MER-R1: MULTIMODAL EMOTION REASONING VIA SLOW-FAST THINKING SYNERGY

Anonymous authors
Paper under double-blind review

ABSTRACT

We find that explicit reasoning does not necessarily translate into better multimodal emotion recognition (MER) accuracy, even though it makes predictions more interpretable. Specifically, for reasoning-based MLLMs, *fast thinking* by triggering direct answers often outperforms *slow thinking* after deliberative reasoning. Our empirical analyses show that fast thinking improves recall with broader and more confident predictions, whereas slow thinking favors precision through conservative filtering of incorrect categories. Building on these insights, we propose `MER-R1`, a reinforcement learning framework that turns slow-fast complementarity into explicit optimization. Dual-objective disentanglement separates recall and precision into two optimization signals, allowing them to be jointly optimized rather than traded off against each other. Slow-fast confidence calibration further aligns the final slow-thinking answer with fast-thinking intuition, strengthening correct emotions while suppressing incorrect ones. In this way, `MER-R1` unifies the recall-oriented intuition of fast thinking with the precision-oriented selectivity of slow thinking. We further provide theoretical justification for this synergy, showing that it mitigates variance-induced interference during optimization. Extensive experiments on MER-UniBench and MME-Emotion show that `MER-R1` achieves state-of-the-art performance and makes reasoning genuinely benefit emotion recognition. Code is provided in the Supplementary Materials.

1 INTRODUCTION

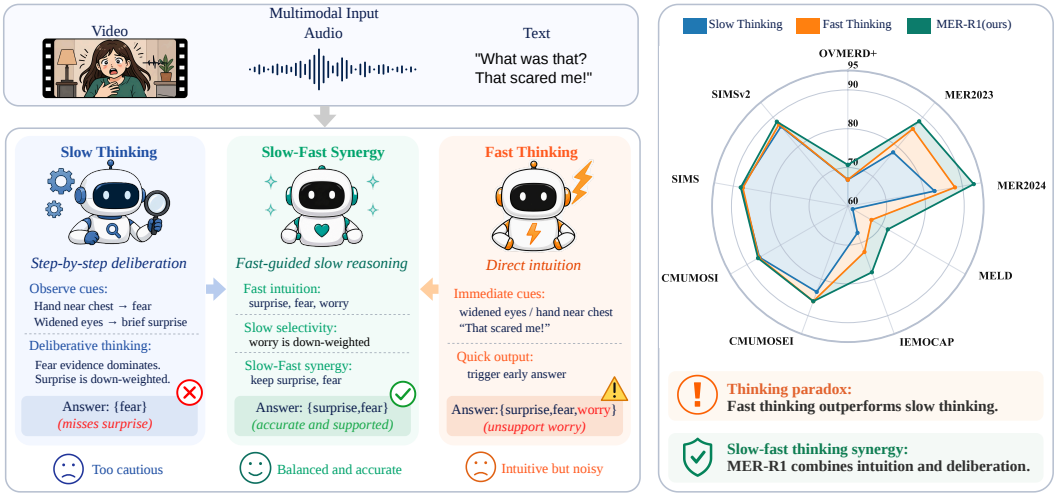
Recent advances in Multimodal Large Language Models (MLLMs) (Hurst et al., 2024; Singh et al., 2025; Xu et al., 2025a) are transforming affective computing (Picard, 2000; Lian et al., 2025b; Shou et al., 2025), shifting multimodal emotion recognition (Lian et al., 2023a; 2024; Poria et al., 2019; Busso et al., 2008) from fixed-label prediction to *Open-Vocabulary Multimodal Emotion Recognition* (OV-MER) (Lian et al., 2025a;c), where models recognize diverse and open-ended human emotions. To improve trustworthiness (Lian et al., 2023b), recent RLVR-based approaches (Guo et al., 2025; Lian et al., 2026; Yang et al., 2025) further equip emotion MLLMs with explicit reasoning chains that ground emotion predictions in visual, acoustic, and textual cues.

Despite this promise, we find that such reasoning does not necessarily lead to stronger emotion recognition. Specifically, we consider two thinking modes for an R1-like emotion reasoning model (Lian et al., 2026): **(1) Slow thinking** produces an answer after standard reasoning, and **(2) Fast thinking** elicits an answer directly without reasoning. Despite well-formed reasoning traces (Figure 1 (a)), slow thinking consistently underperforms fast thinking on current OV-MER benchmarks such as MER-UniBench (Lian et al., 2025a;c) (Figure 1 (b)). This reveals a *thinking paradox* in MER:

Reasoning improves interpretability, yet fails to improve recognition accuracy itself.

To understand this paradox, we analyze fast and slow thinking in Section 3 from prediction and confidence perspectives. At the prediction level, fast thinking exhibits stronger System-1 intuition (Zhang et al., 2025), producing broader emotion coverage and higher recall, whereas slow thinking makes more concentrated predictions and favors precision. At the confidence level, fast thinking is more confident on correct emotions, while slow thinking more conservatively suppresses incorrect categories but may also lower confidence in correct ones. These findings suggest two **slow-fast synergy**

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) Motivating example (b) Evaluation summary on **9** datasets

Figure 1: **Thinking paradox and slow-fast synergy in MER.** (a) Motivating example: slow thinking is cautious but may miss valid emotions, whereas fast thinking improves coverage but introduces noise. Slow-fast synergy combines their complementary strengths. (b) Evaluation summary on 9 datasets: Fast thinking counter-intuitively outperforms slow thinking.

desiderata for effective MER reasoning: preserving the recall-oriented coverage and correct-category confidence of fast thinking, while retaining the precision-oriented selectivity of slow thinking.

To realize these desiderata, we propose `MER-R1`, a reinforcement learning (RL) framework with two components. At the prediction level, *dual-objective disentanglement* separates recall and precision into two objectives and preserves them in both reward and advantage spaces. Unlike prior MER reasoning methods that directly optimize F_1 reward (Lian et al., 2026), our design prevents recall and precision interference during advantage normalization. Our theoretical analysis further shows that standard F_1 optimization can be biased toward the objective with larger group-wise variation, whereas the dual-objective advantage establishes a balanced coupling to both recall and precision signals. At the confidence level, *slow-fast confidence calibration* transfers fast thinking's confidence to the slow-thinking answer while preserving slow thinking's selectivity. Specifically, we compare category-level confidence between the two thinking modes and calibrate correct and incorrect categories in opposite directions: correct categories are encouraged to retain or exceed fast-thinking confidence, whereas incorrect categories are encouraged to remain suppressed. This enables `MER-R1` to strengthen correct emotions without absorbing the noisy over-coverage of fast thinking.

Our contributions are three-fold: **(1) Findings:** We reveal a *thinking paradox* in MER: for reasoning-based MLLMs, fast thinking can outperform slow thinking. We diagnose this paradox from prediction and confidence perspectives and derive two desiderata for slow-fast synergy. **(2) Methodology:** We propose `MER-R1`, an RL framework with dual-objective disentanglement and slow-fast confidence calibration, enabling joint recall–precision optimization and confidence calibration over correct and incorrect categories. **(3) Performance:** `MER-R1` achieves state-of-the-art results on MER-UniBench and MME-Emotion, and validates that satisfying the two slow-fast synergy desiderata makes reasoning genuinely beneficial for emotion recognition rather than merely improving interpretability.

2 Related Work

MLLM reasoning. Recent multimodal large language models (MLLMs) (Hurst et al., 2024; Xu et al., 2025b;a) have acquired reasoning ability through reinforcement learning with verifiable rewards (RLVR) (Guo et al., 2025; Zhang et al., 2025), allowing them to explicitly reason over visual, acoustic, and textual signals during the reasoning process. However, recent studies suggest that slow thinking is not always better than fast thinking. VideoAuto-R1 (Liu et al., 2026) observes this phenomenon on video understanding tasks and addresses it with an *answer-think-answer* paradigm that rewards both the early and final answers while assigning a larger weight to the latter. However, this design

remains at the output level and does not model why fast and slow thinking differ, let alone how to realize synergy between them. In contrast, our work studies their underlying complementarity and turns it into an explicit synergy framework for multimodal reasoning.

Multimodal emotion understanding. Multimodal emotion understanding has recently shifted from closed-set classification to more open and reasoning-oriented settings (Lian et al., 2023b; Cheng et al., 2024; Han et al., 2025; Lian et al., 2025a). Open-Vocabulary Multimodal Emotion Recognition (OV-MER) (Lian et al., 2025c) requires models to predict free-form emotion words beyond a fixed label space, while MME-Emotion (Zhang et al., 2026) further evaluates both recognition and reasoning ability with LLM-based judges. Recent emotion MLLMs (Zhao et al., 2025; Yang et al., 2025; Lian et al., 2026; Rha et al., 2026) increasingly introduce explicit reasoning to enhance interpretability and reinforcement learning to improve emotion recognition. A central objective, therefore, is to **improve emotion recognition while retaining the interpretability afforded by reasoning**; yet whether reasoning itself genuinely enhances recognition remains unclear. We identify and systematically study the *thinking paradox* in multimodal emotion reasoning, where fast thinking can outperform slow thinking on recognition, and resolve it through an explicit slow-fast thinking synergy framework.

3 Understanding the Thinking Paradox

This section investigates the counter-intuitive finding that fast thinking outperforms slow thinking in MER tasks (Figure 1 (b)). We show that fast thinking improves recall with more diverse and confident predictions, whereas slow thinking is more conservative, favoring precision by suppressing incorrect categories but lowering confidence in correct ones. Under current evaluation, which mainly relies on recall or F1 metrics, the advantage of slow thinking is not fully reflected. We begin by introducing the emotion reasoning formulation and the emotion-wheel-based evaluation protocol.

3.1 Problem Formulation and Emotion Wheel-Based Evaluation

Given a multimodal input x, the policy π_θ generates an output $y = (y^{\text{cot}}, y^{\text{ans}})$, where y^{cot} denotes the chain-of-thought reasoning trace and y^{ans} denotes the final emotion answer. Slow thinking follows the standard think-then-answer format, whereas fast thinking keeps the user instruction unchanged but uses an assistant-side empty-thought prefill, `<think></think><answer>`, to directly generate y^{ans}, yielding $y^{\text{cot}} = \emptyset$ (Jiang et al., 2025). Since both the model answer y^{ans} and the ground-truth emotions y^{gt} may contain multiple free-form emotion words, we follow prior work (Lian et al., 2025a;c) and map them through the emotion wheel to level-1 categories. For example, as shown in Figure 2, the **innermost sectors define level-1 emotion categories**, such as `angry`, `happy`, and `sad`. Fine-grained emotion words in the outer sectors are then mapped to these categories; for instance, `joyful`, `content`, and `cheerful` are mapped to the level-1 category `happy`. Further details on all five emotion wheels are provided in Appendix B.1.



Figure 2: Emotion wheel.

Let $\hat{\mathcal{Y}}$ and $\mathcal{G}$ denote the predicted and ground-truth level-1 category sets obtained from y^{ans} and y^{gt}, respectively. We evaluate the prediction by set-level matching. Specifically, recall, precision, and F_1 are defined as: $\mathsf{R} = \frac{|\hat{\mathcal{Y}} \cap \mathcal{G}|}{|\mathcal{G}|}, \mathsf{P} = \frac{|\hat{\mathcal{Y}} \cap \mathcal{G}|}{|\hat{\mathcal{Y}}|}, \mathsf{F} = \frac{2\mathsf{PR}}{\mathsf{P}+\mathsf{R}}$. Current OV-MER evaluation mainly relies on recall-based metrics (Lian et al., 2025a) or F1-based metrics (Lian et al., 2025c). We further examine precision and category-level confidence to diagnose the thinking paradox.

3.2 Precision–Recall Trade-off

In Figure 3 (a), we report the performance gaps between fast and slow thinking in terms of precision, recall, and F1 on five widely used datasets (Lian et al., 2025c;b; 2023a; 2024; Poria et al., 2019; Busso et al., 2008). Fast thinking consistently leads in recall, while slow thinking leads in precision; overall, fast thinking retains a slight advantage in F1. Figure 3 (b) reports the average number of emotion categories predicted by fast and slow thinking. The results show a clear difference in

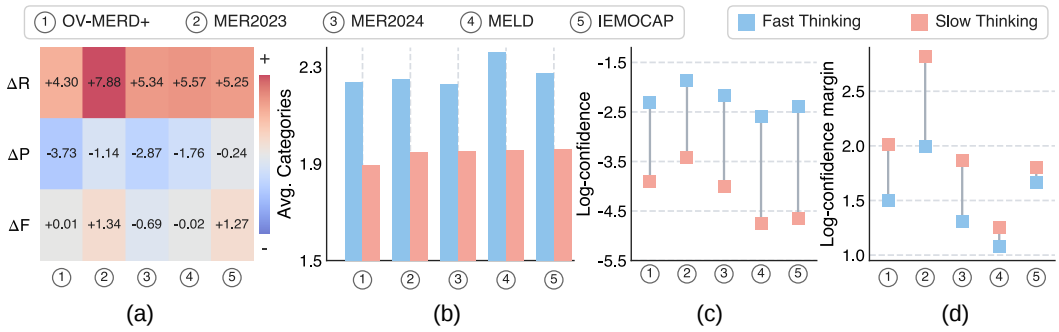


Figure 3: **Analyses of the thinking paradox on five MER benchmarks.** (a) Performance gaps in recall (R), precision (P), and F_1 (F), where $\Delta \mathsf{M} = \mathsf{M}_{\mathsf{fast}} - \mathsf{M}_{\mathsf{slow}}$ for $\mathsf{M} \in \{\mathsf{R}, \mathsf{P}, \mathsf{F}\}$. (b) Prediction diversity by the average number of emotion categories. (c) Confidence on ground-truth emotion categories. (d) Confidence margin between ground-truth and hard-negative emotion categories.

prediction diversity: fast thinking covers more emotion categories, whereas slow thinking produces more concentrated predictions. Together, these two analyses lead to the following finding:

> ***Finding 1:*** *Fast thinking improves recall with broader emotion coverage, whereas slow thinking favors precision with more concentrated predictions.*

Previous benchmarks such as MER-UniBench (Lian et al., 2025a) mainly use hitrate as the evaluation metric, which reflects recall but does not penalize over-prediction (low precision). This helps explain why fast thinking appears more favorable than slow thinking under current evaluation.

3.3 Confident vs. Conservative Predictions

To compare the confidence behavior, we first aggregate the probability mass of free-form emotion words mapped to the level-1 category (Definition 1). We then contrast this confidence between ground-truth and hard-negative categories to define a confidence margin (Definition 2).

Definition 1 (Category-level log-confidence). *For a level-1 emotion category c, let $\mathcal{V}(c)$ denote the set of emotion words associated with c defined by the emotion wheel (Lian et al., 2025c). Given a thinking mode $m \in \{\mathsf{fast}, \mathsf{slow}\}$, we define the category-level confidence of c as:*

$$\mathsf{conf}^m(c) = \log \sum_{e \in \mathcal{V}(c)} p^m(e), \tag{1}$$

where $p^m(e)$ denotes the probability assigned to the first token of emotion word e.

Definition 2 (Confidence margin). *Let $\mathcal{G}$ denote the set of ground-truth level-1 categories for a sample, and let $\mathcal{H}^m$ denote the set of top-3 hard-negative categories. We define $\mathsf{conf}^m_{\mathsf{gt}} = \frac{1}{|\mathcal{G}|} \sum_{c \in \mathcal{G}} \mathsf{conf}^m(c)$ and $\mathsf{conf}^m_{\mathsf{neg}} = \frac{1}{|\mathcal{H}^m|} \sum_{c \in \mathcal{H}^m} \mathsf{conf}^m(c)$ as the average confidence assigned to correct and incorrect categories, respectively. Their relative confidence margin is defined as:*

$$\gamma^m = \mathsf{conf}^m_{\mathsf{gt}} - \mathsf{conf}^m_{\mathsf{neg}}. \tag{2}$$

Figure 3(c) reports the category-level log-confidence on ground-truth emotion categories, showing that fast thinking consistently assigns higher confidence to correct emotions. Figure 3(d) reports the confidence margin between ground-truth and hard-negative categories, showing that slow thinking yields larger margins by suppressing hard negatives. Together, these results lead to Finding 2:

> ***Finding 2:*** *Fast thinking is more confident on correct emotions, while slow thinking is more conservative, better suppressing incorrect categories but under-confidence correct ones.*

3.4 Desiderata for Slow-Fast Thinking Synergy

These observations suggest that an effective emotion reasoning model should not simply prefer one thinking mode over the other. Instead, it should satisfy two desiderata.

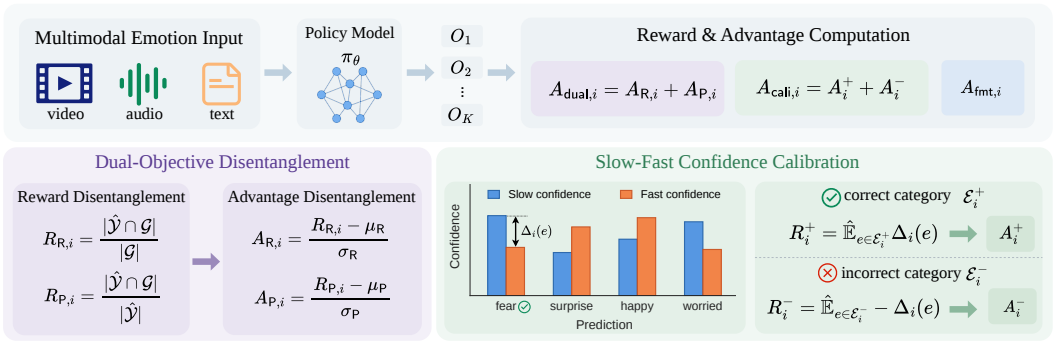


Figure 4: **Framework of MER-R1.** Our MER-R1 consists of dual-objective disentanglement for Desideratum I and slow-fast confidence calibration for Desideratum II.

- **Desideratum I**: At the prediction level, the model should preserve the recall-oriented coverage of fast thinking while maintaining the precision-oriented selectivity of slow thinking.
- **Desideratum II**: At the confidence level, the model should preserve fast thinking's strong confidence on correct categories, while retaining slow thinking's suppression of incorrect ones.

These desiderata motivate our slow-fast thinking synergy framework in Section 4.

4 METHOD

We propose MER-R1, an RL framework for slow-fast thinking synergy. Section 4.1 introduces the GRPO baseline and its limitations. For prediction-level synergy (Desideratum I), Section 4.2 introduces dual-objective disentanglement, which preserves fast thinking's recall-oriented coverage and slow thinking's precision-oriented selectivity as separate optimization signals. For confidence-level synergy (Desideratum II), Section 4.3 introduces slow-fast confidence calibration, which preserves fast thinking's strong confidence on correct categories while retaining slow thinking's suppression of incorrect ones. The overall pipeline is illustrated in Figure 4.

4.1 PRELIMINARY ON GRPO BASELINE

Reasoning-oriented MLLMs typically follow two-stage training: supervised fine-tuning (SFT), followed by GRPO-style RL. Given input x, GRPO samples K outputs $\{y_i\}_{i=1}^K$ from $\pi_{\theta_{\text{old}}}$, computes rewards $\{R_i\}_{i=1}^K$, normalizes them into relative advantages $A_i = \frac{R_i - \mu}{\sigma}$, and optimizes:

$$\mathcal{L}_{\text{GRPO}}(\theta) = -\frac{1}{K}\sum_{i=1}^{K} \min\big(r_i A_i,\ \text{clip}(r_i, 1-\eta, 1+\eta) A_i\big) + \beta D_{\text{KL}}(\pi_\theta \,\|\, \pi_{\text{ref}}), \tag{3}$$

where $r_i = \frac{\pi_\theta(y_i|x)}{\pi_{\theta_{\text{old}}}(y_i|x)}$ is the importance ratio, and $\beta \geq 0$ controls the KL regularization. As a typical reward design (Lian et al., 2026), the baseline combines an emotion-wheel-based F_1 reward with a format reward: $R_{\text{base}} = R_{\text{F}} + R_{\text{fmt}}$, where R_{fmt} encourages <think> then <answer> structure.

However, the GRPO baseline has two limitations: the F_1-style reward entangles recall and precision, allowing the model to trade off one for the other, and the training objective does not explicitly preserve the recall-oriented intuition of fast thinking.

4.2 DUAL-OBJECTIVE DISENTANGLEMENT FOR DESIDERATUM I

To avoid the ambiguous trade-off induced by a single F_1-style reward, we explicitly separate the two optimization goals behind emotion recognition: covering correct categories and avoiding spurious ones. We first decompose the task reward into recall and precision rewards, and then compute their group-normalized advantages separately as the optimization signals.

Reward-level disentanglement. We decompose F_1 reward into recall and precision rewards:

$$R_{\mathsf{R}} = \frac{|\hat{\mathcal{Y}} \cap \mathcal{G}|}{|\mathcal{G}|}, \qquad R_{\mathsf{P}} = \frac{|\hat{\mathcal{Y}} \cap \mathcal{G}|}{|\hat{\mathcal{Y}}|}. \tag{4}$$

This decomposition makes the hidden bi-objective structure of F_1 explicit: R_{R} promotes coverage of correct categories, while R_{P} penalizes spurious predictions. Since different recall–precision trade-offs can yield the same F_1 value, separating R_{R} and R_{P} avoids an ambiguous optimization target.

Advantage-level disentanglement. Given the two rewards R_{R} and R_{P}, a straightforward GRPO pipeline would first aggregate them into a scalar reward and then apply group normalization to obtain the advantage. However, when the two rewards have different group-wise variances, the combined scalar reward is dominated by the higher-variance objective, causing the lower-variance objective to be underrepresented in the final advantage. We therefore first compute group-normalized advantages for R_{R} and R_{P} separately then compute the final advantage A_{dual}:

$$A_{\mathsf{R},i} = \frac{R_{\mathsf{R},i} - \mu_{\mathsf{R}}}{\sigma_{\mathsf{R}}}, \qquad A_{\mathsf{P},i} = \frac{R_{\mathsf{P},i} - \mu_{\mathsf{P}}}{\sigma_{\mathsf{P}}}, \qquad A_{\mathsf{dual},i} = A_{\mathsf{R},i} + A_{\mathsf{P},i}. \tag{5}$$

Theoretical justification. The following proposition shows that standard F_1 optimization can be biased toward the objective with larger normalized variation, whereas our dual-objective disentanglement removes this correlation imbalance (proof in Appendix A).

Proposition 1. *Let $\mu_{\mathsf{R}}, \mu_{\mathsf{P}}$ and $\sigma_{\mathsf{R}}, \sigma_{\mathsf{P}}$ denote the group means and standard deviations of the recall and precision rewards, and let $\tau = \frac{\sigma_{\mathsf{R}}/\mu_{\mathsf{R}}^2}{\sigma_{\mathsf{P}}/\mu_{\mathsf{P}}^2}$ be the normalized group-wise variance ratio. The correlations between each reward and the standard F_1 advantage approximately satisfy*

$$\begin{cases} \mathsf{corr}(R_{\mathsf{R}}, A_{\mathsf{F}}) > \mathsf{corr}(R_{\mathsf{P}}, A_{\mathsf{F}}), & \text{if } \tau > 1, \\ \mathsf{corr}(R_{\mathsf{R}}, A_{\mathsf{F}}) \leq \mathsf{corr}(R_{\mathsf{P}}, A_{\mathsf{F}}), & \text{if } \tau \leq 1. \end{cases} \tag{6}$$

*In contrast, our dual-objective advantage gives **balanced correlations**:*

$$\mathsf{corr}(R_{\mathsf{R}}, A_{\mathsf{dual}}) = \mathsf{corr}(R_{\mathsf{P}}, A_{\mathsf{dual}}). \tag{7}$$

4.3 Slow-fast confidence calibration for Desideratum II

For Desideratum II, we propose slow-fast confidence calibration, which guides the final slow-thinking answer to preserve fast thinking's confidence on correct categories while retaining its suppression of incorrect ones. For each response $y_i = (y_i^{\mathsf{cot}}, y_i^{\mathsf{ans}})$, we construct two answer-generation distributions:

$$p_i^{\mathsf{slow}}(e) = \pi_\theta(\cdot \mid x, y_i^{\mathsf{cot}}, y_{i,<e}^{\mathsf{ans}}), \qquad p_i^{\mathsf{fast}}(e) = \pi_\theta(\cdot \mid x, y_{i,<e}^{\mathsf{ans}}). \tag{8}$$

For each generated emotion word $e \in y_i^{\mathsf{ans}}$, we map it to its level-1 category $c(e)$ and compute its slow-fast confidence gap using Eq. (1):

$$\Delta_i(e) = \mathsf{conf}_i^{\mathsf{slow}}(c(e)) - \mathsf{conf}_i^{\mathsf{fast}}(c(e)). \tag{9}$$

We split the emotion words parsed from y_i^{ans} into correct ($\mathcal{E}_i^+$) and incorrect ($\mathcal{E}_i^-$) sets:

$$\mathcal{E}_i^+ = \{e \in y_i^{\mathsf{ans}} \mid c(e) \in \mathcal{G}\}, \qquad \mathcal{E}_i^- = \{e \in y_i^{\mathsf{ans}} \mid c(e) \notin \mathcal{G}\}. \tag{10}$$

We use the *slow-fast confidence gap* to calibrate the answer in two opposite directions. For emotion words mapped to correct categories, we reward a positive gap so that slow thinking preserves or exceeds the strong confidence of fast thinking. For incorrect emotion words, we reward a negative gap so that slow thinking retains its conservative suppression of spurious predictions:

$$R_i^+ = \frac{1}{|\mathcal{E}_i^+|} \sum_{e \in \mathcal{E}_i^+} \Delta_i(e), \qquad R_i^- = \frac{1}{|\mathcal{E}_i^-|} \sum_{e \in \mathcal{E}_i^-} -\Delta_i(e). \tag{11}$$

Following the same disentanglement strategy, we first compute group-normalized advantages for R_i^+ and R_i^- separately, denoted as A_i^+ and A_i^-, and then combine them as $A_{\mathsf{cali},i} = A_i^+ + A_i^-$. The overall advantage is combined as:

$$A_i = A_{\mathsf{dual},i} + \lambda_{\mathsf{cali}} A_{\mathsf{cali},i} + \lambda_{\mathsf{fmt}} A_{\mathsf{fmt},i}. \tag{12}$$

Finally, we substitute A_i into Eq. (3) for policy optimization.

Table 1: **Main results on MER-UniBench.** MER-R1 achieves the best overall performance under the official benchmark protocol. Entries marked with † are reproduced results, with details in Appendix C.

Model	Fine (F1)	Basic Emotion (Hitrate)				Sentiment Analysis (WAF)				Mean
	OV-MERD+	MER23	MER24	MELD	IEMOCAP	MOSI	MOSEI	SIMS	SIMSv2	
Qwen-Audio (Chu et al., 2023)	32.36	41.85	31.61	49.09	35.47	70.09	46.90	70.73	65.26	49.26
SALMONN (Tang et al., 2024)	45.00	55.53	45.38	45.62	46.84	81.00	67.03	68.69	65.93	57.89
VideoChat2 (Li et al., 2024a)	39.21	33.67	54.50	36.64	48.70	66.84	54.32	69.49	70.66	52.67
LLaMA-VID (Li et al., 2024b)	45.01	50.72	57.60	42.75	46.02	61.78	63.89	69.35	67.48	56.07
Chat-UniVi (Jin et al., 2024)	48.00	57.62	65.67	45.61	52.37	54.53	63.18	68.15	66.36	57.94
mPLUG-Owl (Ye et al., 2023)	48.18	56.86	59.89	49.11	55.54	72.40	72.91	72.13	75.00	62.45
PandaGPT (Su et al., 2023)	37.12	40.21	51.89	37.88	44.04	61.92	67.61	68.38	67.23	52.92
R1-Omni (Zhao et al., 2025)	55.24	64.17	67.43	43.20	51.58	58.02	56.48	71.82	68.58	59.61
Emotion-LLaMA (Cheng et al., 2024)	52.97	59.38	73.62	46.76	55.47	66.13	67.66	78.32	77.23	64.17
AffectGPT (Lian et al., 2025a)	62.52	78.54	78.80	55.65	60.54	81.30	80.90	**88.49**	86.18	74.77
AffectGPT-RL (Lian et al., 2026)	68.05	84.32	**93.75**	63.12	74.26	79.39	79.24	88.25	84.97	79.48
AffectGPT†	59.89	73.69	76.90	52.09	57.74	78.30	78.51	85.28	85.31	71.96
AffectGPT-RL†	64.08	81.88	80.89	57.53	65.71	80.29	80.64	85.70	83.75	75.60
Baseline	66.86	78.16	82.70	61.45	67.28	83.48	86.31	87.78	86.84	77.87
MER-R1	**70.68**	**88.60**	92.96	**71.90**	**78.10**	**86.06**	**86.73**	88.04	**88.48**	**83.50**

Table 2: **Main results on MME-Emotion.** We report the CoT score for each task and all three metrics for the overall average. We include VideoAuto-R1 (Liu et al., 2026) as a video reasoning baseline. MER-R1 achieves the best overall performance.

Methods	ER-Lab	ER-Wild	FG-ER	FG-SA	ML-ER	Noise-ER	IR	SA	Mean		
									Rec.	Rea.	CoT
PandaGPT	28.5	23.2	25.2	41.4	28.9	26.3	29.2	37.7	21.6	38.4	30.0
Emotion-LLaMA	31.7	19.7	31.7	36.4	29.5	49.3	27.3	40.7	21.1	42.4	31.8
AffectGPT	35.0	32.8	32.6	**44.2**	31.3	50.7	30.0	47.9	19.7	57.4	38.6
Baseline	47.0	40.8	35.0	35.6	37.7	59.0	32.4	58.7	27.9	62.7	45.3
VideoAuto-R1	49.4	44.4	**40.2**	33.1	37.3	64.8	32.8	56.7	34.5	58.8	46.6
MER-R1	**57.3**	**49.6**	38.7	37.2	**42.2**	**72.1**	**33.9**	**62.9**	**38.4**	**64.6**	**51.5**

5 EXPERIMENTS

Datasets and metrics. We evaluate MER-R1 on two multimodal emotion benchmarks: MER-UniBench (Lian et al., 2025a) and MME-Emotion (Zhang et al., 2026). On **MER-UniBench**, we follow its three-task setting with 9 datasets: fine-grained emotion recognition on OV-MERD+ (Lian et al., 2025c), evaluated by the *F1-score* metric; basic emotion recognition on MER23 (Lian et al., 2023a), MER24 (Lian et al., 2024), MELD (Poria et al., 2019), and IEMOCAP (Busso et al., 2008), evaluated by *Hitrate*; and sentiment analysis on MOSI (Zadeh et al., 2016), MOSEI (Zadeh et al., 2018), SIMS (Yu et al., 2020), and SIMSv2 (Liu et al., 2022), evaluated by *Weighted Average F1-score (WAF)*. Since *Hitrate* mainly reflects recall and does not penalize over-prediction, we re-evaluate the four basic emotion recognition datasets using *F1-score*. We further evaluate on **MME-Emotion**, which covers 8 tasks: lab, wild, noisy, fine-grained, and multi-label emotion recognition (ER-Lab, ER-Wild, Noise-ER, FG-ER, ML-ER), together with sentiment analysis (SA), fine-grained sentiment analysis (FG-SA), and intent recognition (IR). MME-Emotion reports three metrics: *Recognition* for emotion recognition, *Reasoning* for emotion reasoning, and *CoT* as their average. As GPT-4o used in the original benchmark is no longer available, we evaluate all methods with `gemini-3.1-flash-lite-preview`. More details are in Appendix B.

Implementation details. We use Qwen2.5-Omni (Xu et al., 2025a) as the backbone for all main experiments and train it on MER-Caption+ (Lian et al., 2025a) with a two-stage pipeline. We first perform SFT on 5k randomly sampled instances for 2 epochs with a learning rate of 2×10^{-5}, and then apply RL on the remaining data for 1 epoch with a learning rate of 2×10^{-6}. We set both λ_{cali} and λ_{fmt} to 0.1. Additional implementation details are provided in Appendix C. We further validate MER-R1 on MiniCPM-o-4.5 (Cui et al., 2026) in Appendix D.1, demonstrating its generalization across backbones and consistent slow–fast synergy.

Table 3: **Slow Thinking vs. Fast Thinking with unified F1 metrics.** The four basic emotion recognition datasets are re-evaluated by F1. **R-Mean** averages the five emotion-recognition datasets. MER−R1 makes the final slow-thinking answer outperform its fast-thinking counterpart.

Model	Mode	Fine (F1)	Basic Emotion (F1)				Sentiment Analysis (WAF)				R-Mean
		OV-MERD+	MER23	MER24	MELD	IEMOCAP	MOSI	MOSEI	SIMS	SIMSv2	
Baseline	Fast	**66.87**	**61.92**	64.37	46.18	**53.13**	**86.03**	**86.47**	87.48	**87.60**	**58.49**
	Slow	66.86	60.58	**65.06**	**46.20**	51.86	83.48	86.31	**87.78**	86.84	58.11 ↓
VideoAuto-R1	Fast	**69.67**	**65.09**	**67.38**	**46.77**	**56.33**	**86.21**	**86.05**	86.20	**87.93**	**61.05**
	Slow	69.49	64.20	67.20	46.36	55.32	85.75	85.86	**86.27**	87.83	60.51 ↓
MER−R1	Fast	70.10	64.41	67.21	46.89	55.71	85.30	**87.43**	**88.57**	87.49	60.86
	Slow	**70.68**	**65.43**	**68.46**	**48.70**	**55.73**	**86.06**	86.73	88.04	**88.48**	**61.80** ↑

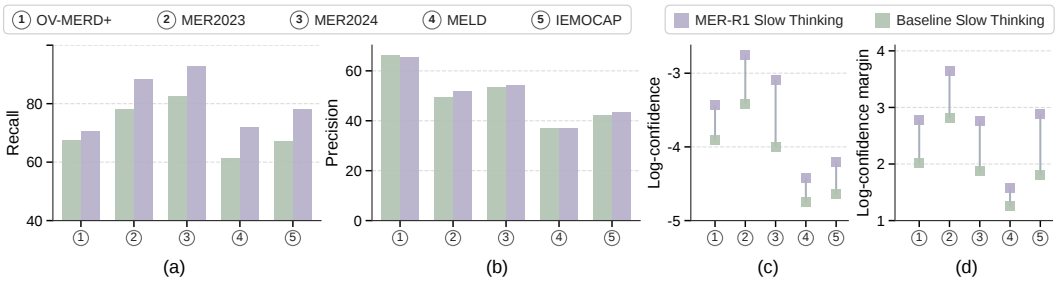


Figure 5: **Effect of slow-fast thinking synergy** (a) Recall. (b) Precision. (c) Confidence on ground-truth categories. (d) Confidence margin. MER−R1 improves recall, preserves precision, boosts confidence on ground-truth categories, and strengthens confidence discrimination.

5.1 Main Results

SoTA performance on MER-UniBench. Table 1 reports the main results under the official MER-UniBench protocol. MER−R1 achieves a new state-of-the-art mean score of 83.50, outperforming the strongest reproduced baseline by 5.63 points. The improvement is consistent across fine-grained emotion recognition, basic emotion recognition, and sentiment analysis. Notably, MER−R1 improves OV-MERD+ from 66.86 to 70.68, obtains large gains on MER23, MELD, and IEMOCAP, and achieves the best results on MOSI, MOSEI, and SIMSv2. These results demonstrate the effectiveness of slow-fast thinking synergy across diverse multimodal emotion tasks.

SoTA performance on MME-Emotion. Table 2 reports the results on MME-Emotion. MER−R1 achieves the best overall CoT mean of 51.5, outperforming all compared methods. Compared with the baseline, MER−R1 substantially improves the overall Recognition score from 27.9 to 38.4, with clear gains on ER-Lab, ER-Wild, ML-ER, Noise-ER, and SA. Meanwhile, the overall Reasoning score also increases from 62.7 to 64.6, indicating that the improvement does not come at the cost of reasoning quality. These results show that slow-fast thinking synergy generalizes beyond MER-UniBench and consistently strengthens both emotion recognition and reasoning.

Fulfilling the Desiderata. MER−R1 fulfills the Desiderata defined in Section 3.4. Table 3 compares slow and fast thinking under a unified F1-based evaluation, where the four basic emotion recognition datasets are re-evaluated by F1 rather than Hitrate. Under this setting, the baseline still exhibits the thinking paradox: its final slow-thinking answer is worse than its fast-thinking counterpart. VideoAuto-R1 (Liu et al., 2026) trains an *answer-think-answer* format with separate answer rewards, but still leaves the final answer weaker than the fast thinking on emotion recognition, with its R-Mean decreasing from 61.05 to 60.51, as it lacks a mechanism for slow-fast complementarity. In contrast, MER−R1 successfully reverses this trend: **the slow-thinking answer outperforms fast thinking on all five recognition datasets**, improving the R-Mean from 60.86 to 61.80 and demonstrating that MER−R1 turns slow thinking into stronger final recognition.

Figure 5 further explains how MER−R1 satisfies the two desiderata. At the prediction level, Figure 5(a) and (b) show that MER−R1 improves recall over the baseline while preserving precision. At the confidence level, Figure 5(c) and (d) show that MER−R1 assigns higher confidence to ground-truth

Table 4: **Main ablation study on MER-UniBench. RD**, **AD**, and **SFCC** denote Reward Disentanglement, Advantage Disentanglement, and Slow-Fast Confidence Calibration, respectively. **Mean$_{\text{official}}$** uses the original bench metrics, while **Mean$_{\text{f1}}$** uses F1 for the basic emotion recognition task.

Variant	RD	AD	SFCC	Fine (F1)	Basic (Hitrate)	Basic (F1)	Sentiment (WAF)	Mean$_{\text{official}}$	Mean$_{\text{f1}}$
Baseline				66.86	72.40	55.93	86.10	77.87	70.55
Ex1	✓			68.06	77.50	57.18	85.89	80.18	71.15
Ex2	✓	✓		69.77	82.40	58.55	85.11	82.20	71.60
Ex3	✓	✓	✓	**70.68**	**82.90**	**59.58**	**87.32**	**83.50**	**73.15**

Table 5: **Ablation on slow-fast confidence calibration.** The first block compares different calibration formulations, including confidence granularity, and the role of the precision calibration term. The second block compares different optimization spaces.

Variant	Fine (F1)	Basic (Hitrate)	Basic (F1)	Sentiment (WAF)	Mean$_{\text{official}}$	Mean$_{\text{f1}}$
Baseline	66.86	72.40	55.93	86.10	77.87	70.55
Study on calibration formulation						
A1: word-level confidence	69.07	80.67	58.89	86.14	81.81	72.13
A2: w/o precision calibration	69.48	82.51	59.09	86.58	82.87	72.46
Study on optimization space						
B1: reward-space calibration	69.03	81.83	59.06	85.19	81.91	71.78
B2: mixed-space calibration	67.81	**83.62**	58.44	86.42	83.11	71.91
MER-R1	**70.68**	82.90	**59.58**	**87.32**	**83.50**	**73.15**

categories and enlarges the margin between ground-truth and hard-negative categories. Together with Table 3, these results demonstrate that MER-R1 fulfills the desiderata, turning slow-fast complementarity into a stronger final slow-thinking answer.

5.2 Ablation Studies and Analyses

Main ablation study. Table 4 presents the main ablation study on MER-UniBench by progressively adding Reward Disentanglement (RD), Advantage Disentanglement (AD), and Slow-Fast Confidence Calibration (SFCC). RD improves the baseline, and AD further strengthens fine-grained and basic emotion recognition by preserving the disentangled optimization signals in the advantage space; together, they increase $\text{Mean}_{\text{official}}$ from 77.87 for the *Baseline* to 82.20 in Ex2. Adding SFCC is key to reversing the thinking paradox and further improves Mean_{F1} from 71.60 in Ex2 to 73.15. These results show that all three components contribute to the final gain of MER-R1.

Ablation on slow-fast confidence calibration. Table 5 ablates the design of Slow-Fast Confidence Calibration. A1 uses word-level confidence instead of category-level confidence and performs worse, showing the importance of calibrating in the emotion-wheel category space. A2 removes precision calibration and is also inferior to the full model, indicating that recall-side calibration alone cannot preserve slow-thinking selectivity. For optimization space, B1 applies calibration in the reward space, while B2 adds corresponding dual and calibration rewards before advantage normalization. Both are weaker than MER-R1 in overall performance, demonstrating that SFCC works best with category-level bidirectional calibration and fully disentangled advantage-space optimization.

6 Conclusion

We reveal a thinking paradox in MER, where explicit reasoning does not necessarily improve recognition accuracy. Our analysis shows that fast thinking provides recall-oriented coverage and correct-category confidence, while slow thinking offers precision-oriented selectivity. Based on this complementarity, we propose MER-R1, an RL framework that realizes slow-fast thinking synergy through dual-objective disentanglement and slow-fast confidence calibration. Experiments on MER-UniBench and MME-Emotion show that MER-R1 achieves SoTA performance and fulfills the Desiderata, making the slow-thinking genuinely beneficial to emotion recognition.

AI USE STATEMENT

Generative AI tools were used for language editing and presentation support. The authors reviewed the resulting text, checked technical statements against the implemented method and reported experiments, and take responsibility for the final content.

ETHICS STATEMENT

This work uses existing publicly available multimodal emotion recognition benchmarks and does not collect new human-subject data or annotations. Nonetheless, automated emotion recognition can raise concerns about privacy, consent, cultural and demographic bias, and misuse, particularly in surveillance, behavioral profiling, or high-stakes decision-making. Our results are limited to controlled benchmark evaluations and should not be interpreted as evidence that the system is suitable for such deployments. Any real-world use should require informed consent, privacy protection, subgroup-level fairness evaluation, and appropriate human oversight.

REPRODUCIBILITY STATEMENT

The method, training objectives, and algorithms are specified in Section 4. The experimental datasets, evaluation protocols, backbone, baselines, training hyperparameters, and compute resources are reported in Section 5, with additional implementation details in Appendix C. Anonymized code and reproduction instructions are provided in the Supplementary Materials.

REFERENCES

Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan. Iemocap: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, 42(4):335–359, 2008.

Zebang Cheng, Zhi-Qi Cheng, Jun-Yan He, Jingdong Sun, Kai Wang, Yuxiang Lin, Zheng Lian, Xiaojiang Peng, and Alexander G Hauptmann. Emotion-llama: Multimodal emotion recognition and reasoning with instruction tuning. In *NeurIPS*, 2024.

Yunfei Chu, Jin Xu, Xiaohuan Zhou, Qian Yang, Shiliang Zhang, Zhijie Yan, Chang Zhou, and Jingren Zhou. Qwen-audio: Advancing universal audio understanding via unified large-scale audio-language models. *arXiv preprint arXiv:2311.07919*, 2023.

Junbo Cui, Bokai Xu, Chongyi Wang, Tianyu Yu, Weiyue Sun, Yingjing Xu, Tianran Wang, Zhihui He, Wenshuo Ma, Tianchi Cai, et al. Minicpm-o 4.5: Towards real-time full-duplex omni-modal interaction. *arXiv preprint arXiv:2604.27393*, 2026.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645(8081):633–638, 2025.

Zhiyuan Han, Beier Zhu, Yanlong Xu, Peipei Song, and Xun Yang. Benchmarking and bridging emotion conflicts for multimodal emotion reasoning. In *ACM MM*, 2025.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024.

Fengqing Jiang, Zhangchen Xu, Yuetai Li, Luyao Niu, Zhen Xiang, Bo Li, Bill Yuchen Lin, and Radha Poovendran. Safechain: Safety of language models with long chain-of-thought reasoning capabilities. In *ACL findings*, 2025.

Peng Jin, Ryuichi Takanobu, Wancai Zhang, Xiaochun Cao, and Li Yuan. Chat-univi: Unified visual representation empowers large language models with image and video understanding. In *CVPR*, 2024.

Kunchang Li, Yali Wang, Yinan He, Yizhuo Li, Yi Wang, Yi Liu, Zun Wang, Jilan Xu, Guo Chen, Ping Luo, et al. Mvbench: A comprehensive multi-modal video understanding benchmark. In *CVPR*, 2024a.

Yanwei Li, Chengyao Wang, and Jiaya Jia. Llama-vid: An image is worth 2 tokens in large language models. In *ECCV*, 2024b.

Zheng Lian, Haiyang Sun, Licai Sun, Kang Chen, Mingyu Xu, Kexin Wang, Ke Xu, Yu He, Ying Li, Jinming Zhao, et al. Mer 2023: Multi-label learning, modality robustness, and semi-supervised learning. In *ACM MM*, 2023a.

Zheng Lian, Licai Sun, Mingyu Xu, Haiyang Sun, Ke Xu, Zhuofan Wen, Shun Chen, Bin Liu, and Jianhua Tao. Explainable multimodal emotion reasoning. *CoRR*, 2023b.

Zheng Lian, Haiyang Sun, Licai Sun, Zhuofan Wen, Siyuan Zhang, Shun Chen, Hao Gu, Jinming Zhao, Ziyang Ma, Xie Chen, et al. Mer 2024: Semi-supervised learning, noise robustness, and open-vocabulary multimodal emotion recognition. In *Proceedings of the 2nd International Workshop on Multimodal and Responsible Affective Computing*, 2024.

Zheng Lian, Haoyu Chen, Lan Chen, Haiyang Sun, Licai Sun, Yong Ren, Zebang Cheng, Bin Liu, Rui Liu, Xiaojiang Peng, et al. Affectgpt: A new dataset, model, and benchmark for emotion understanding with multimodal large language models. In *ICML*, 2025a.

Zheng Lian, Rui Liu, Kele Xu, Bin Liu, Xuefei Liu, Yazhou Zhang, Xin Liu, Yong Li, Zebang Cheng, Haolin Zuo, et al. Mer 2025: When affective computing meets large language models. In *ACM MM*, 2025b.

Zheng Lian, Haiyang Sun, Licai Sun, Haoyu Chen, Lan Chen, Hao Gu, Zhuofan Wen, Shun Chen, Siyuan Zhang, Hailiang Yao, et al. Ov-mer: Towards open-vocabulary multimodal emotion recognition. In *ICML*, 2025c.

Zheng Lian, Fan Zhang, Lan Chen, Yazhou Zhang, Rui Liu, Jinyang Wu, Haoyu Chen, Xiaobai Li, Xiaojiang Peng, Bin He, et al. Affectgpt-rl: Revealing roles of reinforcement learning in open-vocabulary emotion recognition. *NeurIPS*, 2026.

Shuming Liu, Mingchen Zhuge, Changsheng Zhao, Jun Chen, Lemeng Wu, Zechun Liu, Chenchen Zhu, Zhipeng Cai, Chong Zhou, Haozhe Liu, et al. Videoauto-r1: Video auto reasoning via thinking once, answering twice. *arXiv preprint arXiv:2601.05175*, 2026.

Yihe Liu, Ziqi Yuan, Huisheng Mao, Zhiyun Liang, Wanqiuyue Yang, Yuanzhe Qiu, Tie Cheng, Xiaoteng Li, Hua Xu, and Kai Gao. Make acoustic and visual cues matter: Ch-sims v2. 0 dataset and av-mixup consistent module. In *ICMI*, 2022.

Rosalind W Picard. *Affective computing*. MIT press, 2000.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. Meld: A multimodal multi-party dataset for emotion recognition in conversations. In *ACL*, 2019.

Hyeongseop Rha, Jeong Hun Yeo, Yeonju Kim, and Yong Man Ro. Emotion-coherent reasoning for multimodal llms via emotional rationale verifier. In *AAAI*, 2026.

Yuntao Shou, Tao Meng, Wei Ai, and Keqin Li. Multimodal large language models meet multimodal emotion recognition and reasoning: A survey. *arXiv preprint arXiv:2509.24322*, 2025.

Aaditya Singh, Adam Fry, Adam Perelman, Adam Tart, Adi Ganesh, Ahmed El-Kishky, Aidan McLaughlin, Aiden Low, AJ Ostrow, Akhila Ananthram, et al. Openai gpt-5 system card. *arXiv preprint arXiv:2601.03267*, 2025.

Yixuan Su, Tian Lan, Huayang Li, Jialu Xu, Yan Wang, and Deng Cai. Pandagpt: One model to instruction-follow them all. In *TLLM*, 2023.

Changli Tang, Wenyi Yu, Guangzhi Sun, Xianzhao Chen, Tian Tan, Wei Li, Lu Lu, Zejun Ma, and Chao Zhang. Salmonn: Towards generic hearing abilities for large language models. In *ICLR*, 2024.

Jin Xu, Zhifang Guo, Jinzheng He, Hangrui Hu, Ting He, Shuai Bai, Keqin Chen, Jialin Wang, Yang Fan, Kai Dang, et al. Qwen2.5-omni technical report. *arXiv preprint arXiv:2503.20215*, 2025a.

Jin Xu, Zhifang Guo, Hangrui Hu, Yunfei Chu, Xiong Wang, Jinzheng He, Yuxuan Wang, Xian Shi, Ting He, Xinfa Zhu, Yuanjun Lv, Yongqi Wang, Dake Guo, He Wang, Linhan Ma, Pei Zhang, Xinyu Zhang, Hongkun Hao, Zishan Guo, Baosong Yang, Bin Zhang, Ziyang Ma, Xipin Wei, Shuai Bai, Keqin Chen, Xuejing Liu, Peng Wang, Mingkun Yang, Dayiheng Liu, Xingzhang Ren, Bo Zheng, Rui Men, Fan Zhou, Bowen Yu, Jianxin Yang, Le Yu, Jingren Zhou, and Junyang Lin. Qwen3-omni technical report. *arXiv preprint arXiv:2509.17765*, 2025b.

Qize Yang, Shimin Yao, Weixuan Chen, Shenghao Fu, Detao Bai, Jiaxing Zhao, Boyuan Sun, Bowen Yin, Xihan Wei, and Jingren Zhou. Humanomniv2: From understanding to omni-modal reasoning with context. *arXiv preprint arXiv:2506.21277*, 2025.

Qinghao Ye, Haiyang Xu, Guohai Xu, Jiabo Ye, Ming Yan, Yiyang Zhou, Junyang Wang, Anwen Hu, Pengcheng Shi, Yaya Shi, et al. mplug-owl: Modularization empowers large language models with multimodality. *arXiv preprint arXiv:2304.14178*, 2023.

Wenmeng Yu, Hua Xu, Fanyang Meng, Yilin Zhu, Yixiao Ma, Jiele Wu, Jiyun Zou, and Kaicheng Yang. Ch-sims: A chinese multimodal sentiment analysis dataset with fine-grained annotation of modality. In *ACL*, 2020.

Amir Zadeh, Rowan Zellers, Eli Pincus, and Louis-Philippe Morency. Mosi: multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos. In *CoRR*, 2016.

AmirAli Bagher Zadeh, Paul Pu Liang, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. Multimodal language analysis in the wild: Cmu-mosei dataset and interpretable dynamic fusion graph. In *ACL*, 2018.

Duzhen Zhang, Zhong-Zhi Li, Ming-Liang Zhang, Jiaxin Zhang, Zengyan Liu, Yuxuan Yao, Haotian Xu, Junhao Zheng, Xiuyi Chen, Yingying Zhang, et al. From system 1 to system 2: A survey of reasoning large language models. *TPAMI*, 2025.

Fan Zhang, Zebang Cheng, Chong Deng, Haoxuan Li, Zheng Lian, Qian Chen, Huadai Liu, Wen Wang, Yi-Fan Zhang, Renrui Zhang, et al. Mme-emotion: A holistic evaluation benchmark for emotional intelligence in multimodal large language models. In *ICLR*, 2026.

Jiaxing Zhao, Xihan Wei, and Liefeng Bo. R1-omni: Explainable omni-multimodal emotion recognition with reinforcing learning. *arXiv preprint arXiv:2503.05379*, 2025.

Appendix

A Theoretical analysis of dual-objective disentanglement

We provide a theoretical analysis of the dual-objective disentanglement in Section 4.2. We first show that the standard F_1-based advantage is affected by both the local sensitivities of F_1 and the reward variances. We then show that our two-step disentanglement removes these factors by separating recall and precision before group normalization.

A.1 Limitation of the F_1 advantage

For the i-th response in a GRPO group, let $R_{\mathsf{R},i}$ and $R_{\mathsf{P},i}$ denote its recall and precision rewards. The baseline first combines them into a scalar F_1 reward:

$$R_{\mathsf{F},i} = \frac{2R_{\mathsf{R},i}R_{\mathsf{P},i}}{R_{\mathsf{R},i} + R_{\mathsf{P},i}}, \tag{13}$$

and then computes the group-normalized advantage:

$$A_{\mathsf{F},i} = \frac{R_{\mathsf{F},i} - \mu_{\mathsf{F}}}{\sigma_{\mathsf{F}}}, \tag{14}$$

where μ_{F} and σ_{F} are the group mean and standard deviation of R_{F}.

Proposition 2. *For $\mathsf{k} \in \{\mathsf{R}, \mathsf{P}\}$ and $\mathsf{l} \neq \mathsf{k}$, the correlation between reward R_{k} and the F_1 advantage A_{F} is approximated by:*

$$\mathrm{Corr}(R_{\mathsf{k}}, A_{\mathsf{F}}) \approx \frac{\alpha_{\mathsf{k}}\sigma_{\mathsf{k}} + \alpha_{\mathsf{l}}\rho\sigma_{\mathsf{l}}}{\sigma_{\mathsf{F}}}. \tag{15}$$

Here, ρ is the correlation between R_{R} and R_{P}; α_{k} is the local F_1 sensitivity to R_{k}; and σ_{k} is the standard deviation of R_{k}.

Proof. For simplicity, we omit the response index i when no ambiguity arises. Let $\mu_{\mathsf{R}}, \mu_{\mathsf{P}}$ be the group means of R_{R} and R_{P} respectively. By the first-order Taylor expansion around $(\mu_{\mathsf{R}}, \mu_{\mathsf{P}})$, the scalar F_1 reward can be locally approximated as:

$$R_{\mathsf{F}} \approx \frac{2\mu_{\mathsf{R}}\mu_{\mathsf{P}}}{\mu_{\mathsf{R}} + \mu_{\mathsf{P}}} + \alpha_{\mathsf{R}}(R_{\mathsf{R}} - \mu_{\mathsf{R}}) + \alpha_{\mathsf{P}}(R_{\mathsf{P}} - \mu_{\mathsf{P}}), \tag{16}$$

where:

$$\alpha_{\mathsf{R}} = \left.\frac{\partial R_{\mathsf{F}}}{\partial R_{\mathsf{R}}}\right|_{(\mu_{\mathsf{R}},\mu_{\mathsf{P}})} = \frac{2\mu_{\mathsf{P}}^2}{(\mu_{\mathsf{R}} + \mu_{\mathsf{P}})^2}, \qquad \alpha_{\mathsf{P}} = \left.\frac{\partial R_{\mathsf{F}}}{\partial R_{\mathsf{P}}}\right|_{(\mu_{\mathsf{R}},\mu_{\mathsf{P}})} = \frac{2\mu_{\mathsf{R}}^2}{(\mu_{\mathsf{R}} + \mu_{\mathsf{P}})^2}. \tag{17}$$

Subtracting the group mean from the local approximation and normalizing by σ_{F} give:

$$R_{\mathsf{F}} - \mu_{\mathsf{F}} \approx \alpha_{\mathsf{R}}(R_{\mathsf{R}} - \mu_{\mathsf{R}}) + \alpha_{\mathsf{P}}(R_{\mathsf{P}} - \mu_{\mathsf{P}}), \qquad A_{\mathsf{F}} \approx \frac{\alpha_{\mathsf{R}}(R_{\mathsf{R}} - \mu_{\mathsf{R}}) + \alpha_{\mathsf{P}}(R_{\mathsf{P}} - \mu_{\mathsf{P}})}{\sigma_{\mathsf{F}}}. \tag{18}$$

We first compute the correlation between R_{R} and A_{F}. By definition:

$$\mathrm{Corr}(R_{\mathsf{R}}, A_{\mathsf{F}}) = \frac{\mathrm{Cov}(R_{\mathsf{R}}, A_{\mathsf{F}})}{\sigma_{\mathsf{R}}\sigma_{A_{\mathsf{F}}}} = \frac{\mathrm{Cov}(R_{\mathsf{R}}, A_{\mathsf{F}})}{\sigma_{\mathsf{R}}}, \tag{19}$$

where the equality follows from the group normalization of A_{F}, which gives $\sigma_{A_{\mathsf{F}}} = 1$. Using the expression of A_{F}, together with $\mathrm{Var}(R_{\mathsf{R}}) = \sigma_{\mathsf{R}}^2$ and $\mathrm{Cov}(R_{\mathsf{R}}, R_{\mathsf{P}}) = \rho\sigma_{\mathsf{R}}\sigma_{\mathsf{P}}$, we obtain:

$$\mathrm{Cov}(R_{\mathsf{R}}, A_{\mathsf{F}}) \approx \frac{\alpha_{\mathsf{R}}\mathrm{Var}(R_{\mathsf{R}}) + \alpha_{\mathsf{P}}\mathrm{Cov}(R_{\mathsf{R}}, R_{\mathsf{P}})}{\sigma_{\mathsf{F}}} = \frac{\alpha_{\mathsf{R}}\sigma_{\mathsf{R}}^2 + \alpha_{\mathsf{P}}\rho\sigma_{\mathsf{R}}\sigma_{\mathsf{P}}}{\sigma_{\mathsf{F}}}. \tag{20}$$

By symmetry, the corresponding result for R_{P} can be obtained in the same way. Substituting the two covariance results into the correlation definition gives:

$$\mathrm{Corr}(R_{\mathsf{R}}, A_{\mathsf{F}}) \approx \frac{\alpha_{\mathsf{R}}\sigma_{\mathsf{R}} + \alpha_{\mathsf{P}}\rho\sigma_{\mathsf{P}}}{\sigma_{\mathsf{F}}}, \qquad \mathrm{Corr}(R_{\mathsf{P}}, A_{\mathsf{F}}) \approx \frac{\alpha_{\mathsf{P}}\sigma_{\mathsf{P}} + \alpha_{\mathsf{R}}\rho\sigma_{\mathsf{R}}}{\sigma_{\mathsf{F}}}. \tag{21}$$

This completes the proof. □

Implication. Proposition 2 shows that the correlation between each reward and the F_1 advantage depends on both the local F_1 sensitivity α_{k} and the reward standard deviation σ_{k}, especially when ρ is small. Therefore, the baseline tends to emphasize the reward with larger local sensitivity or larger standard deviation, even when both recall and precision are important.

A.2 Effect of dual-objective disentanglement

Proposition 3. *The correlation between each reward and the disentangled advantage is:*

$$\mathrm{Corr}(R_{\mathsf{R}}, A_{\mathsf{dual}}) = \mathrm{Corr}(R_{\mathsf{P}}, A_{\mathsf{dual}}) = \sqrt{\frac{1+\rho}{2}}. \tag{22}$$

Proof. By definition, A_{R} and A_{P} are standardized rewards:

$$A_{\mathsf{R}} = \frac{R_{\mathsf{R}} - \mu_{\mathsf{R}}}{\sigma_{\mathsf{R}}}, \qquad A_{\mathsf{P}} = \frac{R_{\mathsf{P}} - \mu_{\mathsf{P}}}{\sigma_{\mathsf{P}}}. \tag{23}$$

Thus, both have zero mean and unit variance, and their covariance is $\mathrm{Cov}(A_{\mathsf{R}}, A_{\mathsf{P}}) = \mathrm{Corr}(R_{\mathsf{R}}, R_{\mathsf{P}}) = \rho$. Since $A_{\mathsf{dual}} = A_{\mathsf{R}} + A_{\mathsf{P}}$, its variance is:

$$\mathrm{Var}(A_{\mathsf{dual}}) = \mathrm{Var}(A_{\mathsf{R}}) + \mathrm{Var}(A_{\mathsf{P}}) + 2\mathrm{Cov}(A_{\mathsf{R}}, A_{\mathsf{P}}) = 2 + 2\rho. \tag{24}$$

We first compute the correlation between R_{R} and A_{dual}. The covariance term can be written as:

$$\begin{aligned} \mathrm{Cov}(R_{\mathsf{R}}, A_{\mathsf{dual}}) &= \mathrm{Cov}(R_{\mathsf{R}}, A_{\mathsf{R}}) + \mathrm{Cov}(R_{\mathsf{R}}, A_{\mathsf{P}}) \\ &= \mathrm{Cov}\left(R_{\mathsf{R}}, \frac{R_{\mathsf{R}} - \mu_{\mathsf{R}}}{\sigma_{\mathsf{R}}}\right) + \mathrm{Cov}\left(R_{\mathsf{R}}, \frac{R_{\mathsf{P}} - \mu_{\mathsf{P}}}{\sigma_{\mathsf{P}}}\right) \\ &= \sigma_{\mathsf{R}} + \rho\sigma_{\mathsf{R}} = (1+\rho)\sigma_{\mathsf{R}}. \end{aligned} \tag{25}$$

Combining this with $\sigma_{A_{\mathsf{dual}}} = \sqrt{2+2\rho}$, we obtain:

$$\mathrm{Corr}(R_{\mathsf{R}}, A_{\mathsf{dual}}) = \frac{\mathrm{Cov}(R_{\mathsf{R}}, A_{\mathsf{dual}})}{\sigma_{\mathsf{R}}\sigma_{A_{\mathsf{dual}}}} = \frac{(1+\rho)\sigma_{\mathsf{R}}}{\sigma_{\mathsf{R}}\sqrt{2+2\rho}} = \sqrt{\frac{1+\rho}{2}}. \tag{26}$$

By symmetry, the same derivation gives:

$$\mathrm{Corr}(R_{\mathsf{P}}, A_{\mathsf{dual}}) = \sqrt{\frac{1+\rho}{2}}. \tag{27}$$

This completes the proof. □

Implication. Proposition 3 shows that dual-objective disentanglement removes the local-sensitivity and standard-deviation factors identified in Proposition 2 by normalizing recall and precision separately. As a result, both rewards are equally correlated with A_{dual}, preventing optimization from improving one objective at the expense of the other.

A.3 Proof of Proposition 1

Restated Proposition (Proposition 1). *Let $\mu_{\mathsf{R}}, \mu_{\mathsf{P}}$ and $\sigma_{\mathsf{R}}, \sigma_{\mathsf{P}}$ denote the group means and standard deviations of the recall and precision rewards, and let $\tau = \frac{\sigma_{\mathsf{R}}/\mu_{\mathsf{R}}^2}{\sigma_{\mathsf{P}}/\mu_{\mathsf{P}}^2}$ be the normalized group-wise variance ratio. The correlations between each reward and the standard F_1 advantage satisfy*

$$\begin{cases} \mathsf{corr}(R_{\mathsf{R}}, A_{\mathsf{F}}) > \mathsf{corr}(R_{\mathsf{P}}, A_{\mathsf{F}}), & \text{if } \tau > 1, \\ \mathsf{corr}(R_{\mathsf{R}}, A_{\mathsf{F}}) \le \mathsf{corr}(R_{\mathsf{P}}, A_{\mathsf{F}}), & \text{if } \tau \le 1. \end{cases} \tag{28}$$

*In contrast, our dual-objective advantage gives **balanced correlations**:*

$$\mathsf{corr}(R_{\mathsf{R}}, A_{\mathsf{dual}}) = \mathsf{corr}(R_{\mathsf{P}}, A_{\mathsf{dual}}). \tag{29}$$

Proof. From Proposition 2, the correlations between recall/precision rewards and the standard F_1 advantage are:

$$\mathrm{Corr}(R_{\mathsf{R}}, A_{\mathsf{F}}) \approx \frac{\alpha_{\mathsf{R}}\sigma_{\mathsf{R}} + \alpha_{\mathsf{P}}\rho\sigma_{\mathsf{P}}}{\sigma_{\mathsf{F}}}, \qquad \mathrm{Corr}(R_{\mathsf{P}}, A_{\mathsf{F}}) \approx \frac{\alpha_{\mathsf{P}}\sigma_{\mathsf{P}} + \alpha_{\mathsf{R}}\rho\sigma_{\mathsf{R}}}{\sigma_{\mathsf{F}}}. \tag{30}$$

Taking their difference gives:

$$\mathrm{Corr}(R_{\mathsf{R}}, A_{\mathsf{F}}) - \mathrm{Corr}(R_{\mathsf{P}}, A_{\mathsf{F}}) \approx \frac{(1-\rho)(\alpha_{\mathsf{R}}\sigma_{\mathsf{R}} - \alpha_{\mathsf{P}}\sigma_{\mathsf{P}})}{\sigma_{\mathsf{F}}}. \tag{31}$$

In the non-degenerate case with $\rho < 1$ and $\sigma_{\mathsf{F}} > 0$, the sign of this difference is determined by $\alpha_{\mathsf{R}}\sigma_{\mathsf{R}} - \alpha_{\mathsf{P}}\sigma_{\mathsf{P}}$. Using the local sensitivities in Proposition 2, we have:

$$\alpha_{\mathsf{R}}\sigma_{\mathsf{R}} - \alpha_{\mathsf{P}}\sigma_{\mathsf{P}} = \frac{2}{(\mu_{\mathsf{R}} + \mu_{\mathsf{P}})^2}\left(\mu_{\mathsf{P}}^2\sigma_{\mathsf{R}} - \mu_{\mathsf{R}}^2\sigma_{\mathsf{P}}\right). \tag{32}$$

Therefore:

$$\alpha_{\mathsf{R}}\sigma_{\mathsf{R}} > \alpha_{\mathsf{P}}\sigma_{\mathsf{P}} \iff \frac{\sigma_{\mathsf{R}}}{\mu_{\mathsf{R}}^2} > \frac{\sigma_{\mathsf{P}}}{\mu_{\mathsf{P}}^2} \iff \tau > 1. \tag{33}$$

Thus, when $\tau > 1$, $\mathrm{Corr}(R_{\mathsf{R}}, A_{\mathsf{F}}) > \mathrm{Corr}(R_{\mathsf{P}}, A_{\mathsf{F}})$; otherwise, $\mathrm{Corr}(R_{\mathsf{R}}, A_{\mathsf{F}}) \le \mathrm{Corr}(R_{\mathsf{P}}, A_{\mathsf{F}})$.

Finally, Proposition 3 directly gives:

$$\mathrm{Corr}(R_{\mathsf{R}}, A_{\mathsf{dual}}) = \mathrm{Corr}(R_{\mathsf{P}}, A_{\mathsf{dual}}) = \sqrt{\frac{1+\rho}{2}}, \tag{34}$$

which proves the balanced-correlation result for the dual-objective advantage. □

B TASK DETAILS

We evaluate `MER-R1` on two multimodal emotion benchmarks: MER-UniBench (Lian et al., 2025a) and MME-Emotion (Zhang et al., 2026). MER-UniBench evaluates open-vocabulary multimodal emotion recognition under emotion-wheel-based metrics, while MME-Emotion provides a llm-as-a-judge based evaluation of both emotion recognition and emotion reasoning.

B.1 MER-UNIBENCH

MER-UniBench (Lian et al., 2025a) unifies nine multimodal emotion datasets into three task groups: fine-grained emotion recognition, basic emotion recognition, and sentiment analysis. Since model outputs are free-form emotion words, direct string matching is insufficient. Following prior work (Lian et al., 2025a;c), we evaluate recognition outputs with an emotion-wheel-based normalization protocol.

Emotion-wheel normalization. The emotion wheel maps diverse fine-grained emotion words into a shared category space. We apply three normalization steps to both predictions and ground-truth labels. First, morphological variants are reduced to their base form, *e.g.*, *happier* and *happiness* are normalized to *happy*. Second, semantically equivalent emotion words are mapped to the same canonical form, *e.g.*, *joyful* and *happy*. Third, fine-grained emotion words are further mapped to their corresponding inner emotion categories according to each emotion wheel. Following prior work (Lian et al., 2025a), we adopt $K = 5$ emotion wheels, as shown in Figure 6. In each wheel, the outer labels correspond to fine-grained emotions, and the inner sectors define the corresponding level-1 emotion categories.

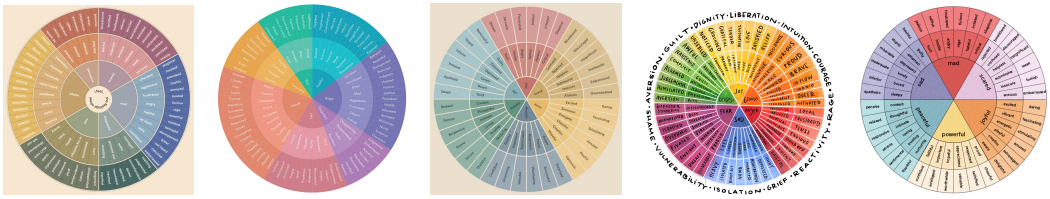
For the k-th wheel w_k, the complete mapping function is defined as:

$$G_{w_k}(\cdot) = F_{l_3}^{w_k}\big(F_{l_2}(F_{l_1}(\cdot))\big), \quad k \in \{1, \ldots, K\}, \tag{35}$$

where F_{l_1}, F_{l_2}, and $F_{l_3}^{w_k}$ denote morphological normalization, semantic normalization, and wheel-specific category mapping, respectively.

Set-level precision, recall, and F1. For sample i, let $\mathcal{Y}_i$ denote the ground-truth emotion set and $\hat{\mathcal{Y}}_i$ denote the predicted emotion set extracted from the `<answer>` field, with duplicate emotion words removed. For each emotion wheel w_k, we compute the set-level precision and recall as:

$$\mathsf{P}_s^k = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|G_{w_k}(\mathcal{Y}_i) \cap G_{w_k}(\hat{\mathcal{Y}}_i)\right|}{\left|G_{w_k}(\hat{\mathcal{Y}}_i)\right|}, \tag{36}$$



(a) W1 (b) W2 (c) W3 (d) W4 (e) W5

Figure 6: **Emotion wheels used in MER-UniBench evaluation.** We adopt five emotion wheels (W1–W5) to map free-form emotion words into level-1 emotion categories.

$$\mathsf{R}_s^k = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|G_{w_k}(\mathcal{Y}_i)\cap G_{w_k}(\hat{\mathcal{Y}}_i)\right|}{|G_{w_k}(\mathcal{Y}_i)|}. \tag{37}$$

The corresponding F1 score is defined as:

$$\mathsf{F}_s^k = \frac{2\mathsf{P}_s^k\mathsf{R}_s^k}{\mathsf{P}_s^k+\mathsf{R}_s^k}. \tag{38}$$

The final emotion-wheel-based F1 score is obtained by averaging over all emotion wheels:

$$\mathsf{F}_{\mathsf{EW}} = \frac{1}{K}\sum_{k=1}^{K}\mathsf{F}_s^k. \tag{39}$$

Fine-grained emotion recognition. Fine-grained emotion recognition requires the model to predict one or more open-vocabulary emotion descriptors. We evaluate this task on OV-MERD+ (Lian et al., 2025c), which extends OV-MERD (Lian et al., 2023b). Following MER-UniBench, we use the emotion-wheel-based set-level F1 score as the primary metric.

Basic emotion recognition. Basic emotion recognition asks the model to infer a basic emotion label from multimodal input. We evaluate this task on MER2023 (Lian et al., 2023a), MER2024 (Lian et al., 2024), MELD (Poria et al., 2019), and IEMOCAP (Busso et al., 2008). Each sample is associated with a single majority-voted emotion label. Under the official MER-UniBench protocol, this task is evaluated by Hitrate (HIT). For each wheel w_k, HIT is defined as:

$$\mathrm{HIT}^k = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\left[G_{w_k}(y_i)\in G_{w_k}(\hat{\mathcal{Y}}_i)\right], \tag{40}$$

where y_i is the ground-truth basic emotion label and $\mathbb{I}[\cdot]$ denotes the indicator function. The final HIT score is averaged over the five emotion wheels. Since HIT mainly reflects recall and does not penalize over-prediction, **we additionally report an F1-based evaluation for these four datasets** after the same emotion-wheel normalization.

Sentiment analysis. Sentiment analysis requires the model to determine the polarity of the multimodal input. We evaluate this task on CMU-MOSI (Zadeh et al., 2016), CMU-MOSEI (Zadeh et al., 2018), CH-SIMS (Yu et al., 2020), and CH-SIMS v2 (Liu et al., 2022). Since our model produces free-form emotion words rather than sentiment labels, we follow the MER-UniBench protocol and use Qwen2.5-7B-Instruct to map the extracted emotion words into a sentiment category, selected from {*positive*, *negative*, *neutral*}. We report the Weighted Average F-score (WAF) as the primary metric. The prompt used for this mapping is shown below.

Sentiment Mapping Prompt

Please assume the role of an expert in emotions. We provide a set of emotion words describing a character. Please choose the most likely sentiment from the candidates: [positive, negative, neutral].

B.2 MME-Emotion

MME-Emotion (Zhang et al., 2026) is a holistic benchmark for evaluating emotional intelligence in multimodal large language models. Different from MER-UniBench, which mainly evaluates open-

vocabulary emotion recognition with emotion-wheel-based metrics, MME-Emotion uses LLM-based judging to assess both the final emotion prediction and the reasoning process.

Tasks. MME-Emotion contains eight evaluation tasks: emotion recognition in laboratory scenarios (ER-Lab), emotion recognition in the wild (ER-Wild), emotion recognition under noisy conditions (Noise-ER), fine-grained emotion recognition (FG-ER), multi-label emotion recognition (ML-ER), sentiment analysis (SA), fine-grained sentiment analysis (FG-SA), and intent recognition (IR). These tasks cover both controlled and in-the-wild scenarios, and evaluate different aspects of affective understanding, including categorical emotion recognition, fine-grained affective interpretation, sentiment polarity, and intent inference.

Evaluation metrics. MME-Emotion reports three metrics. *Recognition* measures whether the final predicted emotion or affective label is correct. *Reasoning* evaluates whether the generated explanation is coherent, grounded, and supportive of the final answer. *CoT* is the average of Recognition and Reasoning. Following the benchmark protocol, task-level results are judged by an evaluator model. Since the original GPT-4o evaluator used by the benchmark is no longer available in our setting, we use `gemini-3.1-flash-lite-preview` as the evaluator for all compared methods to ensure a consistent evaluation protocol. In the Table 2, we report the CoT score for each task and the averaged Recognition, Reasoning, and CoT scores across all tasks.

C Implementation Details

We use Qwen2.5-Omni (Xu et al., 2025a) as the backbone for our main experiments and train only its Thinker module. All models are trained on MER-Caption+ (Lian et al., 2025a) with a two-stage SFT-RL pipeline based on GRPO-style optimization. For evaluation, all of our models are reported using the checkpoint saved at the final training step, without dataset-specific checkpoint selection. The original AffectGPT (Lian et al., 2025a) releases checkpoints from multiple training steps and selects the best-performing checkpoint separately for each evaluation dataset, which can result in stronger reported performance. To ensure a consistent evaluation protocol, we report AffectGPT† using its publicly released final checkpoint instead. Since AffectGPT-RL (Lian et al., 2026) is not publicly available, we reproduce AffectGPT-RL† using Qwen2.5-Omni as the backbone while following the SFT/RL data allocation of the original paper, with only 1k samples used for GRPO. On top of this reproduction, we construct a stronger *Baseline* by changing only the SFT/RL data allocation: we randomly sample 5k instances for SFT and use all remaining 26327 instances for RL, while keeping the backbone, training objectives, and other optimization settings unchanged.

For both the *Baseline* and `MER-R1`, we train the SFT stage for 2 epochs with a learning rate of 2×10^{-5}, followed by 1 epoch of RL with a learning rate of 2×10^{-6}. During RL, we sample $G = 4$ responses per prompt and train for 3262 optimization steps, with a batch size of 1 and 2 gradient accumulation steps. We set both λ_{cali} and λ_{fmt} to 0.1. All experiments are conducted on 16 NVIDIA H100 GPUs, and the RL stage takes approximately 18 hours.

D Additional Experiments

D.1 Generalization to MiniCPM-o-4.5

We further demonstrate that `MER-R1` generalizes to other backbones. Specifically, we replace Qwen2.5-Omni with MiniCPM-o-4.5 (Cui et al., 2026) and evaluate SFT, the standard GRPO *Baseline*, and `MER-R1`, while keeping the training data and hyperparameters unchanged.

As shown in Table 6, `MER-R1` consistently outperforms the GRPO *Baseline* across all nine MER-UniBench datasets, improving the overall mean from 78.28 to 82.75. Table 7 further shows that SFT and standard GRPO exhibit the same thinking paradox observed with Qwen2.5-Omni: slow thinking reduces R-Mean by 1.52 and approximately 0.84 points, respectively. In contrast, `MER-R1` reverses this trend, with slow thinking outperforming fast thinking on all five emotion-recognition datasets and improving R-Mean by 1.14 points. These consistent results demonstrate that the slow–fast thinking synergy learned by `MER-R1` generalizes across different omni-modal backbones.

Table 6: **Main results on MER-UniBench with MiniCPM-o-4.5.** `MER-R1` consistently improves over SFT and the standard GRPO baseline across different emotion understanding tasks.

Model	Fine (F1)	Basic Emotion (Hitrate)				Sentiment Analysis (WAF)				Mean
	OV-MERD+	MER23	MER24	MELD	IEMOCAP	MOSI	MOSEI	SIMS	SIMSv2	
SFT	60.84	70.63	71.84	51.99	53.91	79.55	79.25	81.62	83.55	70.35
Baseline	67.18	83.90	83.17	62.43	67.37	82.20	84.07	86.50	87.72	78.28
MER-R1	**70.60**	**89.04**	**91.98**	**71.90**	**76.50**	**84.12**	**84.23**	**88.04**	**88.38**	**82.75**

Table 7: **Slow Thinking vs. Fast Thinking on MiniCPM-o-4.5 under unified F1 metrics.** The four basic emotion recognition datasets are re-evaluated by F1. **R-Mean** averages the five emotion recognition datasets. `MER-R1` makes slow thinking consistently outperform fast thinking.

Model	Mode	Fine (F1)	Basic Emotion (F1)				R-Mean
		OV-MERD+	MER23	MER24	MELD	IEMOCAP	
SFT	Fast	59.86	**54.44**	**61.24**	**41.99**	**43.67**	**52.24**
	Slow	**60.84**	54.39	56.13	40.08	42.16	50.72 ↓
Baseline	Fast	**67.67**	62.88	**68.41**	**47.87**	53.08	**59.98**
	Slow	67.18	**63.37**	64.84	47.11	**53.23**	59.15 ↓
MER-R1	Fast	69.38	64.63	68.91	48.93	55.63	61.50
	Slow	**70.60**	**66.24**	**69.81**	**49.80**	**56.75**	**62.64** ↑

Table 8: **Controlling for prediction cardinality.** Each cell reports mean P/R/F1 (%) over the five emotion-recognition datasets after retaining at most the first K predicted emotion labels. ΔF1 denotes $\mathrm{F1}_{\mathrm{Slow}} - \mathrm{F1}_{\mathrm{Fast}}$.

Model	Mode	$K = 1$	$K = 2$	$K = 3$	$K = 4$	$K = 5$
Baseline	Fast	55.00/50.66/52.40	52.89/67.16/58.61	48.72/74.55/58.38	48.03/76.27/58.39	47.90/76.69/58.42
	Slow	50.09/45.83/47.54	50.38/59.88/54.22	50.88/68.69/57.94	50.02/70.69/58.03	49.86/71.27/58.12
	ΔF1	-4.86	-4.39	-0.44	-0.36	-0.30
MER-R1	Fast	62.24/57.48/59.40	57.04/70.96/62.60	52.25/76.82/61.54	50.14/79.77/60.92	49.23/81.29/60.69
	Slow	59.96/55.27/57.16	57.47/68.63/62.01	54.78/74.27/62.48	52.49/77.77/62.08	51.19/79.64/61.72
	ΔF1	-2.23	-0.58	$+0.94$	$+1.16$	$+1.03$

D.2 Controlling for Prediction Cardinality

One potential explanation for the thinking paradox is that fast thinking achieves higher recall simply because it predicts more emotion labels. **Our analysis shows that the slow–fast discrepancy instead stems from different category-selection behaviors, rather than merely differences in prediction cardinality.** Under matched cardinality constraints, the *Baseline* still exhibits distinct slow–fast decisions, whereas `MER-R1` substantially narrows this gap and ultimately reverses it, better reconciling slow thinking's selectivity with fast thinking's recall advantage.

Table 8 shows that prediction cardinality cannot explain the slow–fast difference. Most notably, at $K = 1$, where both modes are reduced to a single predicted emotion, fast thinking still substantially outperforms slow thinking in precision, recall, and F1 under the *Baseline*. This indicates that the two modes differ in **which emotion categories they prioritize, rather than merely in how many categories they predict**. As K increases, the characteristic trade-off re-emerges: slow thinking becomes more selective, while fast thinking retains higher recall. In contrast, `MER-R1` substantially narrows the slow–fast gap at $K = 1, 2$ and reverses it for $K \geq 3$, indicating that it reconciles their category-selection behaviors rather than simply controlling prediction cardinality.

D.3 Controlling for Thinking Effort

Another potential explanation for the thinking paradox is that an intermediate amount of reasoning, rather than either Fast or full Slow Thinking, may achieve a better precision–recall balance. **Our**

Table 9: **Controlling for thinking effort.** We evaluate the same *Baseline* checkpoint under different maximum thinking-token budgets. P/R/F1 are averaged over the five emotion-recognition datasets.

Method	Avg. Think Tokens	Precision	Recall	F1	R-Mean	Mean
Fast ($K = 0$)	0.00	47.84	**77.10**	59.04	58.50	**71.11**
Short ($K = 64$)	63.93	46.82	75.37	57.76	57.24	70.49
Medium ($K = 128$)	119.13	49.98	72.28	**59.10**	**58.58**	70.62
Long ($K = 256$)	127.18	**49.99**	72.29	**59.10**	58.57	70.68
Baseline Slow	127.19	49.78	71.43	58.67	58.12	70.55
MER-R1 Slow	**130.26**	**50.50**	**81.67**	**62.41**	**61.80**	**73.14**

Table 10: **Sensitivity to λ_{cali} on MER-UniBench.** Performance is stable across different calibration weights, with a slight decrease as λ_{cali} increases.

λ_{cali}	Fine (F1)	Basic Emotion (Hitrate)				Sentiment Analysis (WAF)				Mean
	OV-MERD+	MER23	MER24	MELD	IEMOCAP	MOSI	MOSEI	SIMS	SIMSv2	
0.3	68.96	89.19	**93.18**	71.22	78.15	**86.66**	86.27	86.02	87.61	83.03
0.2	69.41	**91.54**	92.74	71.44	**78.62**	83.78	86.45	**88.56**	87.50	83.34
0.1	**70.68**	88.60	92.96	**71.90**	78.10	86.06	**86.73**	88.04	**88.48**	**83.50**

analysis shows that adjusting thinking effort merely moves the *Baseline* along the same precision–recall trade-off, whereas MER-R1 learns to improve both sides of the trade-off. To test this, we evaluate the same *Baseline* checkpoint with maximum thinking budgets of 0, 64, 128, and 256 tokens, while keeping the prompt and evaluation protocol unchanged.

As shown in Table 9, the clearest transition occurs from 64 to 128 tokens: increasing the thinking budget raises precision from 46.82 to 49.98, but simultaneously reduces recall from 75.37 to 72.28. Consequently, the best intermediate budget improves R-Mean over Fast Thinking by only 0.08 points, while its overall Mean remains lower. In contrast, MER-R1 achieves 61.80 R-Mean and 73.14 overall Mean, substantially exceeding all intermediate-budget variants and also outperforming Fast Thinking. These results show that the gains of MER-R1 cannot be reproduced by manually adjusting reasoning length, but arise from the learned slow–fast synergy.

D.4 Ablation on Slow–Fast Confidence Calibration

We further examine the sensitivity to the slow–fast confidence calibration weight λ_{cali} on MER-UniBench. As shown in Table 10, the overall performance remains stable when λ_{cali} varies from 0.1 to 0.3, with only a slight decrease in the overall Mean from 83.50 to 83.03. We therefore use $\lambda_{\text{cali}} = 0.1$ as the default setting in all experiments.

D.5 Training Dynamics of Dual Rewards

Figure 7 compares the recall and precision reward dynamics of the *Baseline* and MER-R1. MER-R1 achieves a substantially higher recall reward, reaching around 0.75 compared with about 0.67 for the *Baseline*. Meanwhile, both methods maintain similar precision rewards, fluctuating around 0.65. This shows that MER-R1 improves recall without sacrificing precision. The dual-objective design therefore better balances the recall–precision trade-off and preserves the recall-oriented coverage that is often suppressed during slow thinking.

D.6 Retained Hitrate Analysis

We define retained hitrate as the fraction of ground-truth emotion categories that are correctly predicted by fast thinking and still preserved in the final slow-thinking answer. As shown in Figure 8, MER-R1 consistently retains more correct fast-thinking predictions than the *Baseline* across all five recognition benchmarks. For example, the retained hitrate increases from 83.5 to 94.7 on OV-MERD+, from 86.5 to 95.4 on MER2023, and from 89.4 to 97.2 on MER2024, with similar gains on MELD and IEMOCAP. These results show that MER-R1 better preserves useful fast-thinking

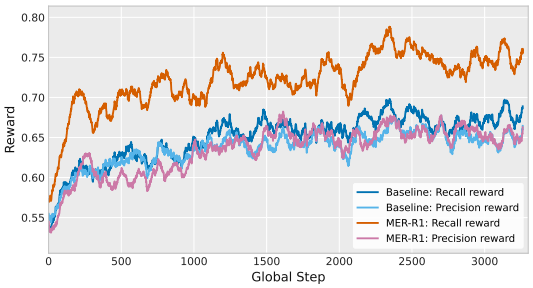


Figure 7: Training dynamics on dual rewards.

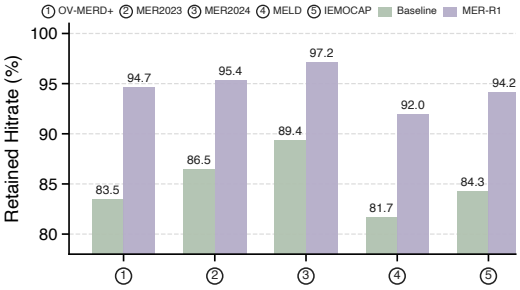


Figure 8: Retained hitrate comparision.

intuition during slow thinking, instead of filtering out correct categories as the *Baseline* often does. This supports the role of slow-fast confidence calibration in selectively retaining recall-oriented signals from fast thinking.

E QUALITATIVE ANALYSIS

We provide qualitative examples to further illustrate how `MER-R1` realizes slow-fast thinking synergy in multimodal emotion recognition. Specifically, we present three types of cases. (i) *Fast-intuition retention* refers to cases where fast thinking already predicts a correct emotion, but the baseline slow-thinking answer removes it after deliberation; `MER-R1` keeps this useful fast-thinking intuition in the final answer. (ii) *Spurious emotion suppression* refers to cases where fast thinking introduces incorrect emotions due to over-coverage; `MER-R1` suppresses these spurious predictions through slow-thinking selectivity. (iii) *Missed emotion recovery* refers to cases where baseline fast thinking fails to cover the ground-truth emotion, while `MER-R1` recovers it after slow reasoning.

For each case type, we show examples from both fine-grained emotion recognition and basic emotion recognition. The fast-intuition retention examples show that `MER-R1` does not simply discard the useful predictions produced by fast thinking; instead, it preserves correct categories such as **angry** or **happy** that are missed by the baseline slow-thinking answer. The spurious emotion suppression examples show that `MER-R1` can remove unsupported emotions such as **angry** or **resentment**, keeping the final prediction focused on the ground-truth emotions. The missed emotion recovery examples further show that `MER-R1` can recover emotions overlooked by baseline fast thinking, such as **surprised**, **worried**, and other positive fine-grained emotions. Together, these examples qualitatively demonstrate that `MER-R1` selectively combines the recall-oriented intuition of fast thinking with the precision-oriented selectivity of slow thinking.

Table 11: Fine-grained emotion recognition example of fast-intuition retention.

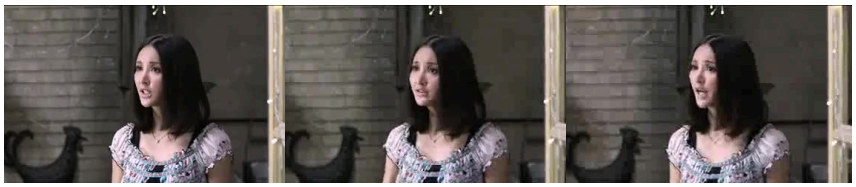
Example of fast-intuition retention	
	
Subtitle (En)	I tell you, it's not easy for a woman who has divorced and has a child to find a partner, it's not as easy as you think.
Ground Truth	Labels: **angry**, dissatisfied
Baseline Fast Thinking	<think> </think> <answer>**angry**, frustrated, anxious, disappointed</answer>
Baseline Slow Thinking	<think>In the text, the subtitle content says, "I'll tell you, it's not easy for a woman who has divorced and is raising a child to find a partner." Based on the woman's relaxed and comfortable posture and facial expression in the video clues, as well as the individual's high level of emotional stress and frustration in the audio clues, we can infer that this sentence may carry a negative or depressed tone. This sentence may be the woman's evaluation or complaint about the current situation or some difficulties. The woman's positive emotions and comfortable posture in the video clues indicate that she may be able to cope with the current situation or solve problems in some way, but based on the individual's negative emotions in the audio clues, we can infer that this sentence may carry a sense of disappointment or frustration. Therefore, this sentence may be the woman's negative evaluation of the current situation, aligning with her overall display of negative emotions.</think> <answer>disappointment, frustration</answer>
MER-R1 Slow Thinking	<think>In the text, the subtitle reads: "It's difficult for a woman to find a partner after divorce." This sentence may be the speaker's response to a certain problem or situation. Based on the audio clues describing the speaker's aggressive tone and high-pitched voice, as well as the audio clue of the speaker expressing feelings of anger or frustration, we can infer that this sentence may carry a tone of complaint or dissatisfaction. The speaker may be angry or dissatisfied about the difficulties or troubles of a woman finding a partner after divorce, hence using this statement to express their emotions.</think> <answer>**angry**, frustrated, anxious, stressed</answer> *Explanation:* Baseline slow thinking misses the ground-truth emotion **angry**, while MER-R1 slow thinking retains it.

Table 12: Basic emotion recognition example of fast-intuition retention.

Example of fast-intuition retention	
	
Subtitle (En)	Thank you all, thank you for not treating me as an outsider. Actually, before this, I thought I only had one friend in my life, Old Tan. I didn't expect to meet you all on the 22nd floor.
Ground Truth	Labels: **happy**
Baseline Fast Thinking	<think> </think> <answer>**happy**, surprised, grateful</answer>
Baseline Slow Thinking	<think>In the text, the subtitle reads: "Thank you all, thank you for not treating me as an outsider. Actually, before this, I thought I only had one friend in my life, Old Tan, but I didn't expect to meet you all on the 22nd floor." This sentence may be the woman expressing gratitude to the people around her. Based on the audio clue of the woman's voice being filled with excitement and enthusiasm, as well as the video clue of the woman smiling and laughing while driving, we can infer that this sentence may carry a tone of excitement and gratitude. The woman may be surprised and grateful for the people around her, feeling that she has made more friends than she thought. Therefore, this sentence aligns with the overall positive emotions displayed by the woman.</think> <answer>excitement, gratitude, surprise</answer>
MER-R1 Slow Thinking	<think>In the text, the subtitle reads: "Thank you all, thank you for not treating me as an outsider. Actually, before this, I thought I only had one friend, Mr. Tan. But I didn't expect to meet you all on the 22nd floor." This sentence might be a woman's expression of gratitude and excitement to those around her. Based on the audio clues, with the woman's voice having a high pitch and slightly faster pace, as well as the video clues showing the woman's happiness and excitement, we can infer that this sentence might be delivered with a positive, appreciative, and joyful tone. The woman might be sharing her happiness and gratitude with those around her, while also feeling unexpected and fortunate for meeting new friends.</think> <answer>**happy**, grateful, appreciative, positive</answer> *Explanation:* Baseline slow thinking misses the ground-truth emotion **happy**, while MER-R1 slow thinking retains it.

Table 13: Fine-grained emotion recognition example of spurious emotion suppression.

Example of spurious emotion suppression	
	
Subtitle (En)	You would feel better this way, wouldn't you?
Ground Truth	Labels: concerned, negative, sad
Baseline Fast Thinking	<think> </think> <answer>**angry**, frustrated, anxious, sad, disappointed</answer>
MER-R1 Slow Thinking	<think>In the text, the subtitle reads, "Can you feel better this way?" This sentence may be a woman's response or concern for someone. Based on the audio clue describing the woman's crying and heavy breathing, as well as the video clue of the woman's sad and crying facial expression, we can infer that this sentence may carry a tone of concern or comfort. The woman may be responding to someone's distress or negative emotions, hoping that the other person can feel better or cope with the situation. Therefore, this sentence may express the woman's concern and care for the other person.</think> <answer>concerned, worried, sad, anxious</answer> *Explanation:* Baseline fast thinking predicts the spurious emotion **angry**, whereas MER-R1 slow thinking suppresses it and keeps the prediction focused on ground-truth emotions.

Table 14: Basic emotion recognition example of spurious emotion suppression.

Example of spurious emotion suppression	
	
Subtitle (En)	Grandpa.
Ground Truth	Labels: sad
Baseline Fast Thinking	<think> </think> <answer>anxiety, nervousness, worry, confusion, frustration, disappointment, stress, sadness, **anger**, **resentment**, urgency, despair, panic</answer>
MER-R1 Slow Thinking	<think>In the text, the caption reads: "Grandpa." This phrase might be a man's way of addressing someone or a title. Based on the video hints, with the man's facial expression showing contemplation and determination, and his body language suggesting relaxation, we can infer that this statement could be the man expressing concern, worry, or asking for help or assistance to his grandfather. The audio clues describing his voice with a slight tremble, at a slightly faster pace, and at a higher pitch, suggest he might be experiencing intense emotions such as anxiety, excitement, or stress. Therefore, this sentence could be the man expressing his emotional distress or seeking help.</think> <answer>anxiety, sadness, concern, worry</answer> *Explanation:* Baseline fast thinking predicts the spurious emotion **anger**, **resentment**, whereas MER-R1 slow thinking suppresses it and keeps the prediction focused on ground-truth emotions.

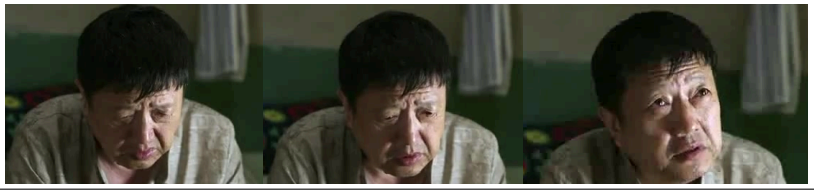
Table 15: Fine-grained emotion recognition example of missed emotion recovery.

Example of missed emotion recovery	
	
Subtitle (En)	It's hard to see it.
Ground Truth	Labels: relaxed, shy, happy, teasing
Baseline Fast Thinking	<think> </think> <answer>excited, anxious, frustrated, stressed, stressed out</answer>
MER-R1 Slow Thinking	<think>In the text, caption content: "It's hard to see it." This sentence might be a woman's response or expression towards the man. Based on the audio clues, which include a woman's laughter and her voice filled with excitement and anticipation, it can be inferred that the woman is experiencing positive emotions and is comfortable around the man. The video hints at the woman's comfortable and relaxed state. From the woman's voice in the audio clues being high-pitched and enthusiastic, it can be speculated that she is experiencing positive emotions and might be responding to the man's words or actions in a playful or teasing manner. Therefore, this sentence could be a woman's sarcastic or humorous way of responding, fitting well with her overall positive emotions.</think> <answer>happy, content, playful, teasing, amused</answer> *Explanation:* Baseline fast thinking misses all ground-truth emotions, whereas MER-R1 slow thinking recovers the correct positive emotions, including **happy**, **content**, and **teasing**.

Table 16: Fine-grained emotion recognition example of missed emotion recovery.

Example of missed emotion recovery	
	
Subtitle (En)	Ah! We won't see each other on Sunday?
Ground Truth	Labels: surprised
Baseline Fast Thinking	<think> </think> <answer>disappointment, frustration, anxiety</answer>
MER-R1 Slow Thinking	<think>In the text, the subtitle reads: "Ah, we're not going to see each other on Sunday." Based on the audio clues, which describe the speaker's voice with a high pitch and excited tone, along with the speaker's use of "ah" to express emotional distress, we can infer that this sentence might convey a sense of anxiety or surprise. The speaker's high pitch and excited tone suggest that they are experiencing strong emotions, while the "ah" in the subtitle could be the expression of surprise or confusion. Therefore, this sentence might reflect the speaker's concern or disappointment over the lack of a Sunday meeting.</think> <answer>confused, anxious, surprised, frustrated</answer> *Explanation:* Baseline fast thinking misses the ground-truth emotion **surprised**, while MER-R1 slow thinking successfully recovers this missed emotion in its final prediction.

Table 17: Basic emotion recognition example of missed emotion recovery.

Example of missed emotion recovery	
	
Subtitle (En)	Are you afraid that my presence here will affect your interaction with Cao Xiaoqiang?
Ground Truth	Labels: worried
Baseline Fast Thinking	<think> </think> <answer>angry, anxious, frustrated</answer>
MER-R1 Slow Thinking	<think>In the text, the caption reads: "Are you really that afraid of me here." This sentence might be a man's response to someone's concern or unease. Based on the audio cues describing a personal voice with a flat and monotonous tone, a higher pitch and slightly more aggressive nature, as well as the video clues showing the man's serious and thoughtful expression, we can infer that the statement might convey a tone of worry or concern. The man might feel uneasy or stressed because of the other person's question or response, thus this sentence could be his way of responding to express his own worry or anxiety.</think> <answer>concerned, fearful, worried, anxious</answer> *Explanation:* Baseline fast thinking misses the ground-truth emotion **worried**, whereas MER-R1 slow thinking recovers precise emotions such as **worried** and **concerned**, with other predictions remaining contextually plausible.