We believe that this work fundamentally shapes the direction of future research in interactive agents.

References

[1] de Barcelos Silva, A., Gomes, M.M., da Costa, C.A., da Rosa Righi, R., Barbosa, J.L.V., Pessin, G., De Doncker, G. and Federizzi, G., 2020. Intelligent personal assistants: A systematic literature review. Expert Systems with Applications, 147, p.113193.

[2] Chin, H., Molefi, L.W. and Yi, M.Y., 2020, April. Empathy is all you need: How a conversational agent should respond to verbal abuse. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-13

[3] Barua, R., Sramon, S. and Heerink, M., 2015. Empathy, compassion and social robots: an approach from Buddhist philosophy. philosophy, 3(4).

[4] Coplan, A. and Goldie, P. eds., 2011. Empathy: Philosophical and psychological perspectives. Oxford University Press.

[5] Deriu, J., Rodrigo, A., Otegi, A., Echegoyen, G., Rosset, S., Agirre, E. and Cieliebak, M., 2021. Survey on evaluation methods for dialogue systems. Artificial Intelligence Review, 54(1), pp.755-810.

[6] Ekman, P., 2004. Emotions revealed. Bmj, 328(Suppl S5).

[7] Ghandeharioun, A., Shen, J.H., Jaques, N., Ferguson, C., Jones, N., Lapedriza, A. and Picard, R., 2019. Approximating interactive human evaluation with self-play for open-domain dialog systems. Advances in Neural Information Processing Systems, 32.

[8] Ghosal, D., Majumder, N., Poria, S., Chhaya, N. and Gelbukh, A., 2019, November. DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 154-164).

[9] Hughes, J., 2012. Compassionate AI and selfless robots: A buddhist approach. Robot ethics: the ethical and social implications of robotics, pp.69-83.

[10] Hutto, C. and Gilbert, E., 2014, May. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media (Vol. 8, No. 1, pp. 216-225).

[11] Kraus, M., Wagner, N., Callejas, Z. and Minker, W., 2021. The role of trust in proactive conversational assistants. IEEE Access, 9, pp.112821-112836.

[12] Leite, I., Castellano, G., Pereira, A., Martinho, C. and Paiva, A., 2014. Empathic robots for long-term interaction. International Journal of Social Robotics, 6(3), pp.329-341.

[13] Lewandowska-Tomaszczyk, B. and Wilson, P.A., 2018. Compassion Cluster Expression Features in Affective Robotics from. Cognitive Infocommunications, Theory and Applications, 13, p.201.

[14] Li, Q., Chen, H., Ren, Z., Ren, P., Tu, Z. and Chen, Z., 2020, December. EmpDG: Multi-resolution Interactive Empathetic Dialogue Generation. In Proceedings of the 28th International Conference on Computational Linguistics (pp. 4454-4466).

[15] Lin, Z., Madotto, A., Shin, J., Xu, P. and Fung, P., 2019, November. MoEL: Mixture of Empathetic Listeners. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 121-132).

[16] Lin, Z., Xu, P., Winata, G.I., Siddique, F.B., Liu, Z., Shin, J. and Fung, P., 2020, April. Caire: An end-to-end empathetic chatbot. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 09, pp. 13622-13623).

[17] Ma, Y., Nguyen, K.L., Xing, F.Z. and Cambria, E., 2020. A survey on empathetic dialogue systems. Information Fusion, 64, pp.50-70.

[18] Madotto, A., Lin, Z., Wu, C.S. and Fung, P., 2019, July. Personalizing dialogue agents via meta-learning. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 5454-5459).

[19] Miller, E. and Polson, D., 2019. Apps, avatars, and robots: The future of mental healthcare. Issues in mental health nursing, 40(3), pp.208-214.

[20] Nikopoulou, V.A., Holeva, V., Kerasidou, M.D., Kechayas, P., Papadopoulou, M., Vrochidou, E., Papakostas, G.A. and Kaburlasos, V.G., 2020. Identifying linguistic cues; towards developing robots with empathy in autism interventions. Journal of Clinical Medicine of Kazakhstan, 2(56), pp.27-33.

[21] Paiva, A., Leite, I., Boukricha, H. and Wachsmuth, I., 2017. Empathy in virtual agents and robots: A survey. ACM Transactions on Interactive Intelligent Systems (TiiS), 7(3), pp.1-40.

[22] Paiva, A., Correia, F., Oliveira, R., Santos, F. and Arriaga, P., 2021. Empathy and Prosociality in Social Agents. In The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition (pp. 385-432).

[23] Park, S., Kim, J., Ye, S., Jeon, J., Park, H.Y. and Oh, A., 2021, November. Dimensional Emotion Detection from Categorical Emotion. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (pp. 4367-4380

[24] Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E. and Mihalcea, R., 2019, July. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 527-536).

[25] Poria, S., Majumder, N., Mihalcea, R. and Hovy, E., 2019. Emotion recognition in conversation: Research challenges, datasets, and recent advances. IEEE Access, 7, pp.100943-100953.

[26] Powell, P.A. and Roberts, J., 2017. Situational determinants of cognitive, affective, and compassionate empathy in naturalistic digital interactions. Computers in Human Behavior, 68, pp.137-148.

[27] Preston, S.D. and De Waal, F.B., 2002. Empathy: Its ultimate and proximate bases. Behavioral and brain sciences, 25(1), pp.1-20.

[28] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. and Sutskever, I., 2019. Language models are unsupervised multitask learners. OpenAI blog, 1(8), p.9.

[29] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. and Liu, P.J., 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research, 21, pp.1-67.

[30] Rashkin, H., Smith, E.M., Li, M. and Boureau, Y.L., 2019, July. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 5370-5381).

[31] Roller, S., Dinan, E., Goyal, N., Ju, D., Williamson, M., Liu, Y., Xu, J., Ott, M., Smith, E.M., Boureau, Y.L. and Weston, J., 2021, April. Recipes for Building an Open-Domain Chatbot. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume (pp. 300-325).

[32] Shuster, K., Komeili, M., Adolphs, L., Roller, S., Szlam, A. and Weston, J., 2022. Language models that seek for knowledge: Modular search & generation for dialogue and prompt completion. arXiv preprint arXiv:2203.13224.

[33] Vallverdú, J. and Casacuberta, D., 2015. Ethical and technical aspects of emotions to create empathy in medical machines. In Machine Medical Ethics (pp. 341-362). Springer, Cham.

[34] Venkatesh, A., Khatri, C., Ram, A., Guo, F., Gabriel, R., Nagar, A., Prasad, R., Cheng, M., Hedayatnia, B., Metallinou, A. and Goel, R., 2018. On evaluating and comparing conversational agents. arXiv preprint arXiv:1801.03625, 4, pp.60-68.

[35] Villarica, R. and Richards, D., 2014, December. Intelligent and empathic agent to support student learning in virtual worlds. In Proceedings of the 2014 Conference on Interactive Entertainment (pp. 1-9).

[36] Yalçın, Ö.N. and DiPaola, S., 2018. A computational model of empathy for interactive agents. Biologically inspired cognitive architectures, 26, pp.20-25.

[37] Yalçın, Ö.N. and DiPaola, S., 2020. Modeling empathy: building a link between affective and cognitive processes. Artificial Intelligence Review, 53(4), pp.2983-3006.

[38] Zhang, Y., Sun, S., Galley, M., Chen, Y.C., Brockett, C., Gao, X., Gao, J., Liu, J. and Dolan, W.B., 2020, July. DIALOGPT: Large-Scale Generative Pre-training for Conversational Response Generation. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations (pp. 270-278).