WORLD LAUREATES FORUM

世界顶尖科学家论坛

国际工程智能大会

2026 International Conference on AI for Engineering

World Laureates Foundation

上海世界顶尖科学家发展基金会

2026

工程智能研究院

INSTITUTE OF AI FOR ENGINEERING

Omni-Perception Policy Optimization for Multimodal Emotion Reasoning

Zhiyuan Han[1,2,4], Beier Zhu[1], Wenwen Tong[2], Pengyang Shao[3], Peipei Song[1], Xinyi Wang[1], Jiangnan Chen[2], Lewei Lu[2], Xun Yang[1]

[1] University of Science and Technology of China [2] SenseTime Research [3] National University of Singapore
[4] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center

01 ABSTRACT

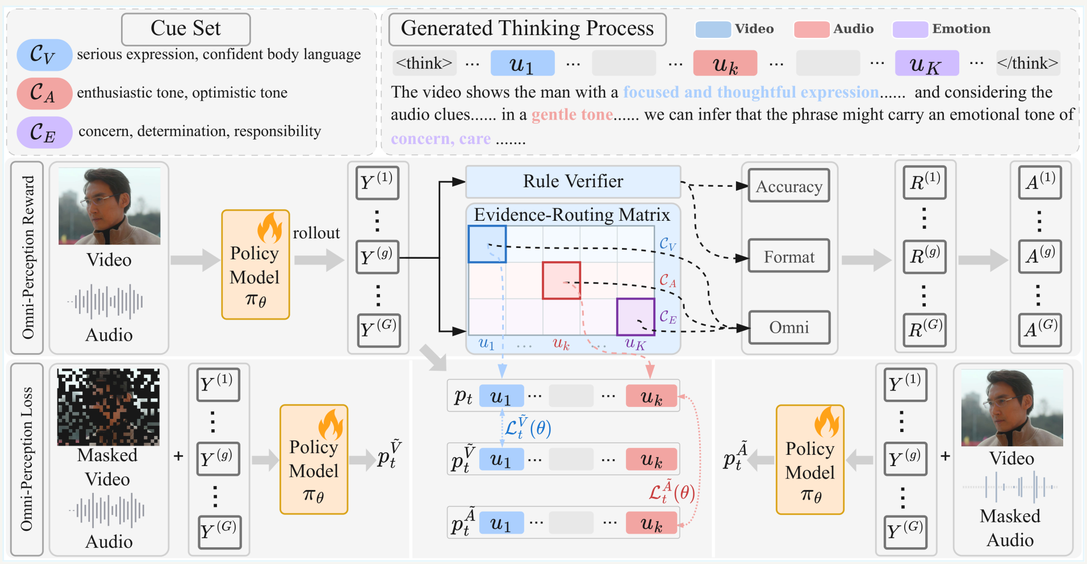
Omni-MLLMs often miss visual cues and invent evidence from absent modalities. OPPO improves cue utilization and reasoning faithfulness.

MEP-Bench: utilization and faithfulness

300 cue-rich samples measure recall. Another 500 masking probes per modality test whether reasoning respects its evidence source.

02 METHOD

OPPO augments GRPO with complementary reward and loss terms, linked by an evidence-routing matrix.



Cue matching guides both reward and modality-targeted loss.

Omni-Perception Reward

Match cues to their best reasoning clauses. Sum the average normalized hit scores of the three cue sets.

Omni-Perception Loss

Mask video or audio. Maximize KL divergence only at cue tokens associated with that modality.

Qwen2.5-Omni. SFT: lr 2e-5, 2 epochs. RL: lr 2e-6, 3,262 steps, G = 4.
α = 0.5, β = 8e-3, ρ = 0.7, δ = 0.5.

03 RESULTS

81.05
MER-UniBench mean

70.44
MEP-Bench cue recall

68.1
MME-Emotion reasoning

MER-UniBench: +5.45 mean over the reproduction

MER-UniBench	AffectGPT-R1 (reproduced)	OPPO
MOSI	80.29	**86.50**
MOSEI	80.64	**84.63**
SIMS	85.70	**86.22**
SIMSv2	83.75	**88.26**
MER23	81.88	**87.73**
MER24	80.89	**90.34**
MELD	57.53	**64.06**
IEMOCAP	65.71	**73.60**
OV-MERD+	64.08	**67.16**
Mean	**75.60**	**81.05**

MME-Emotion

Model	Recog.	Reason.	CoT
Baseline	27.9	62.7	45.3
OPPO	**31.0**	**68.1**	**49.5**

MEP-Bench

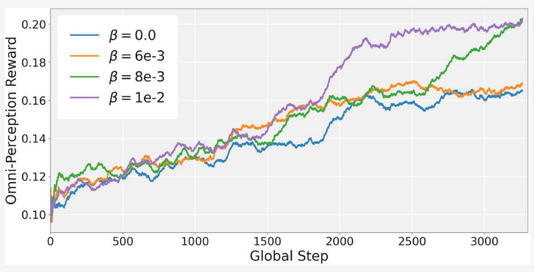
Model	Recall	Acc@V	Acc@A
Baseline	58.00	59.20	33.00
OPPO	**70.44**	**76.40**	**50.60**

Ablation: targeted loss strengthens the reward

Variant	Mean	Recall	Acc@V	Acc@A
Baseline	77.87	58.00	59.20	33.00
+ Reward	79.45	62.55	64.80	39.40
+ PAPO	79.34	58.29	62.00	34.60
+ Loss	**81.05**	**70.44**	**76.40**	**50.60**

Ablation uses its own baseline. Acc@V/A test video/audio faithfulness.
MME-Emotion comparisons use a common evaluator.

04 DISCUSSION



Reward trajectories across loss weights.

Small β leads to early reward saturation. At β = 8e-3, reward keeps rising. The loss opens a faithful information channel so reward optimization can reach modality-sensitive regions.

05 CONCLUSIONS

MEP-Bench diagnoses missed cues and cross-modal hallucination. OPPO pairs cue-level reward with a targeted masking loss to ground reasoning in visual and audio evidence. It improves cue coverage, faithfulness and emotion reasoning. Reliable perception is essential for robust MER.

CORRESPONDENCE

Xun Yang: xyang21@ustc.edu.cn
Beier Zhu: beier.zhu@ustc.edu.cn

SOURCE PAPER

Han et al., Omni-Perception Policy Optimization.
Figures and results from the supplied manuscript.