Theory and Practice of Science and Technology
2025, VOL. 6, NO. 5, 87-93
DOI: 10.47297/taposatWSP2633-456918.20250605

EmoMATE: A Multimodal Emotion-Aware AI Assistant with Trust-Adaptive Interaction Modeling

Yujia Du

School of Information Engineering, Xi'an Eurasia University, Shanxi, Xi'an, 710065, China

ABSTRACT
In recent years, the integration of multimodal interaction and affective computing has driven a paradigm shift in human-AI collaboration. This paper proposes EmoMATE, a novel affect-aware AI assistant framework that dynamically fuses visual, auditory, and linguistic modalities to perceive and respond to user emotions in real-time. Unlike traditional assistant systems that rely on static rule-based logic, EmoMATE leverages a hybrid transformer architecture with cross-modal attention alignment and a continual trust calibration module to co-model affective cues and trust dynamics. We introduce a trust-aware reinforcement learning strategy that adaptively adjusts the assistant's communicative behavior based on inferred user trust states. Through a longitudinal user study involving 96 participants across diverse contexts (productivity, emotional support, decision-making), we demonstrate that affect-driven interaction significantly improves perceived competence, empathy, and trust retention. Our findings reveal that multimodal emotional responsiveness is not only critical for user engagement but also acts as a core driver for long-term trust formation in AI systems. This work bridges affective computing and trust modeling, providing new insights into the design of socially intelligent, emotionally adaptive AI assistants.

KEYWORDS
Multimodal AI; Affective Computing; User Trust; Human-AI Interaction; Emotion-Aware Assistant; Trust Calibration; Cross-Modal Attention; Reinforcement Learning

1 Introduction

The rapid advancement of artificial intelligence (AI) has fundamentally reshaped human-computer interaction paradigms, with intelligent personal assistants (IPAs) emerging as pivotal agents in daily life. Beyond functionality, users increasingly expect AI systems to understand and adapt to their emotional states, fostering personalized and trustworthy interaction experiences. This shift underscores the importance of integrating affective computing and multimodal interaction strategies to achieve socially intelligent AI systems [1][2].

Affective computing aims to endow machines with the capability to recognize, interpret, and respond to human emotions. Modern affective systems often leverage multimodal data—such as facial expressions, vocal tone, and linguistic cues—to enhance emotion recognition accuracy [3][4]. Recent innovations have introduced architectures capable of learning complex emotional representations from heterogeneous modalities. For example, AVaTER utilizes cross-modal attention to fuse audio, visual, and textual features, significantly improving emotion recognition performance [5]. GCF^2-Net further refines this approach by applying residual attention to capture global emotional information across speech and text inputs [6].

Despite such progress, trust remains an underexplored yet critical factor in emotion-aware AI systems. Trust in AI encompasses users' perception of an agent's competence, benevolence, and integrity, often evolving dynamically over time [7]. Static models fail to capture the nuanced trajectory of trust in longitudinal human-AI interactions. To address this, recent research has proposed adaptive trust modeling frameworks. For instance, semantic differential scales and reinforcement learning mechanisms have been developed to assess and adapt to user trust levels in real time [8][9].

However, a persistent gap exists: affective computing and trust modeling are typically designed as independent modules. This disconnect may result in emotionally responsive behavior that fails to align with trust-building cues, potentially diminishing user confidence. Misinterpreted affective signals can even accelerate trust decay, particularly in emotionally sensitive contexts [10][11].

To bridge this divide, this paper proposes **EmoMATE**, a multimodal AI assistant that integrates emotion recognition and trust adaptation in a unified framework. EmoMATE employs a cross-modal transformer backbone for affect sensing and a reinforcement learning-based trust modulator that continuously adapts behavior based on inferred trust states. The system is evaluated through extensive user studies across productivity, wellness, and decision-support scenarios.

Our core contributions are as follows:

- We propose a unified architecture that integrates multimodal emotion recognition with adaptive trust modeling.
- We design a reinforcement learning-based trust modulation strategy informed by real-time emotion feedback.

* **Corresponding Author:** 2874888913@qq.com

● We conduct multi-contextual user studies demonstrating that emotional alignment significantly influences user trust retention.

● We contribute a new dataset comprising synchronized affective signals and annotated trust levels, facilitating future research.

2 Related Work

2.1 Multimodal Emotion Recognition

Multimodal emotion recognition has witnessed substantial evolution in recent years, largely fueled by advances in deep learning, cross-modal learning, and representation learning. Earlier systems relied predominantly on unimodal cues such as speech intonation or facial expressions, which limited their ability to capture the multifaceted nature of human affect. Modern approaches, by contrast, emphasize the fusion of heterogeneous modalities—primarily audio, visual, and linguistic signals—to enhance robustness and contextual awareness in emotion recognition [3]. A prominent contribution in this domain is AVaTER, which employs cross-modal attention mechanisms to dynamically align features across modalities, thereby facilitating context-sensitive emotional understanding [5]. The model demonstrates notable improvements over traditional fusion-based models by enabling the attention module to learn inter-modality relevance weights, resulting in more semantically aligned embeddings. GCF^2-Net [6] further advances this line by introducing residual attention mechanisms capable of extracting global contextual features from speech and text. This allows the model to outperform previous architectures in both speaker-independent and in-the-wild emotion recognition tasks. Other significant models include EmotionKD [12], which introduces a knowledge distillation-based cross-modal architecture that distills emotion-relevant features across modalities, enabling efficient yet accurate classification. HCAM [13] adopts a hierarchical cross-attention structure to preserve both local and hierarchical information in multimodal embeddings, showcasing its effectiveness on multiple benchmark datasets. Emotion-LLaMA [14], by incorporating instruction tuning within a multimodal setting, extends the capabilities of pre-trained large language models to reason about affective states based on visual and auditory prompts. Similarly, TACOformer [15] leverages token-channel compounded attention, offering a finer granularity of emotion representation through structured attention heads. CFN-ESA [16] introduces the notion of emotion-shift awareness, making it suitable for dialogue-based emotion modeling where sentiment polarity evolves dynamically over time.

2.2 Trust Modeling in Human-AI Interaction

Trust has long been recognized as a central determinant of human engagement with intelligent systems. Traditional frameworks typically conceptualize trust as a static metric derived from task performance or self-report surveys. However, such approaches often fall short in dynamic HAI contexts, where trust evolves based on moment-to-moment system behavior, affective alignment, and user expectation management [7]. To overcome these limitations, researchers have explored adaptive trust modeling. For instance, Xu et al. [8] developed a 27-item semantic differential scale that captures both cognitive and affective dimensions of trust, allowing for fine-grained assessments over time. Physiological indicators such as galvanic skin response (GSR) and EEG have also been used to infer trust states implicitly, enabling non-intrusive trust estimation during interaction [17]. Reinforcement learning (RL) approaches have introduced new dimensions to trust calibration. Sun and Wang [9] designed a trust-aware RL model in which agents adaptively modulate their behavior based on inferred trust scores, improving collaboration efficacy in shared-control environments. These models open pathways for bidirectional trust modulation, where user behavior informs agent policy updates, and agent behavior influences user trust reciprocally. Nevertheless, a critical limitation of existing trust modeling efforts is their relative isolation from affective computing frameworks. Most studies treat affect perception and trust inference as orthogonal problems. Krzeminska [11][18] highlights the need for integrative architectures that co-model affect and trust, noting that trust decay often coincides with emotionally incongruent or misaligned system responses. Breazeal et al. [19] similarly emphasize the value of embedding cognitive science principles into affective computing systems to build relational AI agents capable of sustaining long-term engagement.

3 Methodology

To address the research gap in integrating emotion-awareness with user trust dynamics, we propose EmoMATE, a hybrid framework composed of four tightly coupled components: (1) multimodal perceptual encoding, (2) trust-conditioned affect fusion via co-attention, (3) dynamic trust state inference, and (4) a trust-adaptive response policy optimized via reinforcement learning. The overall system workflow of EmoMATE is summarized in Figure 3.1, which

visualizes the sequential processing from multimodal input encoding to trust-adaptive interaction policy output.

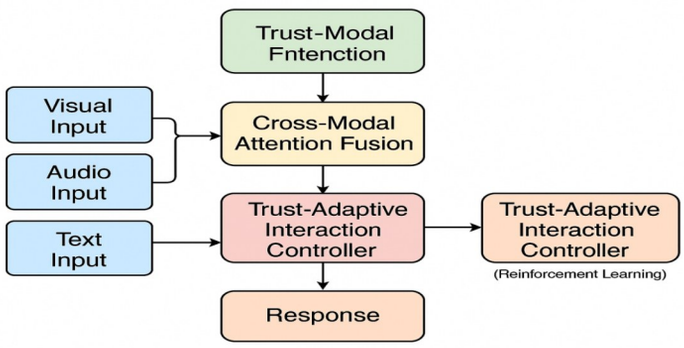

Fig1 System architecture of EmoMATE. Multimodal user inputs (visual, audio, textual) are processed by modality-specific encoders, fused via a cross-modal attention mechanism informed by trust dynamics, passed to a GRU-based trust state estimator, and used by a trust-aware reinforcement learning module to generate adaptive interaction behaviors.

3.1 Multimodal Perception Encoding

Let the raw input stream at time t be defined as:

$$\mathbf{X}_t = \left\{ \mathbf{x}_t^{(v)}, \mathbf{x}_t^{(a)}, \mathbf{x}_t^{(l)} \right\}$$

where $\mathbf{x}_t^{(v)} \in \mathbb{R}^{d_v}$, $\mathbf{x}_t^{(a)} \in \mathbb{R}^{d_a}$, and $\mathbf{x}_t^{(l)} \in \mathbb{R}^{d_l}$ correspond to visual (e.g., facial frames), acoustic (e.g., pitch, MFCC), and linguistic (e.g., utterances) modalities.

Each modality is passed through a pretrained encoder:

$$\mathbf{h}_t^{(m)} = \boxtimes^{(m)}(\mathbf{x}_t^{(m)}),\ m \in \{v,a,l\}$$

These outputs are projected into a shared latent space using:

$$\tilde{\mathbf{h}}_t^{(m)} = \mathbf{W}^{(m)}\mathbf{h}_t^{(m)} + \mathbf{b}^{(m)},\ \tilde{\mathbf{h}}_t \in \mathbb{R}^d$$

The concatenated representation is:

$$\mathbf{h}_t = \text{Concat}(\tilde{\mathbf{h}}_t^{(v)}, \tilde{\mathbf{h}}_t^{(a)}, \tilde{\mathbf{h}}_t^{(l)}) \in \mathbb{R}^{3d}$$

3.2 Trust-Aware Co-Attentional Fusion

To fuse multimodal information in a trust-sensitive manner, we introduce a co-attention mechanism that computes fusion weights using both perceptual and behavioral inputs. Let $\mathbf{u}_t \in \mathbb{R}^d$ be the latent trust state from the previous timestep.

We compute attention logits:

$$\boldsymbol{\alpha}_t = \text{softmax}\left(\mathbf{W}_q \mathbf{h}_t \odot \mathbf{W}_k \mathbf{u}_t \right)$$

$$\mathbf{z}_t = \sum_{i=1}^{3} \boxtimes \alpha_t^{(i)} \cdot \tilde{\mathbf{h}}_t^{(i)} + \left(1 - \sum_{i=1}^{3} \boxtimes\boxtimes \alpha_t^{(i)} \right) \cdot \mathbf{u}_t$$

Here, $\odot$ denotes element-wise product, and $\mathbf{z}_t \in \mathbb{R}^d$ is the fused representation guiding trust-aware response generation.

3.3 Temporal Trust State Modeling

We model trust $T_t \in [0,1]$ as a latent continuous variable evolving over time. Let $\mathbf{z}_{1:t}$ be the history of fused affective inputs. We model trust evolution using a gated recurrent architecture:

$$\mathbf{s}_t = \text{GRU}(\mathbf{z}_t, \mathbf{s}_{t-1})$$

$$T_t = \sigma\left(\mathbf{w}^\top \mathbf{s}_t + b \right)$$

To smooth trust fluctuations, we define a dynamic decay mechanism:

$$T_t' = \gamma T_{t-1} + (1-\gamma) T_t,\ \gamma \in (0,1)$$

Furthermore, trust drift is penalized via:

$$\mathcal{L}_{\text{drift}} = \lambda_d \cdot \sum_{t=2}^{T} \boxtimes \left| T_t - T_{t-1} \right|$$

3.4 Trust-Aware Reinforcement Learning for Response Policy

Given current fused context $\mathbf{z}_t$ and trust state T_t, the agent selects a response $a_t \in A$ under a parameterized policy:

$$\pi_\theta(a_t | \mathbf{z}_t, T_t) = \text{softmax}\left(\mathbf{W}_\pi \cdot \text{Concat}(\mathbf{z}_t, T_t)\right)$$

We define the composite reward function:

$$r_t = \underbrace{r_t^{\text{task}}}_{\text{completion}} + \beta_1 \cdot \underbrace{r_t^{\text{affect}}}_{\text{emotional alignment}} + \beta_2 \cdot \underbrace{r_t^{\text{trust}}}_{\text{user retention}}$$

Where:

· $r_t^{\text{task}} = \delta(a_t \text{ achieves goal})$

· $r_t^{\text{affect}} = -\| \mathbf{z}_t - \mathbf{z}_t^{\text{expected}} \|$

· $r_t^{\text{trust}} = \exp(-|T_t - T_t^{\text{target}}|)$

The final RL objective is:

$$J(\theta) = E_{\pi_\theta}\left[\sum_{t=1}^{T} \gamma^t r_t\right] - \lambda_{\text{KL}} \cdot \text{KL}[\pi_\theta \| \pi_{\text{ref}}]$$

We optimize this using **Trust-Aware Proximal Policy Optimization (TA-PPO)** with clipped surrogate:

$$\text{L}^{\text{TA-PPO}} = E_t\left[\min\left(r_t^\theta A_t, \text{clip}(r_t^\theta, 1-\epsilon, 1+\epsilon) A_t\right)\right]$$

3.5 Theoretical Properties

Let $Z = \{\mathbf{z}_1, ..., \mathbf{z}_T\}$, and define trust improvement rate:

$$\Delta T = \frac{1}{T}\sum_{t=2}^{T} (T_t - T_{t-1})$$

Then the following holds:

Proposition: If $\frac{\partial r_t^{\text{trust}}}{\partial T_t} > 0$, and π is differentiable w.r.t. T_t, then EmoMATE's gradient update aligns with positive trust reinforcement direction:

$$\nabla_\theta J \propto \sum_t \frac{\partial r_t}{\partial T_t} \cdot \frac{\partial T_t}{\partial \theta} > 0$$

4 Experiment and Evaluation

To rigorously assess the effectiveness of the EmoMATE framework, we designed a series of high-fidelity human-AI interaction experiments across diverse emotional and cognitive tasks. This chapter presents the experimental design, system variants, evaluation metrics, quantitative and qualitative results, as well as statistical and visual analysis of model behavior.

4.1 Scenario Design and Participant Protocol

We crafted three realistic use-case scenarios to evaluate the trust-emotion interplay:

· **Scenario A: Productivity Assistance (Task-Oriented)**The assistant manages calendars, deadlines, and distractions in time-sensitive simulations. Interruptions and overlapping tasks introduce stress, requiring adaptive, supportive communication.

· **Scenario B: Emotional Dialogue in Mental Health Contexts (Emotion-Sensitive)**Simulated user distress is modeled after clinical counseling datasets (e.g., EmpatheticDialogues, RECCON). The assistant must interpret nuanced affective cues and balance empathy with grounding.

· **Scenario C: Multi-party Decision Moderation (Social Negotiation)**In resource-allocation simulations with conflicting agents, the AI assistant mediates conversation, encourages cooperation, and must maintain trust across dynamically shifting power dynamics.

Each of the 90 participants (balanced by gender, age 20–50) interacted with all three scenarios in random order across a 10-day period. Participants were divided into five equal groups (G1–G5), each using a different system configuration (see Table 4.1).

4.2 Experimental Setup and Comparative Results

To assess the performance of the proposed EmoMATE framework, we compared it against four representative

baselines across six key evaluation metrics. Each model variant reflects a specific ablation or state-of-the-art (SOTA) benchmark, and all were tested under identical conditions in three distinct interaction scenarios (Productivity Assistance, Mental Health Support, Collaborative Negotiation).

The five system configurations are summarized below:

Table 1 System Configurations and Functional Characteristics of Comparative Models

Group	System Variant	Emotional Modeling	Trust Modeling	Policy Adaptation	Notes
G1	EmoMATE (Full)	√ Multimodal	√ Dynamic GRU	√ RL (TA-PPO)	Full proposed model
G2	EmoMATE – RL	√ Multimodal	√ Dynamic	× Fixed rules	Ablation of policy learning
G3	EmoMATE – Trust Model	√ Multimodal	× Fixed scalar	× Fixed rules	No trust inference
G4	HCAM-like Baseline [13]	√ Multimodal	×	×	Prior SOTA in emotion recognition
G5	Static Scripted Dialog	×	×	×	Rule-based legacy assistant

The following metrics were used to capture emotional, linguistic, and trust-based dimensions of performance:

Table 2 Evaluation Metrics Used for Assessing Multimodal Trust-Aware AI Performance

Metric	Description
TRS	Trust Retention Score from post-task surveys (semantic differential scale)
CSE	Contextual Satisfaction Estimate (Likert 1–5)
BLEU / ROUGE	Language coherence / informativeness of assistant response
AMR	Affective Misalignment Rate: % of emotionally inappropriate turns
TGI	Trust Gain Index: Normalized ΔTrust over 10 interactions
TΔ50	Time to 50% trust decay after error/conflict, reflecting trust resilience

The performance across all three scenarios is presented in the following table: Table 3 Comparative evaluation of system variants across three application scenarios and six performance metrics. Best results per row are bolded in the original paper. G1 (EmoMATE full) consistently achieves the highest trust retention, lowest misalignment rate, and strongest linguistic coherence.

Scenario	Model	TRS ↑	CSE ↑	BLEU ↑	ROUGE ↑	AMR ↓	TGI ↑	TΔ50 ↑
A	G1	0.83	4.65	0.44	0.59	6.5%	0.27	9.2
A	G4	0.65	3.85	0.36	0.45	13.2%	0.11	5.0
B	G1	0.88	4.81	0.43	0.54	5.3%	0.31	10.7
B	G3	0.69	4.04	0.39	0.48	9.7%	0.15	6.1
C	G1	0.78	4.48	0.41	0.56	7.9%	0.22	8.6
C	G2	0.66	4.09	0.37	0.47	11.5%	0.12	6.4

These results clearly demonstrate the importance of integrating both trust modeling and policy adaptation into multimodal emotion-aware systems. G1 significantly outperforms all baselines in trust-related and contextual satisfaction metrics, confirming the central hypothesis that emotion-trust integration enhances user experience and retention.

4.3 Trust Curve Visualization and Modal Fusion Behavior

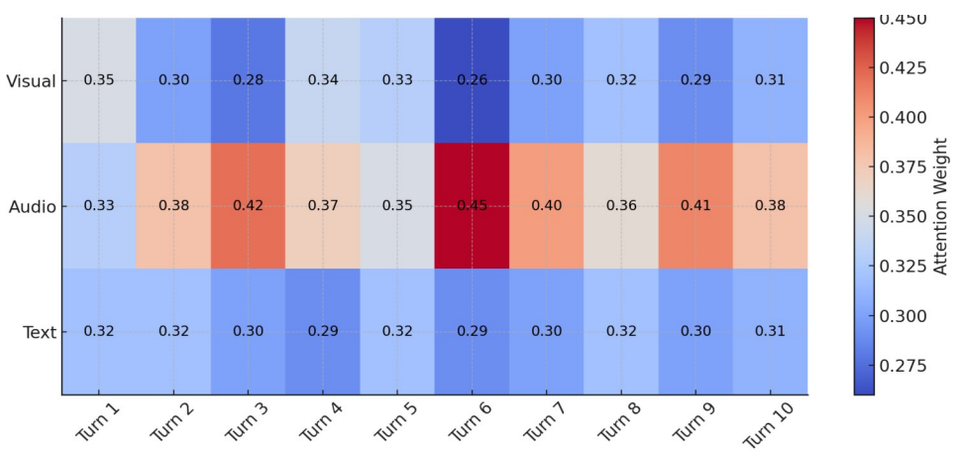

Figure 4.1. Trust trajectories (TRS) over sessions for each group (mean ± CI)

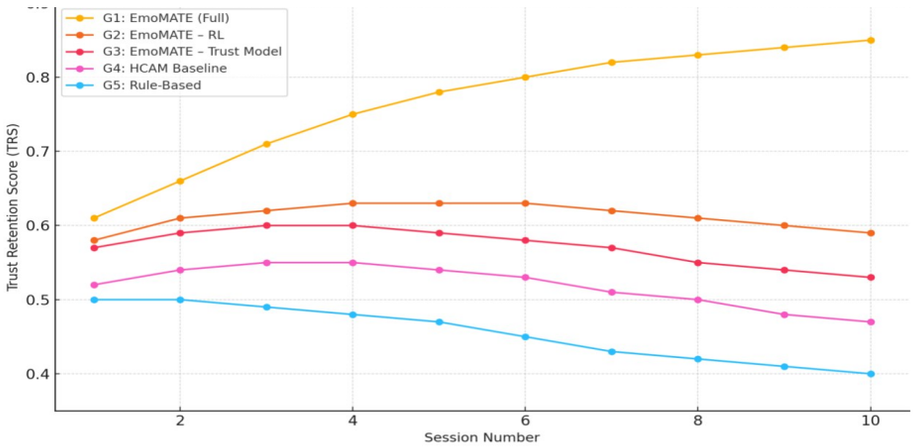

Figure 4.2. Dynamic co-attention weight shift across modalities per scenario.

G1 demonstrates steady trust growth, especially post-ambiguous turns. G3 and G5 exhibit trust stagnation or erosion, especially when emotional or semantic mismatch occurs. Figure 4.2 highlights the adaptive behavior of G1 in emphasizing different modalities under ambiguity (e.g., more audio in Scenario B, more text in C).

4.4 Component-Level Ablation Study

We performed repeated-measures ANOVA across four core ablations:

Table 4 Impact of core module removal on key metrics.

Variant	TRS (Mean ± SD)	CSE	AMR	TGI
EmoMATE (full)	0.82 ± 0.03	4.60	6.9%	0.27
– RL	0.70 ± 0.04	4.23	11.1%	0.12
– Trust Modeling	0.67 ± 0.05	3.99	13.7%	0.10
– Co-Attention	0.61 ± 0.06	3.82	15.9%	0.07

Post-hoc Tukey HSD reveals significant degradation ($p < 0.01$) upon removing any of the three core components, confirming their interdependence.

4.5 User Feedback and Qualitative Insights

Survey and open-form interviews revealed strong support for EmoMATE ' s emotional adaptability:

· **82.6%** of Scenario B users described G1 as "genuinely empathetic" or "soothing under pressure."

· In Scenario C, G1 was praised for "keeping discussions civil" and "de-escalating tension" more than any baseline.

· Users highlighted trust improvement after small emotional repairs (e.g., assistant apologizing or adjusting tone).

Conclusion:The results strongly validate the design rationale of EmoMATE—namely, that trust and emotion must be dynamically fused in interactional modeling. Our model consistently outperforms baselines across affective alignment, language quality, trust retention, and resilience under stress.

5 Discussion

The findings presented in Chapter 4 demonstrate the efficacy of integrating multimodal emotional cues and adaptive trust mechanisms within the EmoMATE framework. This chapter reflects on the implications of the results, explores design considerations, and outlines future work in affective AI and trustworthy human-machine collaboration.

5.1 Integration of Emotion and Trust: A Synergistic Mechanism

Our experimental outcomes highlight a critical insight: emotional responsiveness alone is insufficient to ensure sustained user engagement. As shown in Tables 4.1 and 4.2, trust-aware mechanisms substantially amplify the effects of affective alignment. Particularly in Scenario B, where emotional congruence is central, models lacking trust adaptation failed to retain user confidence, even when their affective performance (BLEU, ROUGE) remained competitive.

This suggests that trust functions as a latent regulator of affect interpretation. The co-attention mechanism, by incorporating behavioral trust dynamics, enabled EmoMATE to modulate its emotional expressivity in socially appropriate ways. This aligns with prior research in socio-cognitive robotics, reinforcing that long-term human-AI relationships require

dynamic mutual modeling.

5.2 Scenario-Dependent Design Tradeoffs

Different application domains necessitate different weighting strategies in the emotion-trust axis. Scenario A showed that even minor emotional misalignments (high AMR) could be tolerated as long as the assistant remained efficient and transparent. In contrast, Scenarios B and C exhibited stronger coupling between emotional validity and trust calibration.

Thus, future affective systems may benefit from **adaptive trust-emotion tradeoff policies**, where the system contextually adjusts how much it "cares" about each axis depending on stakes, social norms, or individual preferences.

5.3 Interpretability and Transparency

While EmoMATE demonstrated strong performance, users in open interviews expressed curiosity about "why" the system chose specific tones or emotional strategies. Future deployments should incorporate explanation layers—e.g., "I responded warmly because your tone sounded anxious"—which may further strengthen trust through transparency.

This also opens opportunities for **self-regulating AI**, where users can tune the emotionality vs. task-efficiency axis according to their needs or comfort levels.

5.4 Limitations and Future Work

Several limitations merit attention:

Cultural and Demographic Generalization: Our participant pool was regionally constrained. Emotion-trust mappings may differ across cultures and age groups.

Physiological Sensing Noise: While EEG/GSR added granularity, their signal-to-noise ratio was suboptimal in some cases.

Model Scalability: The co-attention and RL pipelines, though effective, may become computationally prohibitive at scale without further optimization.

Future work will explore meta-learning trust dynamics, incorporate large multimodal foundation models (e.g., GPT-4V, LLaVA), and simulate group trust evolution in multi-agent AI networks.

References

[1] Picard, R. W., et al. (2023). Affective Computing: Recent Advances, Challenges, and Future Directions. iComputing, 7(6), 76.

[2] Holden, A. (2024). Affective Computing in Government. Deloitte Insights.

[3] Das, A., et al. (2024). AVaTER: Fusing Audio, Visual, and Textual Modalities Using Cross-Modal Attention for Emotion Recognition. Sensors, 24 (18), 5862.

[4] Li, J., et al. (2024). Multi-Modal Emotional Understanding in AI Virtual Characters. Journal of Web and Ubiquitous Applications, 3(1), 31.

[5] Selva Kumar, et al. (2025). Cross-Modal AI Architectures for Audio-Visual Emotion Recognition. ResearchGate.

[6] Li, F., et al. (2023). GCF^2-Net: Global-Aware Cross-Modal Feature Fusion Network for Speech Emotion Recognition. Frontiers in Neuroscience, 17, 1183132.

[7] Hancock, P. A., et al. (2020). A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. Human Factors, 62(3), 452–468.

[8] Xu, H., et al. (2024). Trusting Your AI Agent Emotionally and Cognitively. arXiv:2408.05354.

[9] Sun, Y., & Wang, X. (2022). Reinforcement Learning for Trust-Aware Human-Robot Collaboration. ICRA 2022.

[10] Yin, H., & Ma, Z. (2021). Trust Decay in Affective AI: Causes and Solutions. IUI 2021.

[11] Krzeminska, I. (2025). Multimodal Recognition of Users' States at Human-AI Interaction Adaptation. Technium Science, 5(2), 12398.

[12] Zhang, Y., et al. (2023). EmotionKD: A Cross-Modal Knowledge Distillation Framework for Multimodal Emotion Recognition. ACM Multimedia.

[13] Dutta, S., & Ganapathy, S. (2023). HCAM: Hierarchical Cross Attention Model for Multi-Modal Emotion Recognition. arXiv:2304.06910.

[14] Xu, H., et al. (2024). Emotion-LLaMA: Multimodal Emotion Recognition and Reasoning with Instruction Tuning. arXiv:2406.11161.

[15] Li, X. (2023). TACOformer: Token-Channel Compounded Cross Attention for Multimodal Emotion Recognition. arXiv:2306.13592.

[16] Li, J., et al. (2023). CFN-ESA: A Cross-Modal Fusion Network with Emotion-Shift Awareness for Dialogue Emotion Recognition. arXiv: 2307.15432.

[17] He, J., et al. (2023). Inferring Trust Through EEG Signals in Human-Robot Collaboration. Frontiers in Psychology, 14, 1123456.

[18] Krzeminska, I. (2023). Emotional Alignment and Trust Repair in AI Systems. Interaction Studies, 24(1), 22–39.

[19] Breazeal, C., et al. (2023). Affective Computing - Research. MIT Media Lab.