MER 2026: Interlocutor Emotion, Fine-Grained Emotion, Emotion Preference, and Emotion Recognition from Physiological Signals

1 Introduction

With the recent advancements in embodied AI, enabling robots to better understand human emotions and enhance their emotional intelligence has emerged as a significant research focus [1, 2]. This development can facilitate their seamless integration into human society. Emotion is an internal human state expressed through various modalities, including audio, video, text, and physiological signals [3, 4]. This has spurred the growth of Multimodal Emotion Recognition (MER), which seeks to integrate diverse cues to achieve a more comprehensive understanding of human emotions [5].

This year, we plan to launch MER 2026, the fourth edition in the MER series of challenges, designed to align with current research trends. Let us briefly revisit the evolution of the MER series. MER2023@ACM Multimedia [6] was our first organized challenge, aiming to advance the recognition of basic emotions through semi-supervised and multi-label learning techniques. MER2024@IJCAI [7] shifted the focus from basic emotions to fine-grained emotions and introduced a new track dedicated to *open-vocabulary emotions* [8]. Building on this progression, MER2025@ACM Multimedia [9] added another novel track, *descriptive emotions*, which moves beyond simple emotion label prediction by emphasizing evidence-based emotion understanding [10].

MER 2026 continues to align with recent research trends: moving from coarse-grained emotions to fine-grained emotions [8], and from categorical emotions to descriptive emotions [11]. Accordingly, MER 2026 will contain the following four tracks:

- **Interlocutor Emotion (MER-Cross).** This track shifts the focus from individual to dyadic interaction scenarios, better aligning with real-world embodied AI settings.
- **Fine-grained Emotion (MER-FG).** This track expands the scope from basic emotions to more nuanced categories, enabling a better understanding of human emotions.
- **Emotion Preference (MER-Prefer).** This task provides two emotion descriptions for each video and requires participants to predict which description is preferred by humans.
- **Physiological Signal-Based Emotion (MER-PS).** Unlike previous tracks that rely on human behavioral signals (such as audio, video, and text) to recognize emotions, this track shifts the focus to physiological signals.

Over the past three years, the MER series of challenges has grown into one of the largest emotion recognition challenges in the research community. It provides a common platform for researchers worldwide to explore recent trends and emerging tasks in the field. From MER2023 to MER2025, participant numbers steadily increased, and our dataset from last year was downloaded over 18,000 times on Hugging Face. This year, we will advertise more widely and aim to attract up to 200 teams. **We believe that the MER series of challenges will serve as a flagship challenge at ACM Multimedia.**

2 MER-Cross: Interlocutor Emotion

MER-Cross is a newly introduced track that shifts the focus from single-person to dyadic interaction scenarios, better aligning with real-world embodied AI settings. Specifically, in single-person scenarios, the character is denoted as s_1, and their corresponding audio, visual, and textual information is represented as a_1, v_1, and t_1, respectively. The goal is to utilize the character's multimodal information to predict their emotional state e_1:

$$P(e_1|a_1, v_1, t_1). \tag{1}$$

In MER-Cross, we focus on dyadic interaction scenarios, where two characters are denoted as s_1 and s_2, respectively. In embodied AI scenarios, a robot can both speak and perceive the human's actions. To emulate such scenarios, we designate the human as s_1 and the robot as s_2. Accordingly, we leverage the robot's (s_2) audio and textual input, along with the visual observation of the human (s_1), to predict the emotional state that the human will exhibit in the next round. We formulate this task as:

$$P(e_1^{next}|a_2, t_2, v_1). \tag{2}$$

In this track, we focus on basic emotion recognition, where predictions are limited to six predefined categories: *neutral*, *anger*, *happiness*, *sadness*, *worry*, and *surprise*. To evaluate performance, we adopt the weighted F1-score as the evaluation metric, following established practices in previous work [5].

3 MER-FG: Fine-grained Emotion

MER-FG was first introduced at MER2024 [7], and this year marks its third edition. The primary motivation behind MER-FG is that human emotion encompasses a vast array of emotion words, extending far beyond basic emotions [12]. Confining such rich emotional expression within a fixed label set can result in inaccurate emotion representation. Therefore, MER-FG aims to go beyond basic emotion categories and enable fine-grained emotion recognition [8]. This relaxed label space presents new challenges and imposes new requirements on both solutions and evaluation metrics. On the solution side, there is a need to shift from discriminative methods to generative ones, leveraging the extensive vocabulary of large language models (LLMs) to broaden the scope of emotion recognition. For evaluation metrics, we need to use grouping information from emotion wheels to mitigate the effects of synonymy, thereby ensuring more reliable evaluation results. Following previous MER challenges, we adopt the average F1-score across different emotion wheels as our evaluation metric, defined as follows:

$$\text{EW-based Metric} = \frac{1}{K}\sum_{k=1}^{K} \mathrm{F}_{\mathrm{S}}^{k}, \tag{3}$$

where K denotes the number of emotion wheels and $\mathrm{F}_{\mathrm{S}}^{k}$ represents the F1-score computed using the k-th emotion wheel. Further details can be found in our previous work [11, 13].

Table 1 summarizes the dataset statistics of MER-FG in previous MER challenges. This year, we will further increase the dataset size to foster progress in fine-grained emotion recognition.

Table 1: Dataset statistics of the FG track across previous MER challenges.

	Train&Val	Test
MER2024-FG [7]	332	200
MER2025-FG [9]	332	1,200

4 MER-Prefer: Emotion Preference

MER-Prefer is a newly introduced track. The concept of emotion preference was first proposed in EmoPrefer [14], designed to identify which of two emotion descriptions is preferred by human annotators. This task plays a crucial role in training reward models that can understand human emotions, laying the foundation for developing emotion-aware multimodal large language models (MLLMs). EmoPrefer comprises two preference datasets: EmoPrefer-Data and EmoPrefer-Data-V2. Table 2 summarizes the statistics of these two datasets. We will utilize them for training and additionally annotate new samples for performance evaluation.

Table 2: Training dataset statistics in MER-Prefer.

Dataset	# Samples
EmoPrefer-Data [14]	574
EmoPrefer-Data-V2 [14]	2,096

We adopt the official evaluation metric [14] for performance evaluation. Specifically, given N examples $\{(x^n, d_1^n, d_2^n)\}_{i=1}^N$, where x^n is a video and d_1^n and d_2^n are two corresponding descriptions. The ground-truth preference and the predicted preference are denoted as o_{12}^n and $\hat{o}_{12}^n$, respectively. Each label falls into one of three categories: (1) d_1^n *is preferred*, (2) d_2^n *is preferred*, or (3) *tie*. Following prior work [14], we employ the two-class (tie excluded) weighted F1-score for performance evaluation.

5 MER-PS: Physiological Signal-Based Emotion

MER-PS shifts multimodal emotion recognition from observable behaviors (e.g., audio, video, and text) to *physiological* evidence, with an emphasis on brain activity. To support this track, we adopt the MER-PS dataset, a synchronized EEG–fNIRS emotion dataset designed for studying affective dynamics under realistic stimuli. MER-PS contains recordings from 30 participants watching 15 emotion-eliciting video clips, together with baseline sessions. It provides 64-channel EEG signals (recorded at 1000 Hz) and 51-channel fNIRS signals (recorded at 47.62 Hz) covering emotion-relevant cortical regions, enabling joint observation of neuroelectrical activity and hemodynamic responses.

A key feature of MER-PS is its *real-time* dynamic emotion annotation. During video viewing, each subject continuously reports their affective state using joystick-based valence and arousal trajectories sampled at 1 Hz, represented as integers in $[1, 255]$ with a neutral origin centered at $(128, 128)$. In addition to these time-varying labels, MER-PS provides post-trial self-assessments using the SAM scale (valence, arousal, dominance, and familiarity, scored in $[1, 9]$), as well as PANAS questionnaires before and after the experiment. This combination of dynamic and static annotations supports both fine-grained affect decoding and reliability analysis across labeling protocols.

In MER-PS, we focus on predicting affective dynamics from physiological signals. Given synchronized brain signals, the goal is to estimate the valence–arousal trajectory over the stimulus:

$$\hat{\mathbf{y}}_{1:T} = f\left(\mathbf{x}_{1:L_e}^{\text{EEG}}, \mathbf{x}_{1:L_f}^{\text{fNIRS}}\right), \quad \mathbf{y}_t = [v_t, a_t], \tag{4}$$

where $\mathbf{x}^{\text{EEG}}$ and $\mathbf{x}^{\text{fNIRS}}$ denote the EEG and fNIRS time series, and $\mathbf{y}_t$ denotes the affect label at time t. Since the raw signals and labels have different temporal resolutions, participants may adopt appropriate alignment strategies, such as resampling, window-based aggregation, or sequence-to-sequence modeling, to bridge high-rate physiological streams and 1 Hz annotations.

Following the protocol suggested in the MER-PS study, we evaluate regression performance by reporting mean absolute error (MAE) for both valence and arousal after normalizing labels to $[0, 1]$:

$$\text{MAE} = \frac{1}{N}\sum_{n=1}^{N} |y_n - \hat{y}_n| \,. \tag{5}$$

To provide a single ranking score for the challenge, we use the average MAE over the two affect dimensions:

$$\text{Score} = \frac{\text{MAE}_v + \text{MAE}_a}{2}. \tag{6}$$

This design encourages models to move beyond coarse binning and directly capture continuous affect dynamics, which is critical for robust affective brain–computer interfaces and neuroadaptive interaction systems.

Ethical Statement. The local Institutional Review Board approved the ethical conduct of this study. All procedures involving human participants strictly adhered to established ethical standards, with particular emphasis on ensuring participant safety, privacy, and informed consent. Prior to participation, all subjects were fully informed about the study's purpose, procedures, and data usage policies.

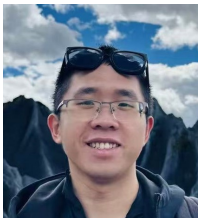
6 Schedule

Data, baseline paper & code available	20 April, 2026
Results submission start	20 June, 2026
Results submission deadline	1 July, 2026
Deadline for paper submission	15 July, 2026
Paper acceptance notification	30 July, 2026
Deadline for camera-ready papers	06 August, 2026

7 People

7.1 Organizers

Jianhua Tao (Senior Member, IEEE) received the Ph.D. degree from Tsinghua University, Beijing, China, in 2001, and the M.S. degree from Nanjing University, Nanjing, China, in 1996. He is currently a Professor with the Department of Automation, Tsinghua University, Beijing, China. He has authored or coauthored more than eighty papers in major journals and proceedings. His current research interests include speech recognition, speech synthesis and coding methods, human–computer interaction, multimedia information processing, and pattern recognition. He is the Chair or Program Committee Member for several major conferences, including ICPR, ACII, ICMI, ISCSLP, NCMMSC, etc. He is also the Steering Committee Member for the IEEE Transactions on Affective Computing, an Associate Editor for the Journal on Multimodal User Interface and the International Journal on Synthetic Emotions, and the Deputy Editor-in-Chief for the Chinese Journal of Phonetics. He was the recipient of several awards from important conferences, such as Eurospeech, NCMMSC, etc.



Zheng Lian (Senior Member, IEEE) is an Associate Professor at the Institute of Automation, Chinese Academy of Sciences (CASIA). He received his Ph.D degree from CASIA, China, in 2021. His research interests primarily center on human-centric AI and affective computing. He co-organized a series of challenges and workshops (MER@IJCAI, MRAC@ACM Multimedia, MEIJU@ICASSP), established benchmark (MERBench) and toolbox (MERTools), and proposed new tasks to enhance accuracy and reliability (EMER, OV-MER, AffectGPT, AffectGPT-R1, EmoPrefer). He (co-)authored more than 100 publications in journals, patents, and conference proceedings, including IEEE TPAMI, NeurIPS, ICML, ICLR, IEEE TNNLS, IEEE TASLP, and IEEE TAFFC, leading to >4,400 citations (h-index: 35). He also serves as Associate Editor of IEEE TAFFC and IEEE TASLP, Area Editor of Information Fusion, Area Chair of ACM Multimedia 2025 and ACL ARR 2025, and ACM Multimedia 2026 Dataset Co-Chair.

Björn W. Schuller (Fellow, IEEE) is a Full Professor of Artificial Intelligence and the Head of GLAM at Imperial College London/UK, Full Professor and Chair of Health Informatics at the TUM University Hospital in Munich/Gemrany, co-founding CEO and current CSO of audEERING – an Audio Intelligence company based near Munich and in Berlin/Germany, amongst other Professorships and Affiliations. He is a Fellow of the ACM and IEEE, and a Golden Core Awardee of the IEEE Computer Society, Fellow of the BCS, Fellow of the ELLIS, Fellow of the ISCA, Fellow and President-Emeritus of the AAAC, and Elected Full Member Sigma Xi. He (co-)authored 1,700+ publications (80,000+ citations, h-index=125), is Field Chief Editor of Frontiers in Digital Health, Editor in Chief of AI Open, and was Editor in Chief of the IEEE Transactions on Affective Computing, amongst manifold further commitments and service to the community. His 50+ awards include being honoured as one of 40 extraordinary scientists under the age of 40 by the WEF in 2015. He served as consultant of companies such as Barclays, GN, Huawei, Informetis, or Samsung.

Guoying Zhao (Fellow, IEEE) is currently an Academy Professor with Academy Finland and with the Center for Machine Vision and Signal Analysis (CMVS), University of Oulu, Finland, and visiting professor with Aalto University and Stanford University. She has authored or co-authored more than 300 papers in journals and conferences, and has served as a reviewer for many journals and conferences. Her papers have currently over 21500 citations in Google Scholar (h-index 69). She is the panel chair for FG 2023, publicity chair for SCIA 2023, and tutorial chair for ICPR 2024. She was general chair of Arctic AI days 2022, 2022 International Conference on Image Processing and Media Computing (ICIPMC 2022) and ICBEA (2019, 2020), co-program chair of ACM International Conference on Multimodal Interaction 2021 (ICMI), co-chair for Late Breaking Results of ICMI 2019, co-publicity chair for FG 2018, has served as area chairs for several conferences and is/was associate editor for Pattern Recognition, IEEE Transactions on Multimedia, IEEE Transactions on Circuits and Systems for Video Technology, and Image and Vision Computing Journals. She is IEEE Fellow, IAPR Fellow, and AAIA Fellow.

Erik Cambria (Fellow, IEEE) is the Founder of SenticNet, a Singapore-based company offering B2B sentiment analysis services, and an Associate Professor at Nanyang Technological University, where he also holds the appointment of Provost Chair in Computer Science and Engineering. Prior to joining NTU, he worked at Microsoft Research Asia (Beijing) and HP Labs India (Bangalore) and earned his PhD through a joint program between the University of Stirling and MIT Media Lab. He is recipient of several awards, e.g., IEEE Outstanding Career Award, was listed among the AI's 10 to Watch, and was featured in Forbes as one of the 5 People Building Our AI Future. He is an IEEE Fellow, Associate Editor of many top-tier AI journals, e.g., Information Fusion and IEEE Transactions on Affective Computing, and is involved in various international conferences as keynote speaker, program chair, and senior program committee member.

7.2 Challenge Chairs

Kele Xu received his doctoral degree in 2016 from Université Pierre et Marie CURIE (UPMC), Paris, France. He is an Associate Professor at the National University of Defence Technology (NUDT), China, and a Senior Member of the IEEE. He has authored or co-authored over 100 publications in peer-reviewed journals and conferences such as ICLR, CVPR, NeurIPS, ICML, AAAI, ASE, IJCAI, SIGIR, ACM MM, ICRA, and IEEE TKDE, TALSP, TOMM, TMI, TCSVT, TGRS, Information Fusion, Speech Communication, and Advanced Science, which have garnered more than 4,000 Google Scholar citations with an h-index of 36. He also served as an Associate Editor for IEEE Transactions on Circuits and Systems for Video Technology.

Xiaojiang Peng (Senior Member, IEEE) is a full professor at the School of Artificial Intelligence at Shenzhen Technology University, and serves as the dean of the Artificial Intelligence Department. He received his Ph.D. degree in Computer Science from Southwest Jiaotong University. He was an associate professor at the Chinese Academy of Sciences and a postdoctoral researcher with Idiap, Switzerland, and Inria THOTH, France. He has published more than 100 top journal/conference papers, and serves as Associate Editor for IEEE Transactions on Affective Computing and Empathic Computing, and also as the reviewer for the IJCV, IEEE TPAMI, IEEE TIP, CVPR, ICCV, ECCV, AAAI, IJCAI, FG, etc. He has been selected by Stanford University as one of the world's top 2% scientists in 2022, 2023, 2024, and 2025. His research interests include computer vision, affective computing, embodied intelligence, and generative models.

Fei Ma is currently a researcher at Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ). Before that, he received the B.S. degree in Communication Engineering from University of Electronic Science and Technology of China in 2017 and the Ph.D. degree in Information and Communication Engineering from Tsinghua University in 2022. So far, he has published over 40 papers in top-tier journals such as IEEE TPAMI, IEEE TMC, IEEE TAFFC, and IEEE TIE, as well as in prestigious conferences including NeurIPS, ACL, AAAI, ACM MM, and IJCAI. His research interests include generative AI and multimodal learning.

Laizhong Cui received the Ph.D. degree in computer science and technology from Tsinghua University, Beijing, China, in 2012. He is currently a professor with the College of Computer Science and Software Engineering, Shenzhen University, China. His research interests include future Internet architecture and protocols, edge computing, and federated learning. He has led more than ten research projects, including the National Key Research and Development Plan of China, the National Natural Science Foundation of China, the Guangdong Natural Science Foundation of China, and the Shenzhen Basic Research Plan. He has published more than 100 papers in leading journals and conferences. He serves as an associate editor or editorial board member for several international journals, including IEEE Internet of Things Journal, IEEE Transactions on Cloud Computing, IEEE Transactions on Network and Service Management, and the International Journal of Machine Learning and Cybernetics. He is a distinguished member of the CCF.

Zebang Cheng is currently a Ph.D. candidate at Shenzhen University, China. His research focuses on affective computing, with particular emphasis on multimodal large models for emotion recognition and reasoning. He has participated in multiple research projects and published several papers in top international venues such as NeurIPS and ACM Multimedia. He also serves as a reviewer for leading journals, including IEEE Transactions on Affective Computing, IEEE Transactions on Multimedia, and Pattern Recognition. His work aims to deepen the understanding of the cognitive mechanisms underlying human emotions and to advance the development of reliable and empathetic human–AI interaction.

Ziyu Jia is an Assistant Professor at the Institute of Automation, Chinese Academy of Sciences. His research focuses on time-series analysis methods and their applications in health and medicine, including multimodal affective computing, sleep stage classification, and brain-computer interfaces. He has published over 50 peer-reviewed papers in venues such as IEEE Transactions on Affective Computing, IEEE Transactions on Multimedia, IEEE Transactions on Neural Systems and Rehabilitation Engineering, KDD, and ICLR. Dr. Jia currently serves as an Associate Editor or Editorial Board Member for prestigious journals, including IEEE Transactions on Affective Computing and Information Fusion, and he is an Area Chair for major AI and machine learning conferences such as IJCAI and IJCNN. In addition to his academic contributions, Dr. Jia has extensive industry experience, having successfully led multiple R&D projects and secured several patents. He has received numerous honors, including the MSRA StarTrack Award and the CIE Young Talent Award.

7.3 Program Committee Members

Jing Han (University of Cambridge), Jianfei Yu (Nanjing University of Science and Technology), Zixing Zhang (Hunan University), Xiaobai Li (Zhejiang University), Jingming Zhao (Renmin University of China), Licai Sun (Oulu University), Ya Li (Beijing University of Posts and Telecommunications), Mingyue Niu (Tianjin Normal University), Mengyue Wu (Shanghai Jiao Tong University), Yong Li (Southeast University), Xie Chen (Shanghai Jiao Tong University), Liumeng Xue (Hong Kong University of Science and Technology).

8 Commitment

If our proposal is accepted, we commit to publishing and maintaining a website dedicated to the Grand Challenge, which will contain all relevant information, datasets, and tasks for at least the next three years. We also pledge to collaborate with the ACM Multimedia Conference organizers to publicize the Grand Challenge tasks and encourage researcher participation. For any questions, please contact Zheng Lian (zheng.lian.zeroqiaoba@gmail.com).

References

[1] Michael A Arbib and Jean-Marc Fellous. Emotions: from brain to robot. *Trends in Cognitive Sciences*, 8(12):554–561, 2004.

[2] Jiafei Duan, Samson Yu, Hui Li Tan, Hongyuan Zhu, and Cheston Tan. A survey of embodied ai: From simulators to research tasks. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(2):230–244, 2022.

[3] Robert Plutchik. A general psychoevolutionary theory of emotion. *Emotion: Theory, research, and experience*, 1, 1980.

[4] Rosalind W Picard. *Affective computing*. MIT press, 2000.

[5] Zheng Lian, Licai Sun, Yong Ren, Hao Gu, Haiyang Sun, Lan Chen, Bin Liu, and Jianhua Tao. Merbench: A unified evaluation benchmark for multimodal emotion recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

[6] Zheng Lian, Haiyang Sun, Licai Sun, Kang Chen, Mngyu 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 *Proceedings of the 31st ACM International Conference on Multimedia*, pages 9610–9614, 2023.

[7] 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*, pages 41–48, 2024.

[8] Zheng Lian, Haiyang Sun, Licai Sun, Haoyu Chen, Lan Chen, Hao Gu, Zhuofan Wen, Shun Chen, Zhang Siyuan, Hailiang Yao, et al. Ov-mer: Towards open-vocabulary multimodal emotion recognition. In *Forty-second International Conference on Machine Learning*, 2025.

[9] 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 *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 13837–13842, 2025.

[10] Zheng Lian, Licai Sun, Mingyu Xu, Haiyang Sun, Ke Xu, Zhuofan Wen, Shun Chen, Bin Liu, and Jianhua Tao. Explainable multimodal emotion reasoning. *arXiv preprint arXiv:2306.15401*, 2023.

[11] 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 *Forty-second International Conference on Machine Learning*, 2025.

[12] Paul Ekman. Are there basic emotions? 1992.

[13] Zheng Lian, Fan Zhang, Yazhou Zhang, Jianhua Tao, Rui Liu, Haoyu Chen, and Xiaobai Li. Affectgpt-r1: Leveraging reinforcement learning for open-vocabulary multimodal emotion recognition. *arXiv preprint arXiv:2508.01318*, 2026.

[14] Zheng Lian, Licai Sun, Lan Chen, Haoyu Chen, Zebang Cheng, Fan Zhang, Ziyu Jia, Ziyang Ma, Fei Ma, Xiaojiang Peng, et al. Emoprefer: Can large language models understand human emotion preferences? In *Proceedings of the International Conference on Learning Representations, ICLR*, 2026.