2022 8th International Conference on Control, Decision and Information Technologies
CoDIT'22 | Istanbul, Turkey / May 17-20, 2022

Empathy scale adaptation for artificial agents: a review with a new subscale proposal

Harika Putta[1], Karl Daher[2], Mira El Kamali[2], Omar Abou Khaled[3], Denis Lalanne[3], Elena Mugellini[3]

2022 8th International Conference on Control, Decision and Information Technologies (CoDIT) | 978-1-6654-9607-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/CODIT55151.2022.9803993

***Abstract*— The communication between humans and artificial agents is becoming crucial and significant in daily life, especially with the advancements in the fields of human-robot and human-computer interaction. For these artificial agents to be recognized as social beings, they should exhibit emotional and empathic behaviors. However, there is no global agreement on measuring the empathic capabilities of these agents. For this reason, the scientific community has paid a significant focus on developing a standardized metric to perceive artificial agents' empathy. In this regard, this article provides a discussion on challenges in artificial empathy evaluation and researches the developments to discuss the factors and recommendations to design a globally accepted metric. It also discusses the qualities required for a globally accepted and standardized metric. Finally, an adaptation to an existing questionnaire is proposed for the evaluation of empathy in artificial agents.**

Keywords—Artificial Empathy; Empathy Evaluation; Human-Robot Interaction; Social Robots

I. INTRODUCTION

Empathy is the "feeling-into" defined by Theodor Lipps [1], the father of scientific theory in 1903. The relationship between ourselves and others is a crucial aspect of emotional intelligence since it is how we perceive what others encounter. The definition of empathy has evolved, and research is defining empathy from different perspectives leading to 43 distinct definitions [2]. According to Wan et al. [3], Empathy is the ability to understand or experience what a person is feeling from the perspective of that person. According to the literature on cognitive psychology and Human-Robot Interaction (HRI), empathy is a complex process by which the frame of reference of an entity is understood and/or communicated and/or responded appropriately without doubts about which frame of reference belongs to whom [4]. It is a critical component of interaction. There are two analytic categories of empathy from the literature: Cognitive empathy and Affective empathy [4], [5]. Cognitive empathy is the capability to understand the feelings and emotional states of others [4]. It improves our communication abilities. Affective empathy, also known as Emotional empathy is the capability to share the feelings of others by eliciting automatic emotional responses [6]. It helps us in developing emotional bonds with people. A new category of empathy has emerged in recent years, called Compassionate empathy. It is the ability to go beyond just understanding and sharing the feelings of others by moving forward to act and help them [7]. There are several subcategories of Cognitive empathy: Empathic understanding and empathic response are the most important ones [4]. Empathic understanding is the capability to understand another person's unique emotions and experiences. The empathic response is the capability to convey the attempts we make to understand someone's feelings and motives from their frame of reference.

As technology advances, the reliance of humans on artificial agents is increasing in society. People are often seeking the help of robots and artificial agents for various reasons which emphasizes the significance of HRI. According to Paiva et al. [8], empathic agents are the agents which give emotional responses that are more congruent with the user's or another agent's situation, or the agents which lead the users to give emotional responses that are more congruent with the agent's situation rather than their own situation. With the wide use of smartphones and the availability of the Internet, personal assistants like Amazon Alexa, Google Assistant, and Apple Siri became more popular. But these social agents should perform in a Human-like manner to accomplish what they are made for. With this, the scientific community worldwide focused on embedding empathy into these artificial agents. Consequently, the need to measure the artificial agents' empathy arose and various attempts were made to evaluate it. However, there is no global agreement on the evaluation of artificial empathy [5], [6]. A standardized and globally accepted metric or evaluation is always necessary to compare across different models and agents. Without a common scale of measuring, one can neither compare nor draw conclusions about the factors, qualities, or quantities being measured in different studies. Bartneck et al. [9] described the importance of developing standardized measurement tools for progressing in the field of HRI. Hence, a globally agreed empathy metric is of utmost need to excel in the domain of artificial empathy. This aids to identify the areas of improvement in the artificial agents and helps to design more realistic empathic artificial agents that can strengthen the human-robot relationship. This paper presents the literature review of the existing methods, the potential challenges in measuring artificial empathy, and finally presented a newly adapted version of an empathy

[1]Harika Putta is with the Master's in Computer Science, BeNeFri Program, University of Bern, Switzerland (e-mail: harika.putta@students.unibe.ch).

[2]Karl Daher is with the HumanTech Institute, University of Applied Sciences of Western Switzerland, Fribourg, Switzerland (e-mail: karl.daher@hes-so.ch).

[2]Mira El Kamali is with the HumanTech Institute, University of Applied Sciences of Western Switzerland, Fribourg, Switzerland (e-mail: mira.elkamali@hes-so.ch).

[3]Omar Abou Khaled is with the HumanTech Institute, University of Applied Sciences of Western Switzerland, Fribourg, Switzerland (e-mail: omar.aboukhaled@hes-so.ch).

[3]Denis Lalanne Kamali is with the Human-IST Institute, University of Fribourg, Fribourg, Switzerland (e-mail: denis.lalanne@unifr.ch).

[3]Elena Mugellini is with the HumanTech Institute, University of Applied Sciences of Western Switzerland, Fribourg, Switzerland (e-mail: elena.mugellini@hes-so.ch).

scale covering the new aspects of artificial empathy measurements that the previous works lacked.

II. Related Work

Yalçin [6] proposed some recommendations for evaluating artificial agents' empathy in a systematic approach. Since there is no global agreement on the evaluation of empathy in artificial agents, the human empathy evaluation methods were examined and the factors to be considered while designing a global metric are suggested.

Charrier et al. [5] constructed a new questionnaire named the RoPE Scale (Robot's Perceived Empathy), which is more specific towards measuring the robot's empathy in HRI. The items are based on human empathy metrics adapted to robots avoiding capabilities that robots don't have. Also, they claimed that the questionnaire is evaluated by experts and achieved reproducibility, validity, and comparability in artificial empathy research.

McQuiggan et al. [10] designed the CARE framework, a methodology for artificial agents to learn empathy models by observing human-human social interactions. They defined Empathic accuracy as the precision with which an empathic agent evaluates another's thoughts and emotions in a social context and then acts empathetically. They described that the empathic accuracy of an empathy model can be determined with two complementary evaluations: (1) Predictive accuracy, which reveals the degree to which an artificial empathy model makes similar empathic decisions to those made by humans, (2) Perceived accuracy, which reveals the degree to which an artificial empathy model makes situationally appropriate empathic decisions.

In their work Thies et al. [11] researched the difference between three chatbots with different personality orientations: productivity, fun, and emotion, where they analyzed the peoples' responses to the chatbots' behavior. Their analysis has shown that people tend to prefer the productivity bot, which can add value to their life, but with a fun and emotional behavior concluding that users preferred empathic chatbots. The results of Khan et al. [12] showed a more positive effect towards attractive agents than unattractive ones. The work presented a foundation for the user evaluation of embodied conversational agents in particular and artificial agents in general. All these works showed the importance of empathy in artificial agents which indirectly points to the need for measuring artificial empathy.

Challenges of Empathy Evaluation in Artificial Agents

There are various reasons for not having a standardized method for HRI empathy evaluation. One primary concern is the differences in the interaction levels and application goals of different artificial agents [6]. E.g., a healthcare professional may require greater compassion, whilst a fitness instructor may concentrate on pushing the interacting partner's limits. Variations in characteristics and non-functional properties like aesthetics, embodiment, fluency, and response time of the agents are another concern [6].

Most of the existing empathy evaluation methods used objective-type questionnaires [5], [6], [13], [14] and avoided subjective reporting methods. Because the subjective self-reporting measures cause uncertainty as people often have problems or reluctance in conveying their emotional state which is inappropriate in HRI [13], [14].

Empathy Evaluation in Human-Human Interaction

Psychology has numerous recognized approaches to measure empathy levels in humans. However, they often tend to be very subjective [6]. Besides, they depend largely on physiological information generated by neural activity, pulse rate, and galvanic skin response, which do not apply to artificial agents. Some metrics treat empathy as a single, comprehensive concept that derives a single global value. On the other hand, certain metrics consider various features and components of empathy like affective and cognitive capacities to calculate the overall empathic response of the artificial agent [6].

There are various measures to evaluate empathy in HHI. Hogan's Empathy Scale [15], a 64 true-false statements self-measure can assess cognitive empathy. Davis's Interpersonal Reactivity Index (IRI) [16] is also a widely used 28-item scale with four sub-scales for measuring multi-dimensional empathy. The Barrett-Lennart Relationship Inventory (BLRI) [17] is one of the widely used client-rated empathy measures in the form of questionnaires. Empathy Quotient (EQ) [18] is an extensively used self-report scale with 40 empathy and 20 filler items. Jefferson Scale of Physician Empathy [19] is another scale created to assess empathy as a primarily cognitive characteristic.

HHI Empathy Evaluation Methods for HRI Empathy Evaluation

In the HRI, robots and artificial agents' empathy is evaluated mostly with the slightly changed versions of HHI questionnaires. This induces an unnecessary bias in the evaluation as the robots do not have the cognitive abilities that humans have. Charrier et al. [5] described that Human empathy and Robot empathy cannot be measured with the same methodology even though they are similar.

Hence, the problem of using modified versions of self-reported measures and the questionnaires from the HHI domain in Artificial empathy measurement must be addressed to avoid biased results. Furthermore, the challenges that are complicating the measurement of artificial empathy should be addressed. RoPE scale [5] is an initiative towards establishing a metric for measuring artificial empathy. However, there is some room for improvement to make it a global and standardized scale. The need for standardized metrics to excel in the HRI domain is also acknowledged at the third HRI conference in 2008. As a result, the purpose of this paper is to contribute towards filling this gap in artificial empathy measurement.

III. Qualities of a Globally Accepted Metric

With quite many definitions given for empathy, the evaluation metrics used to assess empathic behavior might be rather different. Also, the variations and differences in the characteristics like aesthetics and embodiment make it challenging to apply the same evaluation method to different artificial agents. To overcome these factors and for an empathy evaluation metric to be globally accepted, the empathy evaluation metric should satisfy some qualities.

After an intense literature study, we came up with a list of qualities that are to be possessed by a standardized global empathy metric. Below is the list and the explanation of why they are important:

M1. Reproducibility

Using objective questionnaires for empathy evaluation is one of the most widely used techniques to date. However, the individuals' mood in answering the surveys/questionnaires at a point in time might affect the empathic evaluation. Hence, the metric should be able to produce consistent results every time it is used. Most of the metrics tend to fail when used at different intervals of time.

M2. Reliability

If tested on similar or equivalent artificial agents under the same conditions, the results must be similar. A measurement metric is only considered reliable when, under consistent conditions, it provides similar results.

M3. Sensibility

Different artificial agents exhibit varying levels of interactions, depending on the intent of use. The mode of communication varies from text, speech, and touch to facial expressions. The characteristics and non-functional properties play a vital role. The empathy evaluation metric should be sensible enough to capture even the small hints produced by the agents.

M4. Generalizability

Different definitions of empathy dramatically alter how it is interpreted. Therefore, there should be a generalization of the measure to suit any definition given to empathy.

M5. Validity

The determined empathy evaluation method should be validated by experts and bound to the concepts and behaviors related to empathy. There are many evaluation metrics in HHI which are considered to be invalid for various reasons.

M6. Comparability

When a new metric is developed, it is essential to compare it with various state of the Art methods or different standards. The empathy evaluation method should easily correlate to the results of other evaluation methods.

M7. Flexibility

The fields of usage and the application objectives of artificial agents are extremely contextual. Some artificial agents need more empathy, whereas others need to be more stubborn and ambitious. Consequently, the empathy evaluation metric should be flexible to fit all those contexts.

IV. Empathy Evaluation in Artificial Agents

According to Yalçin [6], many empathy evaluation methods from psychology literature involve measuring the specific cognitive and behavioral capabilities separately and also the overall empathy. Similarly, while constructing an affective empathy evaluation method, entire computational systems' performance depends on the discrete components' performance and precision as well as their integration at the system level. However, due to the multi-component nature of artificial agents and their complexity, it is recommended to perform the feature-level and system-level evaluations separately.

System-level evaluation concentrates on the measurement of the systems' behavior in a broader sense and focuses on capturing the artificial agent's global empathic behavior. This evaluation can be affected by various factors that are to be considered and controlled, such as user-related factors, context-related factors, and system-related factors [6]. User-related factors include human empathizing factors like gender, mood, personality, similarity, and social capabilities [20]–[22], individual traits like culture, socio-economic background, and computer experience. Context-related factors include efficiency, effectiveness, acceptability, user satisfaction, and utility that will impact the intensity and expression of empathic behavior [23]. System-related factors corresponding to the aesthetics, human-likeness, efficiency of movement, and trustworthiness of the artificial agents show a drastic influence on the perception of empathy [24], [25]. To regulate the effect of these factors, Human-Computer Interaction (HCI) research has already created various evaluation metrics [9].

Apart from the system-level evaluation, individual component evaluations are also required to evaluate the artificial agents' overall empathic capacity. The deficits in one capacity may have a significant impact on the other. For the global evaluation of empathy, the empathy traits to be evaluated are divided into three hierarchical mechanisms: emotional communication, emotion regulation, and cognitive processes [6]. Emotional communication is further classified as emotion recognition and emotion expression. The appropriate recognition of the emotions by the artificial agent will affect its empathic behavior at every stage of the communication. Similarly, the emotional expression ability of the agent would also directly influence the empathic behavior and its evaluation. Emotion regulation is based on the personality and mood factors of the person. Cognitive processes such as appraisal, re-appraisal, and perspective-taking behaviors are suggested to be included in the higher levels of empathy [26], [27]. However, no standardized method has evaluated these capabilities in artificial agents to date [6].

V. Questionnaire to Evaluate Empathy in Artificial Agents

This section presents a questionnaire to evaluate the artificial agents' empathy. Based on our literature review, we believe that the RoPE scale [5] is the only metric aimed at measuring artificial empathy. Our questionnaire presented in this section is an improvement of the RoPE scale with new items added to cover more aspects of empathic measurement which will be discussed later. We have made a detailed analysis of the RoPE scale and figured out the aspects it is covering from the previous discussions in the paper. The new adaptations made to RoPE aim to make the scale more appropriate and precise in measuring artificial empathy by covering the unattended aspects and factors of artificial empathy measurement. For these new adaptions, we explored other scales in the HRI domain focusing on the sub-scales that are aimed specifically to measure a particular quality of artificial agents. The new items added are inspired by Barrett-

Lennard Relationship Inventory (BLRI) [28], Godspeed questionnaire [29], Companionship Scale for Artificial Pets [30], and AttrakDiff questionnaire [31]. We chose to make improvements to the RoPE scale instead of making a new one for contributing to the comparability (M6) factor in building a standardized scale. The existing RoPE scale claimed improved reproducibility (M1), validity (M5), and comparability (M6). We decided to add the new items from the already validated existing scales to avoid the Psychometrician validation. Each new item is adopted to fulfill the uncovered aspects of artificial empathy in the RoPE scale based on the theoretical analysis. This will ensure the Face validity of the scale [32]. since all the items are from the validated scales, it is necessary to check the internal consistency of the scale which can be done by calculating the Cronbach Alpha coefficient. The 20 items of the RoPE scale are divided into two sub-scales called the Empathic Understanding (EU) and the Empathic Response (ER) each with 8 items along with four filler questions that avoid the subjects from guessing what the test is measuring. Since the RoPE scale is already validated, we took the two sub-scales without any modifications. One of our ideas is to use the filler items in a way that indirectly contributes to the empathic behavior analysis instead of asking random questions. So, we took only two of the filler items (Q17, Q18) from the RoPE scale and added two new items (Q19, Q20) of our own (see Tab. 1). The new scale has a total of 25 items (18 from the RoPE scale and 7 new items). Item (Q19) is inspired by the Focus-on-Empathy forms and item (Q20) is inspired by the Basic 64-item forms of the BLRI scale [28]. Items (Q21, Q25) are designed with reference to the semantic differentials of Foolish-Sensible and Unintelligent-Intelligent of the Perceived Intelligence scale of the Godspeed questionnaire [29]. Items (Q22, Q23) are persuaded from the Enjoyment companionship factor of the Companionship Scale for Artificial Pets [30]. And finally, item (Q24) is framed with reference from the Attractiveness dimension of the AttrakDiff questionnaire with semantic differential word pairs unpleasant-pleasant, bad-good, and rejecting-inviting [31]. The newly added items are grouped as the Empathic Relationship subscale (see Tab. 1). This subscale will measure the bond the artificial agent creates with the user. The relation is based on the interaction between the artificial agent and the user, the comfort that the artificial agent would make the user feel, the prediction that can be made, and the perception that the user feels. The questions should be answered with a number between -3 and 3 on a Likert scale (-3, -2, -1, 1, 2, 3). An answer -3 indicates that "NO, I strongly feel that it is not true" and 3 indicates "YES, I strongly feel that it is true". The items marked with (-) are the questions asked in a negative sense and the filler items are marked with (F). Tab. 1 shows the total questionnaire where the items marked (N) are the new additions (items from Q19 to Q25) and the remaining are from the RoPE scale (items from Q1 to Q18) [5]:

TABLE I. QUESTIONNAIRE FOR MEASURING ARTIFICIAL EMPATHY

Item	Empathic Understanding subscale
Q1	The artificial agent/robot appreciates exactly how the things I experience feel to me.
Q2	The artificial agent/robot knows me and my needs.
Q3	The artificial agent/robot cares about my feelings.
Q4	The artificial agent/robot does not understand me. (-)
Q5	The artificial agent/robot perceives and accepts my individual characteristics.
Q6	The artificial agent/robot usually understands the whole of what I mean.
Q7	The artificial agent/robot reacts to my words but does not see the way I feel. (-)
Q8	The artificial agent/robot seems to feel bad when I am sad or disappointed.
Item	**Empathic Response subscale**
Q9	Whether thoughts or feelings I express are "good" or "bad" makes no difference to the artificial agent/robot's actions toward me. (-)
Q10	No matter what I tell about myself, the artificial agent/robot acts just the same. (-)
Q11	The artificial agent/robot comforts me when I am upset.
Q12	The artificial agent/robot encourages me.
Q13	The artificial agent/robot praises me when I have done something well.
Q14	The artificial agent/robot helps me when I need it.
Q15	The artificial agent/robot knows when I want to talk and lets me do so.
Q16	The artificial agent/robot's response to me is so fixed and automatic that I do not get through to it. (-)
Item	**Empathic Relationship subscale**
Q17	The way the artificial agent/robot acts feels natural. (F)
Q18	When I interact with the artificial agent/robot, I feel anxious. (F)
Q19	The artificial agent/robot respects me. (N) (F)
Q20	I feel comfortable to express what is in my mind with the artificial agent/robot, including feelings about myself or about the artificial agent/robot. (N) (F)
Q21	The artificial agent/robot makes appropriate/correct suggestions to me. (N)
Q22	I think that the artificial agent/robot is a living creature. (N)
Q23	The artificial agent/robot always gives me an immediate response. (N)
Q24	The artificial agent/robot's appearance/audio is pleasant, good, and inviting. (N)
Q25	I believe that the artificial agent/robot has a very appropriate behavior concerning its application or domain. (N)

The authors of the RoPE scale have not provided the reason for adding particular items to the scale. So, to understand the importance of each item, it is important to figure out which factors or qualities of empathy are being analyzed by it. Here, we tried to investigate the contribution of each item to identify the uncovered aspects and factors. This helped us in designing the new items to cover those untouched aspects found in the literature review. The RoPE scale has the items (Q11, Q12, Q13) which measure the perceived accuracy by clarifying if the artificial agent can make situationally appropriate decisions as mentioned in section 2. However, there are no items to measure the predictive accuracy. Hence, we added a new item (Q21) which says how much an artificial agent is making human-like decisions. There are also no items to measure the varying interactive capabilities and variations in domain-specific characteristics which are also crucial in measuring the empathy of an artificial agent. So, we added items (Q23) and (Q25) to measure each of them. As proposed by Yalçin [6] it is important to measure the user-related, context-related, and system-related factors while measuring empathy as pointed out in section 4. The RoPE scale has items that measure user-related factors like personality (Q5) and social capabilities (Q9). For the context-related factors, item (Q6) can measure efficiency, items (Q12, Q13) can measure user satisfaction, and item (Q18) (filler item in RoPE) can measure the utility of the artificial agent. We added item (Q20) which also

measures the utility of the artificial agent. In the system-related factors, the RoPE scale has only one item (Q17) (filler item) which measures the behavioral factor. Since the RoPE scale is lacking items to measure the context-related factors of artificial empathy, we added item (Q22) to measure human-likeliness and believability, item (Q23) to measure fluency, item (Q24) to measure aesthetics, and item (Q19) to measure the behavioral factors. Finally, the purpose of the newly proposed metric is to improve flexibility (M7) and sensibility (M3) along with reproducibility (M1), validity (M5), and comparability (M6) which are already claimed by the RoPE scale.

VI. Conclusion and Future Work

This article discussed the various aspects of evaluating empathy in artificial agents, like the challenges in artificial agents' empathy evaluation and problems of applying the Human-Human Interaction (HHI) empathy metrics to HRI. It provides an aggregation of the global qualities required for a globally accepted metric from a broad literature review. It discussed in detail the factors to be considered and various recommendations addressed in the HRI literature. Finally, it proposed an adaptation of an existing questionnaire to evaluate empathy in artificial agents, which together can achieve the qualities like reproducibility, validity, comparability, flexibility, and sensibility.

In the future, we would like to develop a generalized framework (work in progress) to evaluate the artificial empathy that can be applied to a wide range of artificial agents with different characteristics and varying levels of interaction capabilities. The aim of this framework is to be used along with the questionnaire so that it provides a standardized protocol to measure artificial empathy. We will also test the Empathic Relationship subscale of the questionnaire for reliability. In addition, we will perform the internal consistency test to ensure the inter-correlation of the items under scale validation.

References

[1] C. Montag, J. Gallinat, and A. Heinz, "Theodor Lipps and the concept of empathy: 1851–1914," *American Journal of Psychiatry*, vol. 165, no. 10, p. 1261, 2008.

[2] B. M. P. Cuff, S. J. Brown, L. Taylor, and D. J. Howat, "Empathy: A review of the concept," *Emotion review*, vol. 8, no. 2, pp. 144–153, 2016.

[3] Q. Wan, L. Jiang, Y. Zeng, and X. Wu, "A big-five personality model-based study of empathy behaviors in clinical nurses," *Nurse education in practice*, vol. 38, pp. 66–71, 2019.

[4] L. Charrier, A. Galdeano, A. Cordier, and M. Lefort, "Empathy Display Influence on Human-Robot Interactions: a Pilot Study Human-Robot Interactions: a Pilot Study. Workshop on Towards Intelligent Social Robots: From Naive Robots to Empathy Display Influence on Human-Robot Interactions: a Pilot Study," 2018.

[5] L. Charrier, A. Rieger, A. Galdeano, A. Cordier, M. Lefort, and S. Hassas, "The rope scale: a measure of how empathic a robot is perceived," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2019, pp. 656–657.

[6] Ö. N. Yalçın, "Evaluating empathy in artificial agents," *arXiv preprint arXiv:1908.05341*, 2019.

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

[8] A. Paiva, "Empathy in social agents," *International Journal of Virtual Reality*, vol. 10, no. 1, pp. 1–4, 2011.

[9] C. Bartneck, E. Croft, and D. Kulic, "Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots," 2008.

[10] S. W. McQuiggan and J. C. Lester, "Modeling and evaluating empathy in embodied companion agents," *International Journal of Human-Computer Studies*, vol. 65, no. 4, pp. 348–360, 2007.

[11] I. M. Thies, N. Menon, S. Magapu, M. Subramony, and J. O'neill, "How do you want your chatbot? An exploratory Wizard-of-Oz study with young, urban Indians," in *IFIP Conference on Human-Computer Interaction*, 2017, pp. 441–459.

[12] R. Khan and A. de Angeli, "The attractiveness stereotype in the evaluation of embodied conversational agents," in *IFIP Conference on Human-Computer Interaction*, 2009, pp. 85–97.

[13] C. L. Bylund and G. Makoul, "Examining empathy in medical encounters: an observational study using the empathic communication coding system," *Health communication*, vol. 18, no. 2, pp. 123–140, 2005.

[14] A. M. Rosenthal-Von Der Pütten *et al.*, "Investigations on empathy towards humans and robots using fMRI," *Computers in Human Behavior*, vol. 33, pp. 201–212, 2014.

[15] R. Hogan, "Development of an empathy scale.," *Journal of consulting and clinical psychology*, vol. 33, no. 3, p. 307, 1969.

[16] M. H. Davis, "Measuring individual differences in empathy: Evidence for a multidimensional approach.," *Journal of personality and social psychology*, vol. 44, no. 1, p. 113, 1983.

[17] G. T. Barrett-Lennard, "The relationship inventory: A technique for measuring therapeutic dimensions of an interpersonal relationship," 1959.

[18] S. Baron-Cohen and S. Wheelwright, "The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences," *Journal of autism and developmental disorders*, vol. 34, no. 2, pp. 163–175, 2004.

[19] M. Hojat, S. Mangione, J. S. Gonnella, T. Nasca, J. J. Veloski, and G. Kane, "Empathy in medical education and patient care," *Academic Medicine*, vol. 76, no. 7, p. 669, 2001.

[20] M. H. Davis and others, "A multidimensional approach to individual differences in empathy," 1980.

[21] M. L. Hoffman, *Empathy and moral development: Implications for caring and justice*. Cambridge University Press, 2001.

[22] F. de Vignemont and T. Singer, "The empathic brain: how, when and why?," *Trends in cognitive sciences*, vol. 10, no. 10, pp. 435–441, 2006.

[23] Z. Ruttkay, C. Dormann, and H. Noot, "Evaluating ecas-what, how and why?," 2006.

[24] C. Misselhorn, "Empathy with inanimate objects and the uncanny valley," *Minds and Machines*, vol. 19, no. 3, p. 345, 2009.

[25] A. B. Loyall, "Believable Agents: Building Interactive Personalities.," 1997.

[26] F. B. M. de Waal and S. D. Preston, "Mammalian empathy: behavioural manifestations and neural basis," *Nature Reviews Neuroscience*, vol. 18, no. 8, pp. 498–509, 2017.

[27] K. R. Scherer, "Componential emotion theory can inform models of emotional competence.," 2007.

[28] G. T. Barrett-Lennard, "The Relationship Inventory: A Complete Resource and Guide, First Edition The Relationship Inventory Forms and Scoring Keys," 2015.

[29] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, no. 1. Kluwer Academic Publishers, pp. 71–81, 2009. doi: 10.1007/s12369-008-0001-3.

[30] D. B. Luh, E. C. Li, and Y. J. Kao, "The development of a companionship scale for artificial pets," *Interacting with Computers*, vol. 27, no. 2, pp. 189–201, Nov. 2015, doi: 10.1093/iwc/iwt055.

[31] T. Walsh, P. Nurkka, J. Varsaluoma, H. Petrie, S. Kujala, and C. Power, "Axe UX: Exploring long-term user experience with iScale and AttrakDiff," in *MINDTREK 2014 - Proceedings of the 18th International Academic MindTrek Conference:* , Nov. 2014, pp. 32–39. doi: 10.1145/2676467.2676480.

[32] A. Aithal and P. S. Aithal, "Development and Validation of Survey Questionnaire & Experimental Data–A Systematical Review-based Statistical Approach," *International Journal of Management, Technology, and Social Sciences (IJMTS)*, vol. 5, no. 2, pp. 233–251, 2020.