Fluid-Xpress: Emotion-Aware Dual-Loop Framework for Empathic Facial Reaction in HRI

KANG, Chen <http://orcid.org/0000-0001-8036-4661>, ALITAI, Madina <http://orcid.org/0009-0002-8358-4736>, WANG, Yiting <http://orcid.org/0009-0002-7266-9830>, CAI, Xiaochi <http://orcid.org/0009-0002-6275-7087>, MA, Ruidong <http://orcid.org/0000-0002-8035-5746>, CANGELOSI, Angelo <http://orcid.org/0000-0002-4709-2243> and SHANGGUAN, Zhegong <http://orcid.org/0000-0002-7948-0531>

Citation:

KANG, Chen, ALITAI, Madina, WANG, Yiting, CAI, Xiaochi, MA, Ruidong, CANGELOSI, Angelo and SHANGGUAN, Zhegong (2026). Fluid-Xpress: Emotion-Aware Dual-Loop Framework for Empathic Facial Reaction in HRI. In: HRI Companion '26: Companion Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction. ACM, 425-429. [Book Section]

PDF Download
3776734.3794429.pdf
24 March 2026
Total Citations: 0
Total Downloads: 87

Published: 16 March 2026

Citation in BibTeX format

HRI '26: 21st ACM/IEEE International Conference on Human-Robot Interaction
March 16 - 19, 2026
Scotland, Edinburgh, UK

Conference Sponsors:
SIGAI
SIGCHI

SHORT-PAPER

Fluid-Xpress: Emotion-Aware Dual-Loop Framework for Empathic Facial Reaction in HRI

CHEN KANG, Beijing Language and Culture University, Beijing, China

MADINA ALITAI, Beijing Language and Culture University, Beijing, China

YITING WANG, The University of Manchester, Manchester, Greater Manchester, U.K.

XIAOCHI CAI, Pennsylvania State University, University Park, PA, United States

RUIDONG MA, Sheffield Hallam University, Sheffield, South Yorkshire, U.K.

ANGELO CANGELOSI, The University of Manchester, Manchester, Greater Manchester, U.K.

View all

Open Access Support provided by:

The University of Manchester

Pennsylvania State University

Sheffield Hallam University

Beijing Language and Culture University

HRI Companion '26: Companion Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction (March 2026)
https://doi.org/10.1145/3776734.3794429
ISBN: 9798400723216

Fluid-Xpress: Emotion-Aware Dual-Loop Framework for Empathic Facial Reaction in HRI

Chen Kang[*]
Madina Alitai[*]
School of Information Science
Beijing Language and Culture University
Beijing, China

Yiting Wang
School of Computer Science
University of Manchester
Manchester, United Kingdom

Xiaochi Cai
Pennsylvania State University
Pennsylvania, USA

Ruidong Ma
School of Computing and Digital Technologies
Sheffield Hallam University
Sheffield, United Kingdom

Angelo Cangelosi
Zhegong Shangguan[†]
Cognitive Robotics Lab
Departmenta of Computer Science
University of Manchester
Manchester, United Kingdom
zhegong.shangguan@manchester.ac.uk

Abstract

Large Language Model (LLM)-driven social robots face two key challenges: inference latency creates unnatural silences during which an expressively static robot appears disengaged, and LLMs rarely account for the user's facial affect as a continuous evolving process. We present **Fluid-Xpress**, an emotion-aware dual-loop framework for empathic facial reactions in human-robot interaction. The framework features: (1) a **Macro-Micro dual-loop architecture** that decouples real-time non-verbal feedback from LLM verbal processing, enabling continuous affective backchanneling during inference latency; (2) a **Temporal Affective Engine** using metrics such as MSSD to capture emotional dynamics and detect complex states like cognitive overload and masked emotions; and (3) a **Risk-Adaptive Strategy** that prioritizes immediate intervention during high-arousal states. A pilot study (N=8) showed that Fluid-Xpress significantly improved arousal stability ($p < .05$), mood improvement ($p < .05$), expression awareness ($p < .01$), and perceived empathy ($p < .05$) compared to baseline, providing preliminary support for emotion-aware non-verbal feedback in embodied social robots.

CCS Concepts

• **Computing methodologies** → **Artificial intelligence**; • **Human-centered computing** → **Interactive systems and tools**; *Empirical studies in HCI*.

Keywords

Human-Robot Interaction, Affective Computing, Large Language Models, Non-verbal Behavior, Dynamic Facial Expressions

ACM Reference Format:
Chen Kang, Madina Alitai, Yiting Wang, Xiaochi Cai, Ruidong Ma, Angelo Cangelosi, and Zhegong Shangguan. 2026. Fluid-Xpress: Emotion-Aware Dual-Loop Framework for Empathic Facial Reaction in HRI. In *Companion Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction (HRI Companion '26), March 16–19, 2026, Edinburgh, Scotland, UK*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3776734.3794429

1 Introduction

In human communication, non-verbal cues such as facial expressions, gaze, and head movements carry rich affective information that shapes trust and social connection [5, 19]. For embodied AI and social robots, the capacity to perceive and produce appropriate non-verbal behaviors is foundational to establishing physical co-presence and empathic engagement [8, 20, 21].

Large Language Models (LLMs) have transformed embodied AI by enabling coherent, contextually rich verbal responses, serving as powerful cognitive cores for social robot decision-making [1, 16, 23]. However, effective human-robot interaction (HRI) demands seamless integration of verbal and non-verbal channels. This raises a critical question: *How can social robots generate timely, emotion-aware facial expressions that complement LLM-driven dialogue for empathic interaction?*

Current LLM-based social robots face two key challenges. First, LLM inference latency creates unnatural silences during which a physically present robot that remains expressively static appears disengaged, undermining the user's sense of being understood. Second, LLMs rarely account for the user's facial affect. Human emotion is a continuous process that evolves over time, making it challenging to sample and summarize temporal patterns such as cognitive overload, emotional suppression, or mood shifts.

We propose **Fluid-Xpress**, a multimodal framework that decouples non-verbal feedback from verbal processing for embodied

[*] Both authors contributed equally to this research.
[†] Corresponding Author

HRI Companion '26, Edinburgh, Scotland, UK

ACM ISBN 979-8-4007-2321-6/2026/03
https://doi.org/10.1145/3776734.3794429

Figure 1: The dyadic interaction setup between the participant and the social robot.

social robots. Our core idea is that asynchronous, low-latency affective responses during LLM inference can enhance perceived empathy and engagement. Our contributions include:

- **Dual-Loop Architecture:** Enables real-time non-verbal back-channeling (e.g., empathic mirroring, attentive expressions) during LLM processing.
- **Temporal Affective Engine:** Captures emotional dynamics using metrics such as MSSD to identify states like cognitive overload and masked emotions.
- **Risk-Adaptive Strategy:** Prioritizes rule-based support during high-arousal states to ensure psychological safety.

We conducted an empirical pilot study to evaluate whether Fluid-Xpress improves perceived responsiveness, contextual appropriateness, and user-reported empathy (see Fig. 1).

2 Related Work

Robots with facial expressions are rated as more intelligent and engaging, while also fostering positive reactions in collaborative settings [4, 14]. Facial expressions enhance perceived anthropomorphism, elicit higher empathy and trust, and increase positive arousal in users [2, 6]. Early approaches relied on rule-based mappings from discrete emotions to predefined templates, while recent methods leverage dimensional models like Russell's Circumplex Model [17] for continuous expression generation. However, most systems treat facial expression as a reactive output triggered after verbal response, neglecting the real-time responsiveness essential for natural empathic interaction.

The integration of machine learning [12, 25] and Large Language Models has transformed social robotics to dynamic, context-aware companionship, improving the semantic depth of HRI. Xpress [1] pioneered expression generation by addressing a critical limitation: the need for simultaneity between verbal and non-verbal channels. However, LLM inference introduces a temporal latency gap between immediate non-verbal affect and delayed verbal reasoning, resulting in a frozen robot state during inference that undermines social presence.

We draw inspiration from neurocognitive models: The Dual Systems Model posits that human behavior is governed by the interplay between a socio-emotional system (reactive) and a cognitive control system (deliberative) [22]. Luo and Yu demonstrate that under cognitive overload, individuals lack resources to suppress impulsive affective responses [15]. We apply this duality to HRI: our Fast Loop provides immediate reflexive feedback, while the Slow Loop computes complex verbal interventions.

Effective empathic HRI also requires active emotion regulation. Grounded in Gross's Process Model [10], strategies include attention deployment and cognitive change. Fartook et al. highlight that users in high-arousal states prefer Positive Emotion Regulation over mere empathy [7]. To drive such interventions, we utilize temporal dynamics and the Mean Squared Successive Difference (MSSD) [13] to quantify affective instability beyond static frame analysis.

3 Methodology

Our framework includes four core components, simulating a continuous "listen-feel-react" loop (see Figure 2). The system employs a specific *Engagement Protocol* for cold starts. Upon face detection (t_0), it analyzes the first 12 frames via a *Temporal Analyzer* to categorize the user's facial emotion as "Positive", "Negative", or "Neutral". Then it generates an adaptive greeting emotion based on the response strategy of Xpress [1]. During the user's speech, the system generates simultaneous Speech-to-Text(STT) data and facial frame extraction, ensuring multi-modal input data without loss for other components.

3.1 The Macro-Micro Dual-Loop

During latency, the robot's face is usually set in an initial facial expression. To mitigate the unnatural experience, we implement two concurrent loops: **Micro-Loop (Non-Verbal):** Operating on a regular time sliding window, this loop samples emotional valence/ arousal and triggers immediate facial actuation. Strategies include *Mirroring* (for positive states) [9], *Support* (for negative), and *De-escalation* (for high arousal) based on reference [7]. This ensures the robot remains responsive even before the user finishes speaking. **Macro-Loop (Verbal):** This loop handles complex semantic understanding and verbal response generation via the LLM.

3.2 Deep Affective Computing Engine

Unlike static frame-by-frame recognition, this module maintains a temporal frame queue to compute multidimensional affective metrics that capture the dynamics of user emotional states and arousal trajectories.

We adopt the instability measurements proposed by Jahng et al. [13] to quantify emotional volatility. The Mean Squared Successive Difference (MSSD) captures moment-to-moment variability:

$$\mathrm{MSSD} = \frac{1}{n-1}\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2 \tag{1}$$

where x_i represents the valence or arousal value of the i-th frame. We then derive an Emotional Stability (ES) score that integrates both MSSD and PAC (Probability of Acute Change):

$$\mathrm{ES} = \max\left(0, 1 - \frac{1}{2}\left[\min\left(\frac{\mathrm{MSSD}_c}{M_{\max}}, 1\right) + \min\left(\frac{\mathrm{PAC}_c}{P_{\max}}, 1\right)\right]\right) \tag{2}$$

Composite metrics are computed by averaging across both affective dimensions: $\mathrm{MSSD}_c = \frac{\mathrm{MSSD}(V)+\mathrm{MSSD}(A)}{2}$ and $\mathrm{PAC}_c = \frac{\mathrm{PAC}(V)+\mathrm{PAC}(A)}{2}$, where V denotes valence and A denotes arousal. The normalization constants $M_{\max}$ and $P_{\max}$ represent the maximum MSSD and PAC values derived from a calibration set.

To provide timely emotional support when users experience heightened distress, we implement an abnormal state detection

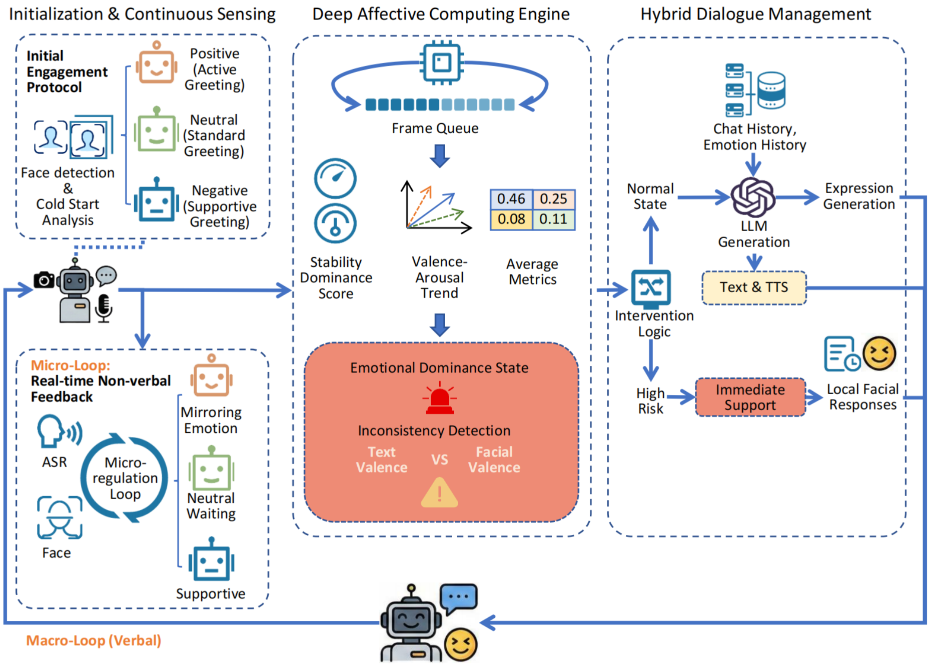
Initialization & Continuous Sensing
Deep Affective Computing Engine
Hybrid Dialogue Management
Initial Engagement Protocol
Positive (Active Greeting)
Neutral (Standard Greeting)
Negative (Supportive Greeting)
Face detection & Cold Start Analysis
Frame Queue
Stability Dominance Score
Valence-Arousal Trend
0.46
0.25
0.08
0.11
Average Metrics
Emotional Dominance State
Inconsistency Detection
Text Valence
VS
Facial Valence
Chat History, Emotion History
Normal State
LLM Generation
Expression Generation
Text & TTS
Intervention Logic
High Risk
Immediate Support
Local Facial Responses
Micro-Loop:
Real-time Non-verbal Feedback
ASR
Micro-regulation Loop
Face
Mirroring Emotion
Neutral Waiting
Supportive
Macro-Loop (Verbal)

Figure 2: The Dual-Loop Architecture. The Micro-Loop (1s in experiment) handles immediate non-verbal cues, while the Macro-Loop manages LLM-based verbal reasoning.

mechanism targeting two high-level states. **Cognitive Overload** is triggered when an increasing arousal trend coincides with low stability scores, indicating the user may be overwhelmed. **Inconsistency Detection** compares semantic sentiment from speech (V_{text}) with facial valence (V_{face}); when divergence exceeds a threshold ($|V_{text} - V_{face}| > 0.5$), the system flags a potential masked emotion (e.g., forced smiling while expressing distress), prompting the dialogue manager to probe deeper with supportive inquiry.

3.3 Hybrid Dialogue Management

We introduce a Risk-Adaptive Decision Tree to balance response speed and emotional depth. Upon detecting the user's affective state, the system routes execution through one of two paths. For **high-risk states** such as cognitive overload or masked emotions, the system bypasses the LLM to deliver low-latency facial responses (e.g., a calming expression accompanied by a breathing guide), preventing further emotional escalation. For **normal states**, the system constructs a prompt containing the user's message, emotional context, and conversation history for LLM processing.

To ensure coherence between local rule-based interventions and LLM-generated responses, we employ prompt augmentation with Response Guidance tokens. When an immediate intervention occurred in the previous turn, the LLM is prompted to follow up contextually (e.g., acknowledging the breathing exercise). Otherwise, the LLM directly generates conversational output via Text-to-Speech synthesis.

4 Pilot Study

We conducted a within-subject pilot study to evaluate whether Fluid-Xpress improves perceived empathy and emotional support in human-robot interaction. We recruited 8 female university students (mean age = 21.3 years) with English proficiency at B2 level or above. All participants were female to reduce gender-related variability in emotional expression and perception. The experiment was conducted in a quiet, enclosed laboratory environment, with psychological counselors available to address any potential emotional crises. All procedures were approved by the institutional ethics review board.

We employed a pre-post measurement design to assess affective states using the Self-Assessment Manikin (SAM) [18], a 9-point pictorial scale measuring valence and arousal, and the Beck Depression Inventory (BDI-13) [3], a 13-item scale assessing psychological states including mood, pessimism, and social interest. Post-interaction, participants rated interaction quality on 5-point Likert scales (1 = strongly disagree, 5 = strongly agree) across seven dimensions shown in Table 1.

Participants completed pre-test questionnaires (SAM and BDI-13), then engaged in two 5-minute conversational sessions: a free-form conversation followed by a recall task discussing recent frustrating experiences. This design evaluated system performance under both neutral and emotionally aroused conditions. Post interaction, participants completed the post-test measures. Each participant experienced both baseline (Xpress) and Fluid-Xpress conditions in counterbalanced order. The system employed Doubao-Seed-1.6-Lite [11] as the conversational LLM, Doubao TTS for speech

Table 1: Interaction Quality Measures

Dimension	Item
Naturalness	The interaction felt like communicating with a real person.
Understanding	The robot understood my emotions.
Appropriateness	The robot responded appropriately to my emotions.
Active Listening	I felt the robot was actively listening to me.
Emotional Relief	Communicating with the robot relieved my emotions.
Expression Awareness	I noticed the robot adjusted its expression based on my facial cues.
Future Use	I would use this robot again or recommend it to others.

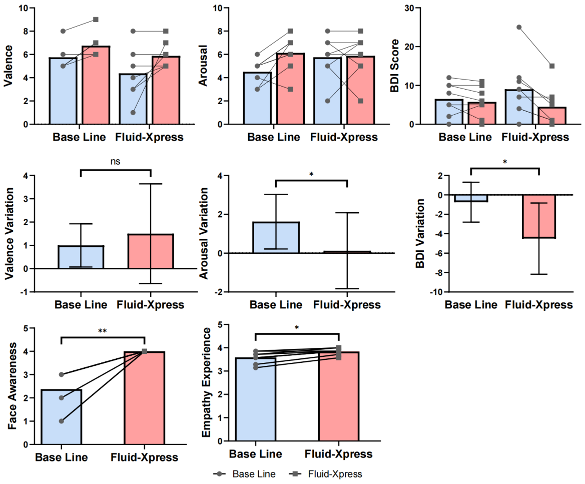

Figure 3: Comparative results across conditions. Asterisks indicate significance: $*p < .05$, $**p < .01$.

synthesis, and EmoNet [24] for facial emotion detection, with calibration parameters $M_{\max} = 0.1$ and $P_{\max} = 0.3$.

5 Results

Figure 3 presents comparative results and we used Wilcoxon matched-pairs tests for analysis (N=8). Both conditions showed pre-post improvements in valence and arousal. While valence variation did not differ significantly between conditions (W = 6.00, p = .69), Fluid-Xpress showed a larger mean improvement (1.50 vs. 1.00), suggesting a trend toward greater emotional enhancement. Arousal variation differed significantly (W = -25.00, p < .05), with Fluid-Xpress producing more stable arousal levels (median difference = -2.00), indicating that real-time non-verbal feedback helps regulate emotional intensity. BDI variation also differed significantly (W = -21.00, p < .05), with Fluid-Xpress showing consistently greater mood improvement (median difference = -4.50).

For interaction quality measures, Expression Awareness showed a highly significant difference (W = 36.00, p < .01), with Fluid-Xpress rated substantially higher than baseline (median difference = 1.00). This confirms that participants clearly perceived the robot's responsive facial expressions during the Micro-Loop. Empathy Experience also improved significantly (W = 28.00, p < .05, median difference = 0.29), with highly effective pairing (Spearman's r = .79, p < .01), suggesting that visible affective responsiveness contributes to perceived empathy.

We also observed scenarios validating the system's behavioral logic. In a **Cognitive Overload** case, a participant expressed: *"You go to travel with her but she suddenly said she won't go and I can't understand her..."* The system detected elevated arousal with low facial stability, triggering Immediate Support with 0.05ms latency. The robot displayed a concerned expression and entered listening mode, effectively de-escalating stress. In a **Masked Emotion** case, a participant stated *"I'm fine, really"* while displaying sadness micro-expressions. The valence discrepancy ($|V_{text} - V_{face}| > 0.5$) triggered the LLM to probe: *"You say you are fine, but you seem a bit down. Do you want to talk about it?"*

6 Discussion

Our pilot study provides preliminary evidence that emotion-aware dual-loop architecture can enhance empathic human-robot interaction, addressing the temporal disconnect between LLM inference and social presence.

The significant improvements in arousal stability and mood align with the Dual Systems Model [22]: the Micro-Loop serves as a reactive system maintaining engagement while the Macro-Loop handles deliberative reasoning. By filling the inference gap with responsive non-verbal feedback, users remain emotionally connected rather than experiencing disengagement. The non-significant difference in valence variation may reflect a ceiling effect, as both conditions improved emotional valence, or may require larger samples to detect subtle differences. The highly significant improvement in Expression Awareness (p < .01) extends previous findings on robot facial expressions for perceived empathy [2, 4]. The observed Cognitive Overload and Masked Emotion cases demonstrate that temporal affective modeling via MSSD [13] captures complex states that static recognition would miss.

Limitations include small sample size (N=8), homogeneous participants, and short interaction duration. Future work should include larger, diverse samples and investigate sustained interactions.

7 Conclusion

This paper presented Fluid-Xpress, an emotion-aware dual-loop framework that decouples non-verbal feedback from verbal processing in LLM-driven social robots. Our pilot study demonstrated significant improvements in arousal stability, mood improvement, expression awareness, and perceived empathy. Future work will extend to diverse populations and longitudinal studies.

8 Acknowledgments

Ethical approval was not required for this study involving non-invasive human–robot interaction with adult participants, and informed consent was obtained. This research was partially supported by the ERC Advanced project "eTALK" (UKRI funded). This research project is also supported by the Fundamental Research Funds for the Central Universities of Beijing Language and Culture University(24QN16).

References

[1] Victor Nikhil Antony, Maia Stiber, and Chien-Ming Huang. 2025. Xpress: A System For Dynamic, Context-Aware Robot Facial Expressions using Language Models. In *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction* (Melbourne, Australia) *(HRI '25)*. IEEE Press, 958–967.

[2] Aryel Beck, Antoine Hiolle, Alexandre Mazel, and Lola Cañamero. 2010. Interpretation of Emotional Body Language Displayed by Robots. In *Proceedings of the 3rd International Workshop on Affective Interaction in Natural Environments*. 37–42.

[3] Aaron Temkin Beck, C Ward, M Mendelson, J Mock, and J Erbauch. 1961. Beck Depression Inventory (BDI).

[4] Cynthia Breazeal. 2003. Emotion and Sociable Humanoid Robots. *International Journal of Human-Computer Studies* 59, 1-2 (2003), 119–155.

[5] Pedro Cárdenas, José García, Rolinson Begazo, Ana Aguilera, Irvin Dongo, and Yudith Cardinale. 2024. Evaluation of robot emotion expressions for human–robot interaction. *International Journal of Social Robotics* 16, 9 (2024), 2019–2041.

[6] Chris Chesher and Fiona Andreallo. 2021. Robotic Faciality: The Philosophy, Science and Art of Robot Faces. *International Journal of Social Robotics* 13, 1 (2021), 83–96.

[7] Ori Fartook, Zachary McKendrick, Tal Oron-Gilad, and Jessica R Cauchard. 2025. Enhancing emotional support in human-robot interaction: implementing emotion regulation mechanisms in a personal drone. *Computers in Human Behavior: Artificial Humans* 4 (2025), 100146.

[8] Laura Fiorini, Federica GC Loizzo, Grazia D'Onofrio, Alessandra Sorrentino, Filomena Ciccone, Sergio Russo, Francesco Giuliani, Daniele Sancarlo, and Filippo Cavallo. 2022. Can I feel you? Recognizing Human's Emotions during Human-robot Interaction. In *International Conference on Social Robotics*. Springer, 511–521.

[9] Vittorio Gallese, Christian Keysers, and Giacomo Rizzolatti. 2004. A unifying view of the basis of social cognition. *Trends in cognitive sciences* 8, 9 (2004), 396–403.

[10] James J Gross. 1998. The emerging field of emotion regulation: An integrative review. *Review of general psychology* 2, 3 (1998), 271–299.

[11] Dong Guo, Faming Wu, Feida Zhu, Fuxing Leng, Guang Shi, Haobin Chen, Haoqi Fan, Jian Wang, Jianyu Jiang, Jiawei Wang, et al. 2025. Seed1. 5-vl technical report. *arXiv preprint arXiv:2505.07062* (2025).

[12] Xiaoxuan Hei, Heng Zhang, and Adriana Tapus. 2023. Robots in education: Influence of Regulatory Focus Theory. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2562–2568.

[13] Seungmin Jahng, Phillip K Wood, and Timothy J Trull. 2008. Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological methods* 13, 4 (2008), 354.

[14] Rachel Kirby, Jodi Forlizzi, and Reid Simmons. 2010. Affective Social Robots. *Robotics and Autonomous Systems* 58, 3 (2010), 322–332.

[15] Jiayi Luo and Rongjun Yu. 2015. Follow the heart or the head? The interactive influence model of emotion and cognition. *Frontiers in psychology* 6 (2015), 573.

[16] Qiaoqiao Ren and Tony Belpaeme. 2025. Touched by ChatGPT: Using an LLM to Drive Affective Tactile Interaction. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 1563–1567.

[17] James A Russell. 1980. A Circumplex Model of Affect. *Journal of Personality and Social Psychology* 39, 6 (1980), 1161.

[18] Zhegong Shangguan, Xiao Han, Younesse El Mrhasli, Nengchao Lyu, and Adriana Tapus. 2025. Factors Influencing Emotional Driving: Examining the Impact of Arousal on the Interplay between Age, Personality, and Driving Behaviors. *Frontiers in Psychology* 16 (2025), 1487493.

[19] Zhegong Shangguan, Xiaoxuan Hei, Fangjun Li, Chuang Yu, Siyang Song, Jianzhuang Zhao, Angelo Cangelosi, and Adriana Tapus. 2025. Learning from Human Conversations: A Seq2Seq based Multi-modal Robot Facial Expression Reaction Framework in HRI. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 7261–7268. doi:10.1109/IROS60139.2025.11247607

[20] Zhegong Shangguan, Yang Liu, Le Song, Tingcheng Li, and Adriana Tapus. 2024. Using a Pneumatic Tactile Steering Wheel to Enhance the Multi-Modal Takeover Request In Smart Vehicle. In *International Conference on Social Robotics*. Springer, Singapore, 122–132.

[21] Jocelyn Shen, Audrey Lee, Sharifa Alghowinem, River Adkins, Cynthia Breazeal, and Hae Won Park. 2025. Social Robots as Social Proxies for Fostering Connection and Empathy towards Humanity. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 989–999.

[22] Elizabeth P Shulman, Ashley R Smith, Karol Silva, Grace Icenogle, Natasha Duell, Jason Chein, and Laurence Steinberg. 2016. The dual systems model: Review, reappraisal, and reaffirmation. *Developmental cognitive neuroscience* 17 (2016), 103–117.

[23] Gemini Robotics Team, Saminda Abeyruwan, Joshua Ainslie, Jean-Baptiste Alayrac, Montserrat Gonzalez Arenas, Travis Armstrong, Ashwin Balakrishna, Robert Baruch, Maria Bauza, Michiel Blokzijl, et al. 2025. Gemini robotics: Bringing AI into the Physical World. *arXiv preprint arXiv:2503.20020* (2025).

[24] Antoine Toisoul, Jean Kossaifi, Adrian Bulat, Georgios Tzimiropoulos, and Maja Pantic. 2021. Estimation of continuous valence and arousal levels from faces in naturalistic conditions. *Nature Machine Intelligence* (2021). https://www.nature.com/articles/s42256-020-00280-0

[25] Chuang Yu, Heng Zhang, Zhegong Shangguan, Xiaoxuan Hei, Angelo Cangelosi, and Adriana Tapus. 2022. Speech-Driven Robot Face Action Generation with Deep Generative Model for Social Robots. In *International Conference on Social Robotics*. Springer, 61–74.

Received 2025-12-08; accepted 2026-01-12