Preliminary Thoughts on Designing and Evaluating Affective Human-Robot Interaction - Part I

Yunjia Sun[1] and Tao Wang[1]

[1]Peking University

Abstract

Affective robots, which interact with humans in fields like education, healthcare, and companionship, require both cognitive and emotional intelligence. Cognitive intelligence enables robots to perceive and understand human behavior, while emotional intelligence allows robots to empathize with and respond to human emotions. This article outlines the affective capabilities necessary for affective robots, including affective recognition and analysis, emotional strategy, emotional expression, state maintenance, proactive reaction, and safety and ethics, as criteria for evaluating these robots. In the challenges and practice section of this article, we also list selected works from the ACM/IEEE International Conference on Human-Robot Interaction (HRI) that focus on affective interaction, providing a preliminary foundation for these criteria. We will continue to update this article with more studies from recent literature and a broader range of studies in the field.

Although this article offers conceptual insights into affective human-robot interaction, it is **NOT** an academic paper.

1 Introduction

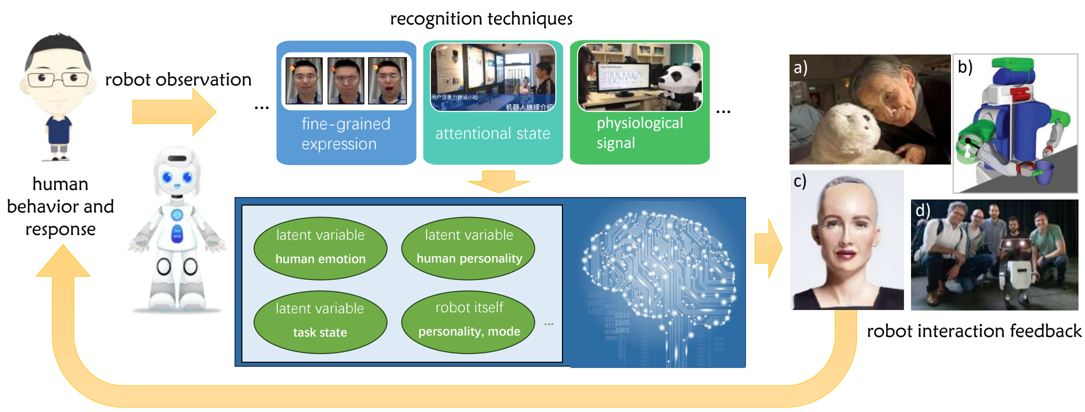


Figure 1: Effective interaction between affective robots and humans. The robot observes human external behavior and infers its internal state. Based on this inferred internal state, the robot selects an appropriate action to respond.[1]

[1]Images on the right are from:
a) https://www.strategy-business.com/article/Robotic-seals-and-bionic-limbs-How-Japan-is-creating-opportunity-for-medtech
b) Kwon *et al.*[1]
c) https://www.hansonrobotics.com/sophia/
d) https://spectrum.ieee.org/disney-robot

Affective robots serve as assistants and companions, engaging frequently with humans in education, healthcare, daily life, entertainment, companionship, and collaboration. Often depicted in science fiction, these robots must recognize, understand, and predict human language and behavior, including goals, intentions, desires, and beliefs. They focus on what is most urgent, relevant, or meaningful to people, deeply understanding human cognition and attention states. Emotional intelligence is equally essential, enabling robots to interpret emotional cues, respond appropriately, and convey their own "emotional state" during interactions.

To achieve the above ability, affective robots require both cognitive and emotional intelligence. Cognitive intelligence encompasses perceiving the environment, people, and objects, understanding human language, completing tasks, and learning from experience. This enables robots to sense, think, and react effectively. Emotional intelligence enables robots to interact with empathy, to understand human emotions and psychology. Human-centered design ensures robots engage in emotionally enriched interactions.

With these capabilities, affective robot finds applications across many fields. Affective robots provide emotional support as companions, ensure safety by monitoring environments, optimize tasks and schedules in management, act as personalized tutors in education, and enhance teamwork by adapting to team dynamics in collaborative settings.

An affective robot's primary task is to interact with humans. Fig. 1 illustrates an overall model of affective robots interacting with humans. The robot maintains hidden variables such as the emotional state, personality, and task status of the human it interacts with, as well as its own "personality" state. Based on the observed behaviors and reactions of the human, the robot processes observable variables, including speech, facial expressions, gestures, movements, eye contact, environmental images, commands, and other task-related or environmental factors. These observations are used to update the hidden variables, such as the human's emotions, personality traits, and task status. Using these updated hidden variables, the robot determines an appropriate interaction approach and engages with the human accordingly. The human, in turn, responds to the robot's interaction feedback with new behaviors and reactions. The robot continuously observes these new observable variables, updates the hidden variables in real-time, and generates a new round of interaction responses, creating an ongoing dynamic cycle.

To help realize the various functions of affective robots, this article outlines a series of affective capabilities that these robots should possess, including 1) affective recognition and analysis, 2) emotional strategy, 3) emotional expression 4) state maintenance, 5) proactive reaction, 6) safety and ethics, as the criteria of affective robots. Furthermore, we present some studies addressing the key points mentioned in the criteria.

2 Criterias for affective human-robot interaction

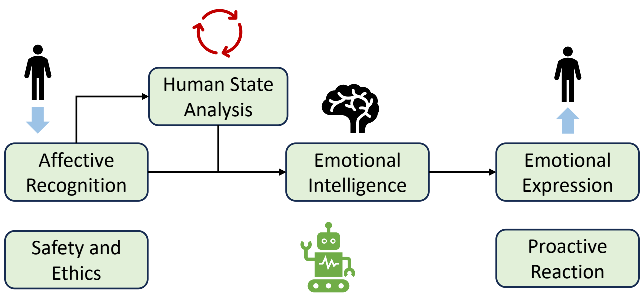


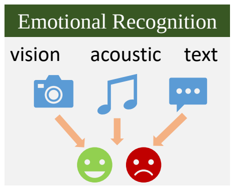
Figure 2: Main factors of human-robot interaction of affective robots, including recognition and analysis, emotional intelligence, emotional expression, proactive reaction, state maintenance, and safety and ethics.

To achieve safe, direct, and effective human-robot interaction (HRI) in affective robots, six key factors must be considered: 1) affective recognition and analysis, 2) emotional intelligence, 3) emotional expression, 4) state maintenance, 5) proactive reaction, 6) safety and ethics. Robots take the results of recognition and analysis as input and use emotional intelligence to choose interaction strategies and express them correctly. They also maintain states of themselves, the environment, tasks, the person they interact with, and other agents. The state maintenance ability provides the robots with the ability to react proactively. In addition, safety and ethics are crucial in HRI too. In the following subsections, we introduce each factor.

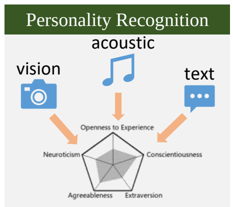
2.1 Affective Recognition

Affective recognition allows robots to interpret affective cues, enabling more adaptive and effective interactions.

Emotional Recognition refers to the ability to identify the emotions of interaction subjects accurately. For example, a sales robot must recognize a customer's emotional state to determine whether it is appropriate to proceed with a sales pitch. This capability allows robots to adapt their behavior based on emotional cues, enhancing user experience and interaction quality. Recognizing emotions is crucial because human emotions significantly influence interactions. Effective emotion recognition enables robots to engage in more natural and polite communication. Moreover, it serves as the foundation for emotional interventions, allowing robots to respond appropriately to users' emotional states and support their needs.



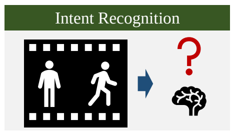
Personality Recognition refers to the ability to accurately identify the personality traits of interaction subjects. For example, a companion robot may recognize whether its user is introverted or extroverted, allowing it to minimize interruptions for introverts while engaging in more frequent conversations with extroverts. The importance of personality recognition lies in enabling robots to develop personalized strategies tailored to the identified personality. This process improves interaction effectiveness, enhances user satisfaction, and fosters more comfortable and engaging experiences.



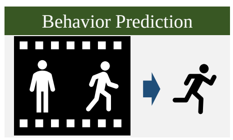
2.2 Human State Analysis

Human state analysis involves the ability of robots to assess and anticipate the actions, intentions, and behaviors of interaction subjects, enabling more proactive and efficient responses.

Intent Recognition refers to the ability to infer the intentions of interaction subjects. For instance, a household robot observing a human coughing while searching through cabinets may infer that the user intends to find water. This capability allows robots to proactively assist users, improving their overall experience and interaction efficiency. Intent recognition is important in enhancing interaction efficiency through proactive planning and optimizing the experience of interaction subjects. By anticipating user needs, robots can streamline tasks and provide more effective assistance.

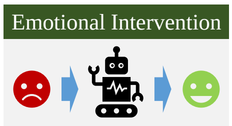


Behavior Prediction refers to the ability to anticipate the future actions of interaction subjects. This capability allows robots to respond proactively and adapt to dynamic situations, enhancing interaction quality and performance. For example, a robot navigating in a shared space may predict that a human is about to suddenly turn around, allowing the robot to avoid collision risks. Behavior prediction is important in improving interaction efficiency through proactive planning. By forecasting actions, robots can prepare appropriate responses, streamline processes, and optimize user experiences. Additionally, behavior prediction contributes to safety by enabling robots to detect and mitigate potential risks before they occur.

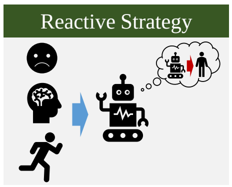


2.3 Emotional strategy

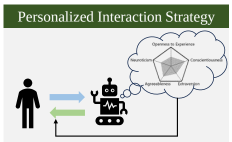
Emotional strategies involve the ability to analyze recognition results and make decisions to generate appropriate expressions or responses based on specific goals. Four abilities related to emotional strategy are listed below.



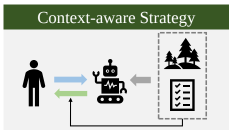
Emotion Intervention refers to the ability to devise strategies for intervening in the emotions of interaction subjects. For example, providing targeted emotional comfort to individuals with anxiety disorders. The importance of emotional intervention lies in its ability to provide emotional value to the interaction subjects, helping them stabilize their emotions and improve their overall well-being. By effectively intervening in emotions, robots can create more positive, supportive interactions and enhance the emotional experience of users.



Reactive Strategy refers to devising appropriate response strategies based on intent recognition and behavior prediction. For example, a robot navigating in a shared space may anticipate that a human might suddenly turn around and develop the optimal evasive route. The importance of reactive strategy planning lies in optimizing the user experience, enhancing adaptability in complex environments, and improving safety. By formulating effective strategies to handle dynamic situations, robots can better manage uncertainties and ensure smoother, more reliable interactions.



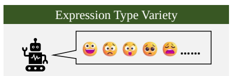
Personalized Interaction Strategy refers to develop tailored interaction approaches based on the characteristics of different subjects. For example, a robot may talk more when interacting with introverted individuals and listen more when engaging with extroverted ones. The importance of personalized interaction strategies lies in their ability to address individual differences, making interactions more relevant and engaging. By adapting to user preferences and behaviors, robots can handle a wider range of users, improving interaction efficiency and reducing communication costs through more targeted and meaningful exchanges.



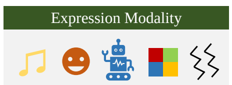
Context-aware Strategy refers to tailored interaction approaches developed based on different contexts, such as environments, tasks, and task states. For example, in a quiet environment, a robot may communicate with users through voice, whereas in a noisy environment, it may incorporate additional modalities, such as visual signals or text displays, to ensure the message is effectively conveyed. Context-aware strategies are important in improving interaction efficiency by aligning responses with situational needs. They also reduce exploration costs, as robots can quickly adapt to changing conditions without requiring extensive trial-and-error adjustments.

2.4 Emotional Expression

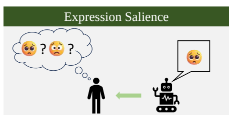
Expression capability refers to the ability to accurately and appropriately convey intended information, ensuring clarity and relevance in communication. Four abilities corresponding to emotional expression are listed below.



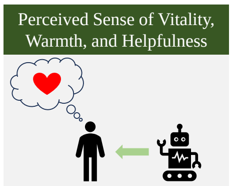
Expression Type Variety refers to the range of emotional types a robot can express. For example, companion robots or pet robots may need to express a wide variety of emotions to provide users with a genuine sense of companionship. The importance of comprehensive expression lies in enhancing a robot's ability to convey emotions, thereby improving emotional communication and strengthening user engagement. It also increases communication efficiency by enabling robots to respond more effectively to different interaction needs.



Expression Modality refers to the various forms through which a robot can express emotions. For example, in human-like communication, a robot may utilize not only language but also vocal tones, gestures, and facial expressions to convey emotions effectively. The importance of having diverse expression modalities lies in enabling more complete emotional expression, which enhances the depth and clarity of communication. Additionally, it allows robots to function effectively in a wider range of scenarios, accommodating varying interaction contexts and user needs.



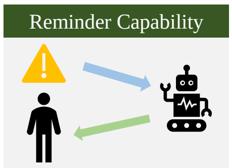
Expression Salience refers to the extent and accuracy with which a robot conveys emotional information. For example, a companion robot should be able to provide appropriate emotional feedback to its interaction partner, ensuring the conveyed response aligns with the partner's emotional state. The importance of accurate emotional expression lies in enhancing a robot's emotional communication capabilities, improving interaction efficiency, and preventing misunderstandings. By delivering clear and precise emotional signals, robots can foster trust and strengthen emotional connections with users.



Perceived Sense of Vitality, Warmth, and Helpfulness refers to a robot's ability to maintain a sense of liveliness, warmth, and helpfulness both during and outside task execution. For example, a robot may exhibit human-like behaviors such as blinking and breathing even when idle, creating a more lifelike presence. The importance of maintaining vitality, warmth, and helpfulness lies in making robots appear more "alive," fostering emotional connections, and increasing user trust. By projecting these qualities, robots can create a more engaging and reassuring interaction experience.

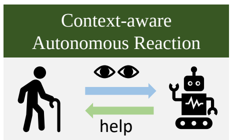
2.5 Proactive Reaction

Proactive reaction refers to the ability to make autonomous decisions during collaboration, addressing tasks and subjects of interaction with initiative and foresight. Three abilities related to proactive reaction are listed below.



Reminder Capability refers to a robot's ability to provide prompts or alerts when humans become distracted, make errors, or encounter unexpected situations. For example, during collaboration, a robot can remind a human partner to stay focused if signs of distraction are detected. The importance of reminder capability lies in improving task efficiency and enhancing safety. By actively identifying and addressing lapses, robots can help maintain smooth workflows and prevent potential errors or accidents.

The ability to Report Failures And Seek Assistance refers to a robot's ability to inform interaction partners when a task has failed and actively seek assistance. For example, during collaboration, if a robot encounters a task it cannot complete, it should notify the human partner and request help. The importance of failure notification and help-seeking capability lies in improving task success rates. By acknowledging limitations and involving humans when necessary, robots can ensure problems are addressed promptly, maintaining workflow efficiency and reliability.



Context-aware Autonomous Reaction refers to a robot's ability to respond proactively based on task conditions and the state of its interaction partner through contextual awareness. For example, in a warehouse setting, if a robot observes that a human partner is struggling to lift a heavy object, it can autonomously step in to assist. The importance of context-aware autonomous reactions lies in improving task efficiency and enhancing user satisfaction. By dynamically adapting to changing circumstances and providing timely assistance, robots can create smoother and more effective collaborative experiences.

2.6 Safety and Ethics

Safety and Ethics refer to principles that ensure robots operate without causing harm to their interaction partners, the environment, or themselves, while adhering to ethical standards. These principles form the foundation for building trust and accountability in human-robot interactions and are integral to the development of responsible robotic systems.

One of the earliest and most influential frameworks for robot safety and ethics is Isaac Asimov's *Three Laws of Robotics*, first introduced in his 1942 short story *Runaround*. These laws are as follows:

1. A robot may not harm a human being, or, through inaction, allow a human being to come to harm.

2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Asimov later introduced the *Zeroth Law*, which takes precedence over the other three laws:

> A robot may not harm humanity, or, by inaction, allow humanity to come to harm.

These laws have become a cornerstone in discussions about roboethics, highlighting the need for clear ethical guidelines in robotic behavior. However, their application in modern robotics presents challenges due to their abstract nature and the complexity of real-world scenarios.

The challenges of applying ethical principles in modern robotics are addressed by the field of *roboethics*, first proposed by Gianmarco Veruggio in 2002. Roboethics seeks to establish a comprehensive ethical framework for the design, development, and deployment of robots.

By adhering to these ethical principles, robots can become reliable partners in a variety of domains, ranging from healthcare and education to industry and social companionship, ultimately contributing to the well-being of individuals and society as a whole.

3 Challenges and Practices

While the elements discussed in Section 2 are all important for affective human-robot interaction, implementing them and applying them to robots is far from easy. This section reviews the challenges associated with each element and highlights some practices found in the literature.

3.1 Affective Recognition

Affective recognition is challenged by context-dependence, partial observability, and subjective interpretation, which complicate accurate identification and understanding.

Emotional Recognition faces several challenges. Human emotions are inherently complex and multidimensional, often involving subtle variations. Partial observability, where some individuals may not outwardly express their emotions, complicates the process. Furthermore, emotional expressions are context-dependent, which requires situational awareness for an accurate interpretation. Finally, emotions are subjective and may vary between individuals, adding further difficulty to the recognition process.

Personality Recognition faces challenges such as partial observability, as personality traits may not always be explicitly expressed and can only be inferred from limited cues. Subjectivity further complicates the process, as personality assessments often rely on interpretation and may vary across observers. Additionally, even interaction subjects themselves may have inaccurate self-perceptions, making it difficult to establish a reliable basis for identification.

3.2 Human State Analysis

Human state analysis faces challenges such as the invisibility of intentions and future behaviors, ambiguity in interpretation, and context-dependence.

Intent Recognition presents several challenges. Intentions are often invisible, making them difficult to detect directly. Additionally, uncertainty in interpreting ambiguous actions and the context-dependent nature of intentions further complicate recognition. Overcoming these challenges requires sophisticated modeling techniques and contextual awareness. Kato *et al.*[2] designed an approach to predict the intention to interact based on the trajectories of the customers. Huang *et al.*[3] predicted the user's task intent based on observed gaze patterns.

Behavior Prediction faces challenges due to the inherent invisibility of future actions, which must be inferred from past behaviors, making accurate predictions difficult. Additionally, uncertainty arises from the unpredictable nature of human behavior, influenced by factors such as emotions, intentions, and environmental changes. Moreover, behavior predictions are context-dependent, requiring a precise understanding of situational factors, which can be complex and highly variable. Ryoo *et al.*[4] presented a method for early activity prediction by analyzing robot-centric (first-person) videos, using the novel "onset" concept to predict human actions at their early stages.

3.3 Emotional strategy

Emotional strategy in affective robots involves creating flexible, context-aware responses that can adapt to individual differences and emotional states. The challenges include developing interventions and reactive strategies that are personalized and dynamically adjusted to each unique situation.

Emotion Intervention faces challenges involving the need for psychological expertise and a deep understanding of context to accurately interpret emotional states and design effective interventions. Additionally, it requires the ability to develop diverse intervention strategies tailored to different forms of emotional expression and often demands multi-turn interactions to achieve meaningful outcomes. Hoffman *et al.*[5] designed a peripheral robotic conversation companion that intervenes in conversations emotionally through simple postures.

Reactive Capabilities face challenges due to the vast sample space of intentions and behaviors, making it difficult to account for all possible scenarios. Additionally, reactions are highly context-dependent, requiring a precise understanding of situational factors to generate appropriate responses. Kato *et al.*[2] selected different strategies for approaching (or not approaching) customers based on their intent to interact, mimicking the behavior of human service providers.

Personalization faces challenges due to the large and diverse sample space of personalities, which makes it difficult to discretize and categorize them into distinct groups for effective modeling and adaptation. Ramachandran *et al.*[6] personalized the robot's timing strategies for providing breaks during tutoring by adjusting break schedules based on individual children's performance, using strategies tailored to either performance gains or drops.

Context-aware Strategy faces challenges due to the large and diverse sample space of contexts, making it difficult to generalize strategies across different situations. Additionally, new strategies may need to be developed when environments or tasks change, requiring adaptability and continuous learning. Hemminahaus *et al.*[7] developed a context-aware strategy by using reinforcement learning to adaptively select and optimize social behaviors based on the user's state and the interaction context during task-oriented interactions.

3.4 Emotional Expression

Emotional expression in affective robots involves creating expressions that are both varied and salient, while ensuring they align with the robot's capabilities and the user's expectations. The challenge lies in balancing complexity with clarity, so that emotional cues are effective and appropriately perceived.

Expanding Emotional Expression Types is challenging due to the large and diverse sample space of emotions, making it difficult to cover all possible emotional states. Furthermore, emotions are difficult to discretize, as they often exist on a spectrum and can be complex and nuanced. Song *et al.*[8] picked discrete emotions from the valence-arousal space as the types of emotional expression. Löffler*et al.*[9] selected joy, sadness, fear, and anger as target emotions because these four primary emotions have distinct facial expressions recognized universally, and developmental psychology research indicates that young children can easily recognize anger and fear. Hu *et al.*[10] used Russell's circumplex model of emotions[11], a widely recognized model for emotional states, to convey emotions in their robot.

Expanding Expressive Modalities presents challenges that stem from the need for corresponding hardware support, such as actuators for gestures, screens for visual displays, and speakers for vocal expressions.. Song *et al.*[8] expressed emotions through color, sound, and vibration with an appearance-constrained social Robot. Löffler*et al.*[9] selected color, motion, and sound as the emotion expression modalities, based on their survey of 59 robots where these modalities were the most commonly used. Hu *et al.*[10] used dynamically changing skin textures, with shape (goosebumps and spikes), amplitude, and frequency, to express emotions in robots.

Enhancing Expression Salience is challenging because the abundance and subjectivity of emotional information makes it difficult to define clearly and discretely. Additionally, the subjectivity of interpretation by interaction partners can lead to variability in recognition. Furthermore, it is difficult to quantify and evaluate the effectiveness of expressions accurately. Song *et al.*[8] conducted a structured study to evaluate the effects of three modalities on a human's emotional perception towards their simple-shaped robot "Maru." Gomez *et al.*[12] presented "Haru," an experimental tabletop robot designed to enhance affective potential through anthropomorphic design, mechanical movements, and visual displays.

Perceived Sense of Vitality, Warmth, and Helpfulness faces challenges in aligning expressive behaviors with the robot's physical and functional capabilities, ensuring that its actions remain consistent with its design and technical limitations. Furthermore, humans may develop (overly) high expectations regarding the robot's competence and emotional depth, potentially leading to disappointment or misunderstandings if these expectations are not met[13]. Carpinella *et al.*[14] developed and validates the Robotic Social Attribute Scale (RoSAS), a tool to measure people's perceptions of robots, identifying three key dimensions: warmth, competence, and discomfort.

3.5 State Maintenance

Maintaining these states presents challenges, as it requires clearly defined state spaces and well-established conditions for state transitions to ensure accurate tracking and management of the different states.

Devin *et al.*[15] presented a framework, inspired by the theory of mind[16, 17], where the robot maintains the mental state, including goals, plans, actions, and the environment's state, of both itself and the other agents to ensure fluent execution of shared plans, particularly in response to unexpected situations like human inattention.

3.6 Proactive Reaction

Proactive reactions in affective robots present challenges in adapting responses to dynamic contexts and individual needs, ensuring effective communication while maintaining flexibility in real-time interactions.

Reminder Capability faces a key challenge in ensuring that reminders are effectively delivered to humans. The robot must adapt to individual preferences and contexts, ensuring the reminder is noticed and acknowledged without overwhelming the user.

Report Failures and Seek Assistance presents a challenge in ensuring that these actions are communicated correctly and appropriately. The robot must use clear and understandable language or signals, tailored to the user's needs, to effectively convey the issue and request assistance. Kwon *et al.*[1] framed the accurate expression of incapability and its reasons as a trajectory optimization problem, while adhering to physical limits, improving user perception and collaboration.

Responding autonomously Based on Situational Contexts presents the challenge of handling the complexity of dynamic environments. In such settings, both the states of interaction partners and task parameters form large, multidimensional spaces, making it infeasible to predefine responses for every possible situation. Robots must therefore develop strategies to generalize from limited data and adapt their actions flexibly. Huang *et al.*[3] implemented anticipatory robot control by using predicted user intent and confidence values to proactively plan and execute motions toward predicted targets, adjusting the robot's actions based on the weighted prediction history and confidence thresholds. Görür *et al.*[18] introduced how the robot uses a partially observable Markov decision process to anticipate a human's state of mind, allowing the robot to plan and react proactively during collaboration.

3.7 Safety and Ethics

Challenges in safety and ethics in human-robot interaction include the difficulty of anticipating all possible situations and the high cost of errors.

Lacey *et al.*[19] argued that the "cute" aesthetic in home robots constitutes a dark pattern in HRI by manipulating users' emotions, prioritizing short-term gains, and reducing conscious agency, thereby calling for more ethical design practices in robotics. Jackson *et al.*[20] argued that language-capable robots, if not properly designed to adhere to human moral norms, can inadvertently weaken human moral perceptions, especially through miscommunication in tasks like clarification requests.

4 Conclusion

This article outlines the essential emotional capabilities required for affective robots to interact effectively with humans, including recognition and analysis, emotional strategy, emotional expression, state maintenance, and proactive reaction. These capabilities serve as criteria for evaluating affective robots and contribute to their development for applications in various fields such as healthcare, education, and companionship.

While this article provides a preliminary foundation for affective robot design, future work will focus on refining these criteria to form a more comprehensive and integrated framework. In addition, a more comprehensive survey of the literature will help identify emerging trends and gaps, guiding future research in the field of emotional capabilities in robots.

References

[1] M. Kwon, S. H. Huang, and A. D. Dragan, "Expressing robot incapability," in *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2018, pp. 87–95.

[2] Y. Kato, T. Kanda, and H. Ishiguro, "May i help you? - design of human-like polite approaching behavior-," in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2015, pp. 35–42.

[3] C.-M. Huang and B. Mutlu, "Anticipatory robot control for efficient human-robot collaboration," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016, pp. 83–90.

[4] M. S. Ryoo, T. J. Fuchs, L. Xia, J. K. Aggarwal, and L. Matthies, "Robot-centric activity prediction from first-person videos: What will they do to me?" in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2015, pp. 295–302.

[5] G. Hoffman, O. Zuckerman, G. Hirschberger, M. Luria, and T. Shani-Sherman, "Design and evaluation of a peripheral robotic conversation companion," in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2015, pp. 3–10.

[6] A. Ramachandran, C.-M. Huang, and B. Scassellati, "Give me a break! personalized timing strategies to promote learning in robot-child tutoring," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, 2017, pp. 146–155.

[7] J. Hemminahaus and S. Kopp, "Towards adaptive social behavior generation for assistive robots using reinforcement learning," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, 2017, pp. 332–340.

[8] S. Song and S. Yamada, "Expressing emotions through color, sound, and vibration with an appearance-constrained social robot," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, 2017, pp. 2–11.

[9] D. Löffler, N. Schmidt, and R. Tscharn, "Multimodal expression of artificial emotion in social robots using color, motion and sound," in *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2018, pp. 334–343.

[10] Y. Hu and G. Hoffman, "Using skin texture change to design emotion expression in social robots," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2019, pp. 2–10.

[11] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[12] R. Gomez, D. Szapiro, K. Galindo, and K. Nakamura, ""haru": Hardware design of an experimental tabletop robot assistant," in *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2018, pp. 233–240.

[13] E. Cha, A. D. Dragan, and S. S. Srinivasa, "Perceived robot capability," in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2015, pp. 541–548.

[14] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, “The robotic social attributes scale (rosas): Development and validation,” in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, 2017, pp. 254–262.

[15] S. Devin and R. Alami, “An implemented theory of mind to improve human-robot shared plans execution,” in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016, pp. 319–326.

[16] D. Premack and G. Woodruff, “Does the chimpanzee have a theory of mind?” *Behavioral and Brain Sciences*, vol. 1, no. 4, p. 515–526, 1978.

[17] S. Baron-Cohen, A. M. Leslie, and U. Frith, “Does the autistic child have a “theory of mind” ?” *Cognition*, vol. 21, no. 1, pp. 37–46, 1985. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0010027785900228

[18] O. C. Görür, B. Rosman, F. Sivrikaya, and S. Albayrak, “Social cobots: Anticipatory decision-making for collaborative robots incorporating unexpected human behaviors,” in *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2018, pp. 398–406.

[19] C. Lacey and C. Caudwell, “Cuteness as a ‘dark pattern’ in home robots,” in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2019, pp. 374–381.

[20] R. B. Jackson and T. Williams, “Language-capable robots may inadvertently weaken human moral norms,” in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2019, pp. 401–410.