RLE-REASONER: BUILDING TRUSTWORTHY EMOTION REASONING FROM COMPLEMENTARY EEG/EOG EVIDENCE THROUGH COUNTERFACTUAL DIAGNOSIS AND DIAGNOSIS-GUIDED REPAIR

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal large language models (MLLMs) can generate natural-language explanations for emotion judgments, yet fluent reasoning may describe physiological patterns absent from the recorded signals. We propose **RLE-Reasoner**, a build–diagnose–repair framework that connects emotion reasoning to verifiable EEG/EOG response evidence. In Stage 1, **RLE-MLLM** combines separate EEG/EOG encoders, Complementarity-Aware Response Fusion (CARF), and a physiological prefix to predict seven-class emotions and generate structured physiological evidence with a natural-language explanation. Seven attributes independently extracted from raw signals supervise the structured EEG/EOG evidence fields without entering the model input; their verbalization supplies the explanation target. In Stage 2, **RLE-Eval** uses controlled replacement of real EEG/EOG recordings to test whether the generated evidence reflects the current signals. In Stage 3, **RLE-PostTrain** turns these diagnoses into GRPO rewards that encourage correct physiological descriptions and responses to changed evidence. On a 5,040-trial, twelve-subject dataset with 1,260 held-out test trials, RLE-MLLM 14B achieves **66.52%** seven-class cross-subject emotion accuracy and **62.08%** matched physiological-state accuracy. Post-training improves overall physiological-grounding accuracy from **63.84%** to **74.15%** and state Macro-F1 from **46.21%** to **58.10%**, with matched emotion accuracy changing from 66.52% to 66.66%. Blinded human assessment of free-form text against supplied physiological reference states finds that consistency increases from **58.33%** to **69.06%**. These results show that physiological evidence can guide the diagnosis and repair of generated emotion reasoning.

1 INTRODUCTION

Multimodal large language models (MLLMs) predict emotions and generate natural-language explanations from visual, audio, and textual cues (Cheng et al., 2024; Lian et al., 2025). A plausible explanation, however, may describe evidence absent from the response being analyzed; even a correct emotion label does not establish that its explanation is faithful (Xing et al., 2025; Rha et al., 2026). We call an explanation's unsupported physiological claims *hallucinated emotional reasoning*.

EEG captures neural responses to emotional stimuli, while EOG records ocular activity during visual engagement. The modalities are complementary, but most physiological emotion models use their signals for classification rather than verifying the content of generated explanations (Ma et al., 2026; Jiang et al., 2024; **?**). In particular, a language model's fluency or agreement with a predicted label cannot establish that its description follows the recorded signals. This motivates three linked questions: can EEG/EOG support both classification and interpretable evidence generation, can controlled signal changes expose unsupported descriptions, and can the resulting diagnoses improve generation?

We propose **RLE-Reasoner**, a build–diagnose–repair framework. **RLE-MLLM** combines separate EEG/EOG encoders with Complementarity-Aware Response Fusion (CARF) and generates a seven-

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.

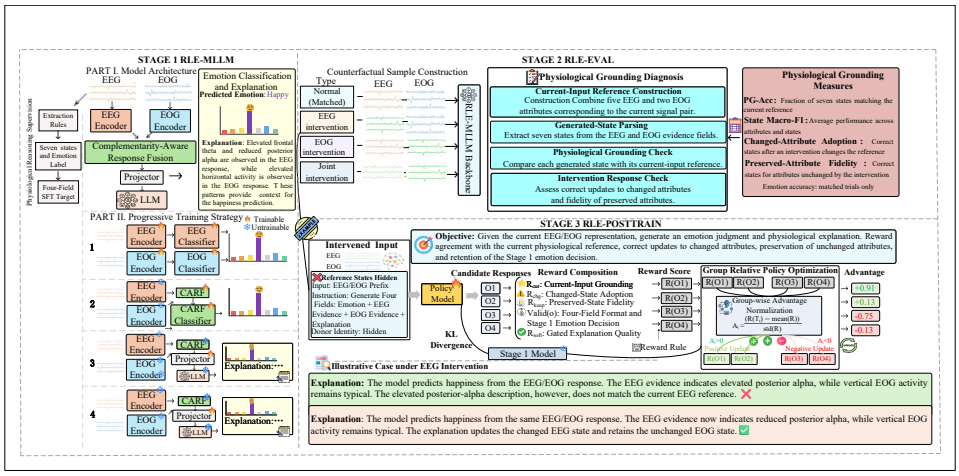


Figure 1: RLE-Reasoner Framework.

class emotion judgment, seven structured physiological states, and an explanation. A fixed, label-independent extractor supplies state targets from raw signals without entering the model input. **RLE-Eval** replaces real EEG and/or EOG recordings and checks whether the generated states follow the signals currently provided. **RLE-PostTrain** uses these checks as rewards to improve physiological grounding while monitoring matched-input classification. Figure 1 summarizes the pipeline.

Contributions. (1) RLE-MLLM jointly generates emotion labels and structured EEG/EOG states supervised by an independent signal extractor, reaching 66.52% seven-class accuracy and 62.08% matched state accuracy with 14B. (2) RLE-Eval diagnoses unsupported structured evidence using matched, EEG, EOG, and joint interventions on recorded signals. (3) RLE-PostTrain improves 14B overall physiological-grounding accuracy from 63.84% to 74.15% and State Macro-F1 from 46.21% to 58.10%; blinded human assessment finds higher evidence consistency in free-form explanations.

2 Related Work

Physiological emotion understanding. Emotion-LLaMA and AffectGPT generate emotion descriptions from external modalities (Cheng et al., 2024; Lian et al., 2025). EEG–language models such as NeuroLM and E^2-LLM connect brain signals with language (Jiang et al., 2024; Ma et al., 2026); EEG/EOG fusion has also improved discriminative emotion recognition (**?**). Our work instead supervises seven explicit EEG/EOG states in the generated response and evaluates the accompanying prose separately.

Diagnosis and repair. EmotionHallucer examines emotion-related hallucinations, and Emotion-Coherent Reasoning rewards rationale–label consistency (Xing et al., 2025; Rha et al., 2026). Such checks do not test whether generated EEG/EOG states follow controlled changes to the recorded signals. GRPO enables group-relative reward optimization without a value model (Shao et al., 2024); here, its reward is anchored to independently extracted physiological states rather than linguistic plausibility alone.

3 Method

3.1 Response–Evidence Formulation

For subject u, affective condition c, and trial t, the response is $r_{u,c,t} = (\mathrm{EEG}_{u,c,t}, \mathrm{EOG}_{u,c,t})$, with stimulus category y_c and reference attributes $\mathbf{a}^*_{u,c,t} = \mathcal{E}_{\mathrm{phys}}(r_{u,c,t})$. The category labels the facial-expression stimulus, not directly the participant's subjective emotion; the stimulus is never shown to the model. The fixed, label-independent $\mathcal{E}_{\mathrm{phys}}$ yields five EEG and two EOG states from raw recordings. Only the response and task instruction enter the model; y_c and $\mathbf{a}^*$ are supervision/evaluation

targets. At inference, RLE-MLLM produces $(\hat{y}, \hat{\mathbf{a}}, \hat{q}) = M_\theta(r)$: a category, seven states, and free-form explanation. We measure physiological grounding by comparing $\hat{\mathbf{a}}$ with the reference for the *current* input, and assess the prose separately.

3.2 RLE-REASONER

3.2.1 OVERVIEW

RLE-MLLM builds response-based emotion explanations, RLE-Eval diagnoses unsupported structured evidence through controlled signal replacement, and RLE-PostTrain uses those diagnoses for repair (Figure 1).

Stage 1: Complementarity-Aware RLE-MLLM. For $r_i = (\mathrm{eeg}_i, \mathrm{eog}_i)$, a pretrained REVE encoder (El Ouahidi et al., 2025) yields h_i^{e}, while multiscale 1D convolutions and a temporal Transformer encode the 4×1000 EOG trial as h_i^{o}. Each branch has a seven-class cross-entropy head. CARF maps the branches to $z_i^m = P_m(h_i^m)$, $m \in \{\mathrm{e}, \mathrm{o}\}$, and constructs discrepancy $d_i = |z_i^{\mathrm{e}} - z_i^{\mathrm{o}}|$ and shared information $s_i = z_i^{\mathrm{e}} \odot z_i^{\mathrm{o}}$:

$$\begin{aligned} g_i &= \sigma(G([z_i^{\mathrm{e}}; z_i^{\mathrm{o}}; d_i; s_i])), \\ z_i^{\mathrm{f}} &= g_i \odot z_i^{\mathrm{e}} + (1 - g_i) \odot z_i^{\mathrm{o}}, \\ u_i &= [z_i^{\mathrm{e}}; z_i^{\mathrm{o}}; z_i^{\mathrm{f}}; d_i]. \end{aligned} \tag{1}$$

The unimodal and CARF heads are trained with seven-class cross-entropy; their logits combine as $\ell_i = \alpha \ell_i^{\mathrm{e}} + \beta \ell_i^{\mathrm{o}} + \gamma C_{\mathrm{carf}}(u_i)$. We first train the unimodal encoders, freeze them to train CARF, and then freeze both encoders and CARF. A two-layer projector maps the *complete* u_i to $p_i = P_{\mathrm{LLM}}(u_i) \in \mathbb{R}^{K \times d_{\mathrm{LLM}}}$. We align the projector with Qwen LoRA frozen, then jointly update the projector and LoRA; the final four-field prediction is generated by Qwen rather than by the upstream classification heads.

Because emotion labels do not specify physiological explanations, the fixed extractor $E_{\mathrm{phys}}(r_i)$ supplies five EEG states (N170 amplitude, LPP amplitude, frontal theta, posterior alpha, frontal alpha asymmetry) and two EOG states (horizontal and vertical RMS activity). N170 and LPP use baseline-corrected amplitudes at 140–220 ms and 300–800 ms; theta/alpha use relative band power. Attributes are computed before encoder normalization. Training-subject medians m_k and MADs set $z_{i,k} = (v_{i,k} - m_k)/(1.4826\,\mathrm{MAD}_k)$; thresholds at ± 0.75 produce reference-relative reduced, typical, or elevated states (N170: more/less negative; asymmetry: lower/higher). These statistics are frozen for validation and testing; the states are not clinical labels.

Each target has *Emotion*, *EEG evidence*, *EOG evidence*, and *Explanation*. The label supplies Emotion, the extractor supplies seven states, and offline Qwen-Max verbalizes the fixed label/states without seeing raw signals. Neither target labels nor states enter the model input. Qwen generates the whole target o_i^* from the learned prefix p_i and instruction q_i using

$$\mathcal{L}_{\mathrm{SFT}} = -\sum_{i,\tau} \log p_\theta(o_{i,\tau}^* \mid o_{i,<\tau}^*, q_i, p_i), \tag{2}$$

with loss only on response tokens. Neither extractor nor Qwen-Max is used at inference; interventions begin only in Stage 2.

Stage 2: Counterfactual Diagnosis with RLE-Eval. Stage 2 introduces **RLE-Eval**, a controlled physiological intervention protocol for diagnosing hallucinated emotional reasoning without further model training. Here, a hallucinated physiological description is a generated structured state inconsistent with the reference derived from the EEG/EOG response currently provided to the model. This attribute-level diagnosis tests whether the structured evidence follows signal changes; it does not establish causal neural mechanisms of emotion or verify every sentence of free-form prose.

For each held-out trial $r_i = (\mathrm{eeg}_i, \mathrm{eog}_i)$, we select another real held-out trial $r_j = (\mathrm{eeg}_j, \mathrm{eog}_j)$ as the replacement source and construct four conditions:

The replacement trial is selected from a different subject and stimulus group and must pass the same signal-quality checks as the original trial. Selection does not use the emotion label. Instead, the replacement trial must differ from the original trial in at least one independently extracted attribute

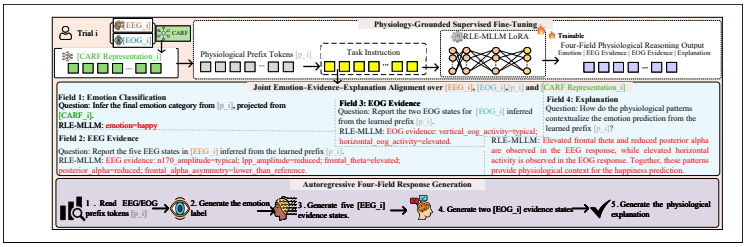


Figure 2: Physiology-grounded instruction tuning in RLE-MLLM. CARF retains modality-specific, fused, and discrepancy representations, which are mapped by the projector into EEG/EOG prefix tokens. During training, independently extracted physiological attributes provide target-side supervision; at inference, Qwen generates the complete four-field response directly from the learned prefix.

of the replaced modality, or across the seven attributes for the joint intervention. This requirement ensures an observable change in the physiological reference rather than assuming that different emotion labels necessarily correspond to different EEG/EOG responses. Single-modality mixtures are diagnostic inputs rather than naturally recorded trials.

For every condition, we construct the corresponding physiological reference using the modality-specific attributes defined in Stage 1:

$$\begin{aligned} a_i^{\text{match}*} &= (a_i^{\text{eeg}*}, a_i^{\text{eog}*}), \\ a_{i\leftarrow j}^{\text{eeg}*} &= (a_j^{\text{eeg}*}, a_i^{\text{eog}*}), \\ a_{i\leftarrow j}^{\text{eog}*} &= (a_i^{\text{eeg}*}, a_j^{\text{eog}*}), \\ a_j^{\text{joint}*} &= (a_j^{\text{eeg}*}, a_j^{\text{eog}*}). \end{aligned} \tag{3}$$

Thus, the reference always corresponds to the EEG and EOG recordings actually provided under the current condition.

For each condition, the Stage 1 model generates the same four-field response: *Emotion*, *EEG evidence*, *EOG evidence*, and *Explanation*. We parse five EEG states and two EOG states from the structured evidence fields and compare them with the current physiological reference. For N evaluated inputs, physiological- grounding accuracy (PG-Acc) is

$$\text{PG-Acc} = \frac{1}{7N} \sum_{i=1}^{N} \sum_{k=1}^{7} \mathbb{I}[\hat{a}_{i,k} = a_{i,k}^{*}], \tag{4}$$

where an absent or invalid state receives no credit. With seven required states per response, this is the overall state accuracy reported in Stage 1 and the matched-input PG-Acc reported here. We also compute State Macro-F1 by averaging the state-wise F1 scores within each attribute and then across the seven attributes. This separates recovery of less frequent states from gains driven by a frequent reference state. All intervention results use the reference corresponding to the EEG/EOG input under that condition.

Emotion Accuracy and Macro-F1 are evaluated only on matched trials, for which the original emotion label remains valid. No counterfactual emotion label is assigned to mixed-modality interventions. The free-form *Explanation* is evaluated separately; it does not enter the deterministic PG-Acc.

Table 1: Controlled physiological interventions in RLE-Eval.

Condition	Physiological input
Matched	$(\text{eeg}_i, \text{eog}_i)$
EEG intervention	$(\text{eeg}_j, \text{eog}_i)$
EOG intervention	$(\text{eeg}_i, \text{eog}_j)$
Joint intervention	$(\text{eeg}_j, \text{eog}_j)$

RLE-Eval thus diagnoses the structured physiological content of generated reasoning against the current EEG/EOG reference, without treating the whole explanation as verified.

Stage 3: Physiology-Constrained GRPO for Reasoning Repair. RLE-PostTrain initializes its policy from the Stage 1 model and applies GRPO to reduce unsupported structured physiological evidence in the generated explanations. Training instances follow the four RLE-Eval conditions introduced in Stage 2 and are constructed exclusively from the training subjects. The physiological encoders, CARF, and projector remain frozen, and only the Qwen LoRA policy is updated.

For every rollout, the policy receives the same learned EEG/EOG prefix and task instruction as in Stages 1–2. The independently extracted physiological attributes and intervention metadata are hidden from the policy and used only to compute rewards. For each input, we retain the emotion category generated by the Stage 1 checkpoint as a reference for the reward. The Stage 3 policy generates its own emotion category; deviations from that reference are penalized during training rather than prevented at inference.

For a generated response o, let R_{cur} denote its seven-state accuracy against the physiological reference corresponding to the current EEG/EOG input, with invalid or missing states scored as incorrect. For an intervention case, let $\mathcal{C}$ and $\mathcal{U}$ denote the sets of attributes whose reference states are changed and preserved relative to the corresponding matched input, respectively. Given the generated state $\hat{a}_k$ and current reference state a_k^*, we define

$$R_{\mathrm{chg}} = \frac{1}{|\mathcal{C}|} \sum_{k \in \mathcal{C}} \mathbb{I}[\hat{a}_k = a_k^*], \qquad R_{\mathrm{keep}} = \frac{1}{|\mathcal{U}|} \sum_{k \in \mathcal{U}} \mathbb{I}[\hat{a}_k = a_k^*]. \tag{5}$$

The intervention construction guarantees $\mathcal{C} \neq \varnothing$. If $\mathcal{U}$ is empty, its term is omitted and the remaining reward weights are renormalized. The deterministic physiological reward is

$$R_{\mathrm{phys}} = \begin{cases} R_{\mathrm{cur}}, & \text{matched}, \\ 0.60R_{\mathrm{cur}} + 0.30R_{\mathrm{chg}} + 0.10R_{\mathrm{keep}}, & \text{intervention}. \end{cases} \tag{6}$$

The changed-attribute term encourages the structured physiological evidence to follow the intervened modality, while the preserved-attribute term discourages unnecessary revisions to unaffected evidence.

We impose a strong penalty in the reward. A candidate is valid only if it follows the four-field output format and its emotion field matches the Stage 1 reference decision for that input. Invalid candidates receive a reward lower than any valid candidate:

$$R(o) = \begin{cases} R_{\mathrm{phys}} + \lambda_{\mathrm{soft}} \mathbb{I}[R_{\mathrm{cur}} = 1] R_{\mathrm{soft}}, & \mathrm{Valid}(o) = 1, \\ -1 + 0.1R_{\mathrm{phys}}, & \mathrm{Valid}(o) = 0, \end{cases} \tag{7}$$

where $\lambda_{\mathrm{soft}} = 0.02$ and $R_{\mathrm{soft}} \in \{0, 0.5, 1\}$. The soft reward is activated only when all seven structured physiological states are correct. Qwen-Max then evaluates whether the free-form explanation clearly organizes the EEG and EOG evidence, avoids causal overstatement, and handles conflicting or uncertain evidence appropriately. It does not determine the physiological states and cannot compensate for an incorrect state, invalid format, or changed emotion decision. Accordingly, the deterministic reward evaluates the structured EEG/EOG evidence fields, while the free-form *Explanation* field enters training only through this gated soft reward.

For each input, the policy samples G responses. Their rewards are normalized within the group to obtain the sequence-level advantage

$$\hat{A}_g = \frac{R_g - \bar{R}}{\sqrt{\frac{1}{G} \sum_{j=1}^{G} (R_j - \bar{R})^2 + \epsilon}}. \tag{8}$$

For a group with zero reward variance, we set $\hat{A}_g = 0$. Let $c_{g,\tau} = (o_{g,<\tau}, q, p)$ denote the generation context and

$$\rho_{g,\tau} = \frac{\pi_\theta(o_{g,\tau} \mid c_{g,\tau})}{\pi_{\mathrm{old}}(o_{g,\tau} \mid c_{g,\tau})} \tag{9}$$

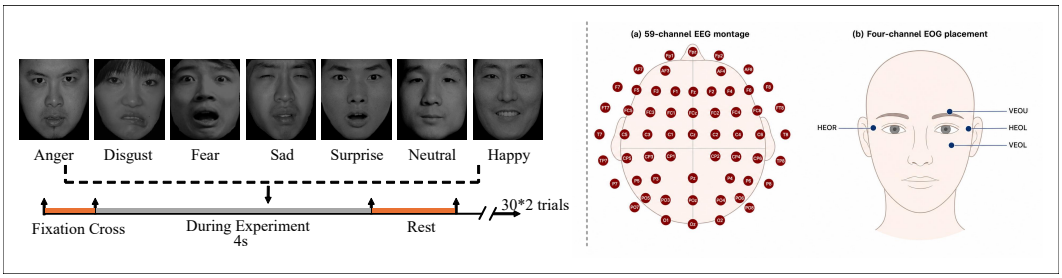


Figure 3: Experimental paradigm and EEG/EOG acquisition setup. Left: facial-expression stimuli and the trial timeline. Right: (a) the 59-channel EEG montage and (b) the four-channel EOG electrode placement.

denote the policy ratio. Because each sampled group is used for a single on-policy update, we optimize the following KL-regularized group-relative objective without additional PPO-style clipping:

$$\mathcal{J}_{\mathrm{GRPO}}(\theta) = \mathbb{E}_{g,\tau}\left[\rho_{g,\tau}\hat{A}_g - \beta D_{\mathrm{KL}}^{g,\tau}\right], \tag{10}$$

where $D_{\mathrm{KL}}^{g,\tau}$ regularizes the policy toward the frozen Stage 1 reference model.

After post-training, the model is evaluated on the same held-out RLE-Eval inputs using the deterministic Stage 2 protocol. Qwen-Max provides only the gated auxiliary training score; it does not score test-set physiological grounding. We evaluate matched-input emotion accuracy separately because the reward penalty does not fix the policy's generated emotion category. RLE-PostTrain thus converts the errors diagnosed in Stage 2 into optimization signals while checking empirically whether emotion classification is maintained.

Human evaluation of free-form explanations. To assess the *Explanation* field beyond the seven structured states, we select 25 held-out base trials and evaluate each under the four RLE-Eval conditions, yielding 100 paired inputs. For each input, the Stage 1 SFT and RLE-PostTrain explanations are presented in random order without model identities, together with the seven reference states corresponding to the current EEG/EOG input. Three reviewers independently assess whether each explanation is consistent with those states, contains unsupported physiological content, and clearly communicates the evidence. Their judgments on free-form text are reported separately from deterministic PG-Acc.

4 Experiments and Results

4.1 Dataset and Experimental Protocol

We use a private visual-stimulus-evoked EEG/EOG dataset from twelve subjects in seven balanced categories (anger, disgust, fear, sadness, surprise, neutral, happiness). Each trial has 1–1.5 s fixation, 4 s facial-expression stimulation, and 2 s rest; the stimulus defines the category but is never provided to the model. We record 59-channel EEG and four-channel EOG (horizontal outer-canthi HEOR/HEOL and vertical VEOU/VEOL) at 1,000 Hz (Figure 3). Each subject has 30 stimuli shown twice per category: 420 trials, or 5,040 overall. Recordings are downsampled to 500 Hz, segmented from -0.5 to 1.5 s into 59×1000 EEG and 4×1000 EOG tensors, and normalized channel-wise per trial.

We assign eight subjects (3,360 trials) to training, one (420) to validation, and three (1,260) to testing. Test subjects are excluded from training, model selection, physiological-reference estimation, and reward construction; all reference thresholds are estimated only on training subjects and frozen. Every held-out trial is tested under four conditions, giving 5,040 test inputs.

(a) Seven-Class Emotion Classification

Ours					Baselines				
Method	**Acc.**	**Macro-F1**	**UAR**	**Kappa**	**Method**	**Acc.**	**Macro-F1**	**UAR**	**Kappa**
EEG-only	56.90	56.72	56.64	0.50	E^2-LLM-style 14B	61.10	61.30	61.05	0.55
EOG-only	42.38	42.16	42.25	0.33	EEGNet	58.81	58.61	58.70	0.52
CARF	62.15	61.27	62.18	0.56	EmT	57.62	56.71	57.00	0.51
Projector	62.40	61.38	62.32	0.56	LGF-EEG	58.52	58.61	58.55	0.52
RLE-MLLM 4B	57.95	58.33	58.10	0.51	NeuroLM	60.51	61.52	61.00	0.54
RLE-MLLM 8B	62.50	62.27	62.83	0.56	MAET	59.76	59.48	59.61	0.53
RLE-MLLM 14B	**66.52**	**66.52**	**66.31**	**0.61**	FAT	60.28	60.06	60.14	0.54

(b) Physiological Grounding of Generated Emotion Reasoning

Model	**Format Acc.**	**N170**	**LPP**	**Frontal Theta**	**Posterior Alpha**	**FAA**	**Horizontal EOG**	**Vertical EOG**	**Overall State Acc.**
RLE-MLLM 4B	100.00	42.54	52.11	45.76	33.65	48.63	42.56	48.53	44.83
RLE-MLLM 8B	100.00	51.67	60.24	51.55	36.90	55.24	53.10	54.52	51.89
RLE-MLLM 14B	100.00	65.42	72.44	57.21	46.52	63.24	66.52	63.24	62.08

Table 2: Stage 1 emotion classification and physiologically grounded emotion reasoning (%). Panel (a) reports seven-class cross-subject emotion recognition on three held-out test subjects. Panel (b) reports response format and seven independently extracted EEG/EOG state accuracies on 1,260 held-out trials. FAA denotes frontal alpha asymmetry.

4.2 Stage 1: Emotion Classification and Physiological Reasoning

Table 2 compares the subject-disjoint results with retrained E^2-LLM-style 14B, NeuroLM, EEGNet, EmT, LGF-EEG, MAET, and FAT baselines (Ma et al., 2026; Jiang et al., 2024; Lawhern et al., 2018; Ding et al., 2025; Zhou et al., 2026; Jiang et al., 2025; Wang et al., 2025).

Method	**Split-1**		**Split-2**		**Split-3**		**Split-4**		**Split-5**		**Split-6**		**Avg.**	**Std.**	**Worst**
	U1	**U2**	**U3**	**U4**	**U5**	**U6**	**U7**	**U8**	**U9**	**U10**	**U11**	**U12**			
EEG-only	54.05	53.40	55.04	59.33	58.33	55.25	58.75	60.84	53.21	60.81	57.75	60.10	57.24	2.89	53.21
EOG-only	38.50	39.40	47.02	41.12	38.81	38.85	45.98	43.66	45.59	44.86	45.66	42.75	42.68	3.21	38.50
CARF	59.83	63.47	64.95	62.31	62.51	61.58	62.16	62.24	60.65	61.19	62.10	62.25	62.10	1.30	59.83
Projector	60.87	61.58	62.04	61.52	63.24	62.12	63.33	62.66	63.68	62.95	62.43	62.00	62.37	0.84	60.87
RLE-MLLM 14B	**64.02**	**63.55**	**65.11**	**64.87**	**65.48**	**66.06**	**66.42**	**67.18**	**68.03**	**66.78**	**67.15**	**66.15**	**65.90**	**1.34**	**63.55**

Table 3: Held-out-subject accuracy of representative Stage 1 variants across six additional subject-disjoint robustness runs (%). Each run evaluates two unseen subjects; U1–U12 denote the twelve held-out subject evaluations. Avg., sample Std., and Worst aggregate these twelve accuracies, separately from the primary 8/1/3 split.

CARF raises accuracy from 56.90% (EEG only) to 62.15%, showing EEG/EOG complementarity. RLE-MLLM 14B reaches 66.52% accuracy and Macro-F1 on held-out matched trials, 5.42 points above the parameter-matched E^2-LLM-style baseline. Across the separate six-run robustness evaluation, its mean subject accuracy is 65.90% (SD 1.34; worst 63.55; Table 3). The 4B model underperforms CARF, while 8B and 14B exceed it, indicating a capacity-dependent difficulty in joint classification and evidence generation.

All three models produce parseable four-field responses, yet their seven-state accuracy rises from 44.83% (4B) to 51.89% (8B) and 62.08% (14B). The 14B model attains 72.44% on LPP but only 46.52% on posterior alpha. Format compliance therefore does not establish physiological correctness; these figures score structured states rather than all free-form sentences.

4.3 Stage 2: EEG/EOG Intervention Evaluation of Hallucinated Emotional Reasoning

RLE-Eval replaces real EEG and/or EOG recordings, rather than adding noise. Each of the 1,260 held-out trials is tested as matched, EEG intervention, EOG intervention, and joint intervention

(5,040 test inputs total). Generated states are always compared with independently extracted attributes of the *current* signals.

Model	Physiological-Grounding Accuracy					State Macro-F1
	Matched	EEG Int.	EOG Int.	Joint Int.	Overall	Overall
RLE-MLLM 4B	44.83	51.34	52.74	52.00	50.23	34.73
RLE-MLLM 8B	51.89	57.52	58.42	57.21	56.26	39.12
RLE-MLLM 14B	62.08	65.36	64.15	63.77	63.84	46.21

Table 4: Intervention-based evaluation of hallucinated emotional reasoning on all 1,260 held-out trials (1,260 inputs per condition; %). Physiological-grounding accuracy measures the fraction of generated physiological states supported by the EEG/EOG signals currently provided to the model. State Macro-F1 macro-averages the reference-relative states across the seven attributes.

Table 4 reports overall PG-Acc of 50.23%, 56.26%, and 63.84% for 4B, 8B, and 14B. The matched values equal Stage 1 state accuracies because inputs and scoring are identical. Intervention accuracy need not decline: replacement also changes the reference. The diagnostic test is whether generated states agree with that current reference; lower State Macro-F1 reveals uneven recovery among reference states.

4.4 Stage 3: Physiology-Constrained Post-Training

RLE-PostTrain expands 3,360 training-subject trials to 13,440 condition-specific instances, using training-subject donors and references only. Table 5 compares Stage 1 SFT, GRPO without counterfactual (CF) terms, and full RLE-PostTrain on the same 5,040 held-out inputs.

Model	Training	PG-Acc.	PG-MacroF1	Changed Adoption	Preserved Fidelity	Matched Emotion Acc.
RLE-MLLM 4B	Stage 1 SFT	50.23	34.73	41.65	51.88	57.95
	GRPO w/o CF terms	56.42	36.52	42.15	52.05	58.05
	RLE-PostTrain	**60.05**	**42.81**	**48.68**	**57.66**	**58.11**
RLE-MLLM 8B	Stage 1 SFT	56.26	39.12	45.79	55.82	62.50
	GRPO w/o CF terms	58.25	44.09	46.43	55.71	62.50
	RLE-PostTrain	**65.22**	**47.12**	**54.21**	**61.91**	**62.66**
RLE-MLLM 14B	Stage 1 SFT	63.84	46.21	51.97	59.12	66.52
	GRPO w/o CF terms	68.00	50.55	53.45	61.11	**66.87**
	RLE-PostTrain	**74.15**	**58.10**	**60.10**	**69.50**	66.66

Table 5: Physiological-reasoning performance before and after RLE-PostTrain on three held-out test subjects (%). PG-Acc and State Macro-F1 use all 5,040 matched and intervention inputs (1,260 per condition). Changed-attribute adoption and preserved-attribute fidelity use the applicable intervention pairs; matched emotion accuracy uses the 1,260 unmodified held-out trials. GRPO w/o CF terms removes the changed-attribute and preserved-attribute reward terms.

RLE-PostTrain improves all four physiological measures for 4B, 8B, and 14B (Table 5). For 14B, PG-Acc rises from 63.84% to 74.15%, State Macro-F1 from 46.21% to 58.10%, changed-state adoption from 51.97% to 60.10%, and preserved-state fidelity from 59.12% to 69.50%. GRPO without CF terms yields smaller gains; its 8B preserved-state fidelity even changes from 55.82% to 55.71%. Matched emotion accuracy remains close to Stage 1 across sizes: 57.95% to 58.11% (4B), 62.50% to 62.66% (8B), and 66.52% to 66.66% (14B). These are independently generated Stage 3 emotion predictions, not frozen Stage 1 outputs.

For free-form text, three blinded reviewers assess 100 paired inputs (25 base trials under four conditions), viewing Stage 1 and Stage 3 explanations in random order alongside the seven current reference states. They judge consistency, unsupported content, and clarity; this assessment is separate from deterministic structured-state scores.

RLE-MLLM 14B	Evidence Consistency	Unsupported Content	Explanation Clarity
Stage 1 SFT	58.33	38.69	76.79
RLE-PostTrain	69.06	28.57	80.00

Table 6: Blinded human assessment of free-form explanations from RLE-MLLM 14B (%). Columns report the proportion of explanations judged consistent with the current EEG/EOG reference, the proportion containing unsupported physiological content, and the proportion judged clear, respectively.

Human evaluation. For 14B, blinded evidence consistency rises from 58.33% to 69.06%, unsupported content falls from 38.69% to 28.57%, and clarity rises from 76.79% to 80.00% (Table 6). These judgments cover prose beyond the seven deterministic states.

5 Conclusion

RLE-Reasoner connects EEG/EOG emotion prediction to independently extracted signal attributes through supervised generation, controlled intervention diagnosis, and physiology-constrained GRPO. For 14B, matched emotion accuracy is 66.52% before and 66.66% after repair; overall grounding improves from 63.84% to 74.15%, with higher blinded consistency against the seven supplied reference states in free-form explanations. The private cohort is small, and seven reference-relative attributes cannot fully establish the physiological mechanisms of emotion. Structured-state scoring and human review support narrower grounding claims; broader cohorts and patterns remain for future work.

References

Zebang Cheng, Zhi-Qi Cheng, Jun-Yan He, Jingdong Sun, Kai Wang, Yuxiang Lin, Zheng Lian, Xiaojiang Peng, and Alexander Hauptmann. Emotion-LLaMA: Multimodal emotion recognition and reasoning with instruction tuning. *arXiv preprint arXiv:2406.11161*, 2024. URL https://arxiv.org/abs/2406.11161.

Yi Ding, Chengxuan Tong, Shuailei Zhang, Muyun Jiang, Yong Li, Kevin JunLiang Lim, and Cuntai Guan. EmT: A novel transformer for generalized cross-subject EEG emotion recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 36(6):10381–10393, 2025. doi: 10.1109/TNNLS.2025.3552603. URL https://doi.org/10.1109/TNNLS.2025.3552603.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. REVE: A foundation model for EEG: Adapting to any setup with large-scale pretraining on 25,000 subjects. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/20a917f77773ac0fa8bea2bdd6606b66-Abstract-Conference.html.

Wei-Bang Jiang, Yansen Wang, Bao-Liang Lu, and Dongsheng Li. NeuroLM: A universal multitask foundation model for bridging the gap between language and EEG signals. *arXiv preprint arXiv:2409.00101*, 2024. URL https://arxiv.org/abs/2409.00101.

Wei-Bang Jiang, Xuan-Hao Liu, Wei-Long Zheng, and Bao-Liang Lu. SEED-VII: A multimodal dataset of six basic emotions with continuous labels for emotion recognition. *IEEE Transactions on Affective Computing*, 16(2):969–985, 2025. doi: 10.1109/TAFFC.2024.3485057.

Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, and Brent J. Lance. EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018. doi: 10.1088/1741-2552/aace8c.

Zheng Lian, Haoyu Chen, Lan Chen, Haiyang Sun, Licai Sun, Yong Ren, Zebang Cheng, Bin Liu, Rui Liu, Xiaojiang Peng, Jiangyan Yi, and Jianhua Tao. AffectGPT: A new dataset, model, and benchmark for emotion understanding with multimodal large language models. *arXiv preprint arXiv:2501.16566*, 2025. URL https://arxiv.org/abs/2501.16566.

Fei Ma, Han Lin, Yifan Xie, Hongwei Ren, Xiaoyu Shen, Wenbo Ding, and Qi Tian. E^2-LLM: Bridging neural signals and interpretable affective analysis. *arXiv preprint arXiv:2601.07877*, 2026. URL https://arxiv.org/abs/2601.07877.

Hyeongseop Rha, Jeong Hun Yeo, Yeonju Kim, and Yong Man Ro. Emotion-coherent reasoning for multimodal LLMs via emotional rationale verifier. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 2029–2037, 2026. doi: 10.1609/aaai.v40i3.37184. URL https://ojs.aaai.org/index.php/AAAI/article/view/37184.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. URL https://arxiv.org/abs/2402.03300.

Jinfeng Wang, Yanhao Huang, Sifan Song, Boqian Wang, Jionglong Su, and Jiaman Ding. A novel fourier adjacency transformer for advanced EEG emotion recognition. *arXiv preprint arXiv:2503.13465*, 2025. URL https://arxiv.org/abs/2503.13465.

Bohao Xing, Xin Liu, Guoying Zhao, Chengyu Liu, Xiaolan Fu, and Heikki Kälviäinen. Emotion-hallucer: Evaluating emotion hallucinations in multimodal large language models. *arXiv preprint arXiv:2505.11405*, 2025. URL https://arxiv.org/abs/2505.11405.

Zheng Zhou, Isabella McEvoy, and Camilo E. Valderrama. Local-global feature fusion for subject-independent EEG emotion recognition. *arXiv preprint arXiv:2601.08094*, 2026. URL https://arxiv.org/abs/2601.08094.