REACT 2026 Challenge: The Fourth Personalised Multiple Appropriate Facial Reaction Generation in Dyadic Interactions

Dr Siyang Song
University of Exeter
United Kingdom
s.song@exeter.ac.uk

Dr Micol Spitale
Politecnico di Milano
Milan, Italy
micol.spitale@polimi.it

Zijian Wu
Nanjing University of Science and Technology
China
wuzijian@njust.edu.cn

Xiangyu Kong
University of Exeter
United Kingdom
xk219@exeter.ac.uk

Cheng Luo
King Abdullah University of Science and Technology
Saudi Arabia
cheng.luo@kaust.edu.sa

Dr Cristina Palmero
King's College London
London, United Kingdom
cristina.palmero@kcl.ac.uk

Dr German Barquero
Universitat de Barcelona
Barcelona, Spain
germanbarquero@ub.edu

Prof Sergio Escalera
Universitat de Barcelona
Barcelona, Spain
sergio@maia.ub.es

Prof Michel Valstar
University of Nottingham
Nottingham, United Kingdom
michel.valstar@nottingham.ac.uk

Prof Mohamed Daoudi
IMT Nord Europe
Villeneuve d'Ascq, France
mohamed.daoudi@imt-nord-europe.fr

Dr Fabien Ringeval
Université Grenoble Alpes
Grenoble, France
fabien.ringeval@imag.fr

Prof Andrew Howes
University of Exeter
Exeter, United Kingdom
andrew.howes@exeter.ac.uk

Prof Elisabeth André
University of Augsburg
Augsburg, Germany
elisabeth.andre@uni-a.de

Prof Hatice Gunes
University of Cambridge
Cambridge, United Kingdom
hatice.gunes@cl.cam.ac.uk

ABSTRACT

According to the Stimulus Organism Response (SOR) theory, for a given external stimulus, individuals may react differently according to their internal state and external contextual factors in a specific period in time. Analogously, in dyadic interactions, a broad spectrum of human facial reactions might be *appropriate* for responding to a specific human speaker behaviour. Following the successful organisation of the REACT 2023, REACT 2024 and REACT 2025 challenge series, a body of generative deep learning (DL) models have been investigated for the problem of multiple appropriate facial reaction generation (MAFRG). While REACT 2023 and 2024 challenges were built on human-human dyadic interaction datasets collected for other purposes, the REACT 2025 challenge provided the first natural and large-scale audio-visual Multiple Appropriate Facial Reaction Generation (MAFRG) dataset (called MARS) recording 137 human-human dyadic interactions containing a total of 3,105 interaction sessions covering five different topics. This year, we are proposing the REACT 2026 challenge encouraging the development and benchmarking of Machine Learning (ML) models that can be used to generate multiple *appropriate*, **diverse**, **realistic** and **synchronised** human-style facial reactions expressed by human listeners in response to each input speaker behaviour expressed by the corresponding speaker. As a key of the challenge, we will continuously provide challenge participants with MARS dataset but additionally providing **individual-level Big-Five personality labels and EEG recordings**. This introduces a new one-to-many personalised reaction generation setting combining behavioural, affective and neurophysiological signals, which remains largely unexplored in current dyadic interaction modelling. We will then invite the challenge participating groups to submit their developed / trained ML models for evaluation, which will be benchmarked in terms of the appropriateness, diversity, realism and synchronisation of their generated facial reactions.

ACM Reference format:
Dr Siyang Song, Dr Micol Spitale, Zijian Wu, Xiangyu Kong, Cheng Luo, Dr Cristina Palmero, Dr German Barquero, Prof Sergio Escalera, Prof Michel Valstar, Prof Mohamed Daoudi, Dr Fabien Ringeval, Prof Andrew Howes, Prof Elisabeth André, and Prof Hatice Gunes. 2026. REACT 2026 Challenge: The Fourth Personalised Multiple Appropriate Facial Reaction Generation in Dyadic Interactions. In *Proceedings of ACM Conference, , (Conference'25),* 5 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 STATE-OF-THE-ART AND CONTRIBUTIONS

Recent years have seen an increasing number of studies targeting automatic human-human dyadic interaction analysis, due to the wide application scenarios and the advancements in pattern recognition, cognitive science, and neural network techniques [7]. Personalised human-style facial behaviours play a key role for people to convey their unique characteristics, attitudes and emotions in human-human interactions. More importantly, the development of personalised facial reaction generation (FRG) models enable interactive systems and humanoid virtual agents to produce human-style facial reactions that better match individual reaction styles, improving naturalness, engagement, and user trust compared to generic facial reactions. Previous FRG solutions [3, 5, 8, 10, 11] aim to generate a specific behavioural reaction that resembles the ground-truth (real) response or reaction for a given input. As a result, most of them proposed deterministic solutions that aim to reproduce the ground-truth reaction – specifically hand gestures [11] and facial reactions [3, 8] - without considering the non-deterministic aspects that different but appropriate human facial reactions could be triggered from the same perceived stimuli.

To bridge this gap, our challenge team recently introduced a new theoretical framework on why and how multiple appropriate facial reactions can be generated for responding to a given speaker behaviour [9], and also successfully organised three multiple appropriate facial reaction generation (MAFRG) challenges respectively at the ACM Multimedia 2023 conference (REACT 2023 [22]), IEEE FG 2024 conference (REACT 2024 [23]), and ACM Multimedia 2025 conference (REACT 2025 [21]). Following the successful organisation of these challenges, an increasing number of approaches [1, 12–20, 24–26, 28] have been proposed for the MAFRG task, where most of them attempted to addressing the 'one-to-many mapping' problem during training (i.e., one input corresponds to multiple different but appropriate facial reactions (AFRs)).

While the REACT 2023 and REACT 2024 challenges only including modified/segmented audio-visual clips originally collected by the NoXI [2], UDIVA [6] and RECOLA [4] datasets, all of which were recorded for the purposes other than the MAFRG task (e.g., personality computing and emotion recognition) and thus limiting to develop more advanced MAFRG systems, the recent REACT 2025 Challenge introduced and shared the novel and well-annotated Multimodal Challenge Dataset (called Multi-modal **M**ultiple **A**ppropriate **R**eaction in **S**ocial Dyads (MARS) Dataset), together with the recorded audio, visual, and objectively-annotated ground-truth (GT) real appropriate facial reactions of each speaker behaviour. However, all these challenges only encourage participants to develop generic MAFRG models that produce general AFRs without considering personalised aspect, i.e., human listeners of different internal disposition (e.g., personality) can express varied AFRs in response to the same speaker behaviour [8].

Despite the rapid progress achieved by recent MAFRG approaches and the previous editions of the REACT challenge series, several key research gaps remain. First, most existing works focus on generic facial reaction modelling and overlook the role of individual differences, while human facial reactions are inherently shaped by internal traits such as personality and cognitive states. Second, current benchmarks rarely integrate neurophysiological signals, limiting the exploration of internal–external alignment between latent user states and observable behavioural responses. Finally, the evaluation of one-to-many appropriate facial reaction generation remains underdeveloped, as traditional deterministic metrics struggle to capture diversity, appropriateness, and temporal synchronisation simultaneously. Addressing these limitations motivates the design of the REACT 2026 challenge, which introduces personalised appropriate facial reaction generation tasks supported by multi-modal behaviours, personality, and EEG signals, together with extended evaluation protocols. This new challenge includes generating not only multiple generic AFRs but also personalised AFRs in response to each speaker audio-visual behaviour expressed at the same time as previous MAFRG challenges required, (i.e., multiple face videos and the corresponding multi-channel facial primitive time-series consisting of 25 facial attributes – e.g., 15 action units (AUs), 8 facial expressions, valence and arousal) as defined in [9]. In summary, the main contributions and novelties of this challenge are listed as follows:

- Introducing and sharing the extended REACT 2026 Multimodal Challenge Dataset (called Multi-modal **M**ultiple **A**ppropriate **R**eaction in **S**ocial Dyads (MARS) Dataset), together with the recorded multi-modal audio, visual, objectively-annotated ground-truth appropriate facial reactions of each speaker behaviour, as well as newly introduced Big-Five personality labels and EEG signals.
- Creating and presenting state-of-the-art baseline personalised MAFRG models for generating personalised multiple AFRs in response to each multi-modal input (i.e., a speaker audio-visual behaviour) in a dyadic interaction setting, which is different from the previous REACT 2023, 2024, and 2025 challenges that only require to develop generic MAFRG models. Also, this challenge requires the developed MAFRG models to be able to process variable-length multi-modal clips. This aims to better simulate real-world human-computer interactions where different users may express the same behaviour with different ways and rates, which is different from REACT 2023 and REACT 2024 challenges that only require to predict equal-length appropriate facial reactions from equal-length speaker behaviours.
- Defining and introducing new objective measures in addition to previous MAFRG metrics to more comprehensively evaluate the developed personalised MAFRG models in terms of their generated personalised AFRs.

2 CHALLENGE TASKS AND BASELINE MODELS

Formally, given a short-term behaviour $b(S_n)^{t_1,t_2}$ expressed by a speaker S_n at the time period $[t_1, t_2]$, the REACT 2026 challenge not only follows the similar purpose and form as the REACT 2025 challenges, focusing on two generic Multiple Appropriate Facial Reaction Generation (MAFRG) tasks (i.e., offline generic MAFRG and online generic MAFRG tasks) but also newly introduces two personalised MAFRG tasks (i.e., offline personalised MAFRG and online personalised MAFRG tasks).

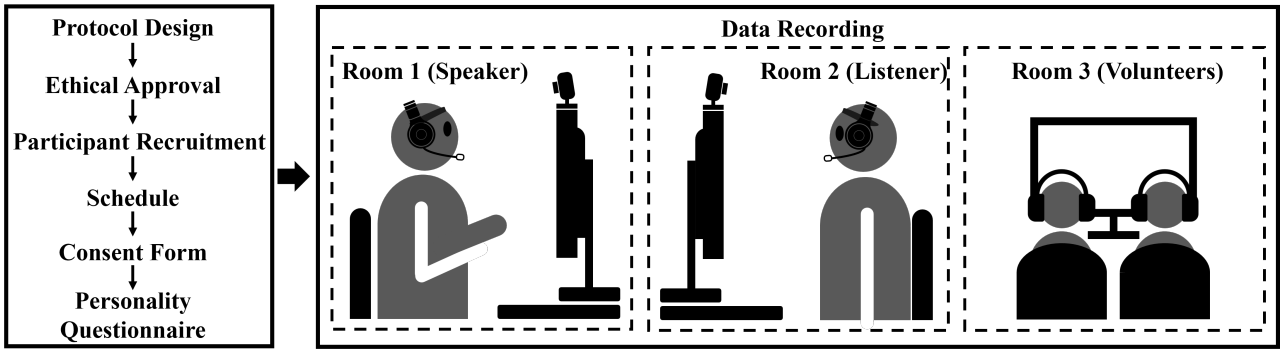

Figure 1: Illustration of the data collection process. The left section outlines the preparatory steps, including protocol design, ethical approval, scheduling, obtaining participant consent, and completing a personality questionnaire. The right section illustrates the physical data collection setup for dyadic interaction data recording.

2.1 Generic Appropriate Facial Reaction Generation

Task 1: Offline Generic Appropriate Facial Reaction Generation: This task aims to develop a machine learning model $\mathcal{H}$ that takes the entire speaker behaviour sequence $b(S_n)^{t_1,t_2}$ as the input, and generates multiple (M) appropriate and realistic / naturalistic spatio-temporal facial reactions $p_f(L|b(S_n)^{t_1,t_2})_1, \cdots, p_f(L|b(S_n)^{t_1,t_2})_M$, where $p_f(L|b(S_n)^{t_1,t_2})_m$ is a multi-channel time-series consisting of AUs, facial expressions, valence and arousal state representing the m_{th} predicted facial reaction. As a result, M facial reactions are required to be generated for the task given each input speaker behaviour.

Task 2: Online Generic Appropriate Facial Reaction Generation: This task aims to develop a machine learning model $\mathcal{H}$ that estimates each frame (i.e., $\gamma_{\text{th}} \in [t_1, t_2]$ frame) of the listener's facial reaction by only considering the γ_{th} frame and its previous frames expressed by the corresponding speaker (i.e., $t_{1\text{th}}$ to γ_{th} frames in $b(S_n)^t$), rather than taking all $t_{1\text{th}}$ to $t_{2\text{th}}$ frames into consideration. The model is expected to gradually generate all facial reaction frames to form multiple (M) appropriate and realistic / naturalistic spatio-temporal facial reactions $p_f(L|b(S_n)^{t_1,t_2})_1, \cdots, p_f(L|b(S_n)^{t_1,t_2})_M$, where $p_f(L|b(S_n)^{t_1,t_2})_m$ is a multi-channel time-series consisting of AUs, facial expressions, valence and arousal state representing the m_{th} predicted facial reaction. As a result, M facial reactions are required to be generated for the task given each input speaker behaviour.

2.2 Personalised Appropriate Facial Reaction Generation

Task 1: Offline Personalised Appropriate Facial Reaction Generation: This task aims to develop a machine learning model $\mathcal{H}$ that takes the entire speaker behaviour sequence $b(S_n)^{t_1,t_2}$ as the input, and generates multiple (M) personalised appropriate and realistic / naturalistic spatio-temporal facial reactions $p_f(l|b(S_n)^{t_1,t_2})_1, \cdots, p_f(l|b(S_n)^{t_1,t_2})_M$, where $p_f(l|b(S_n)^{t_1,t_2})_m$ is a multi-channel time-series consisting of AUs, facial expressions, valence and arousal state representing the m_{th} predicted personalised facial reaction, **which is expected to be similar to real AFRs expressed by the target listener l in response to human behaviours that are similar to the given speaker behaviour B^s.** As a result, M personalised facial reactions are required to be generated for the task given each input speaker behaviour.

Task 2: Online Personalised Appropriate Facial Reaction Generation: This task aims to develop a machine learning model $\mathcal{H}$ that estimates each frame (i.e., $\gamma_{\text{th}} \in [t_1, t_2]$ frame) of the listener's facial reaction by only considering the γ_{th} frame and its previous frames expressed by the corresponding speaker (i.e., $t_{1\text{th}}$ to γ_{th} frames in $b(S_n)^t$), rather than taking all $t_{1\text{th}}$ to $t_{2\text{th}}$ frames into consideration. The model is expected to gradually generate all facial reaction frames to form multiple (M) appropriate and realistic / naturalistic spatio-temporal facial reactions $p_f(L|b(S_n)^{t_1,t_2})_1, \cdots, p_f(L|b(S_n)^{t_1,t_2})_M$, where $p_f(L|b(S_n)^{t_1,t_2})_m$ is a multi-channel time-series consisting of AUs, facial expressions, valence and arousal state representing the m_{th} predicted personalised facial reaction, **which is expected to be similar to real AFRs expressed by the target listener l in response to human behaviours that are similar to the given speaker behaviour B^s.** As a result, M personalised facial reactions are required to be generated for the task given each input speaker behaviour.

Baseline models: To enable comparison for the participating teams, we will provide two deep learning baselines that are specifically designed for the online generic and personalised MAFRG task, i.e., the open-source ReactFace (i.e., Trans-VAE based model) [14] and PerFRDiff (i.e., Diffusion-based model) [28] frameworks, as well as two specifically designed offline generic and personalised MAFRG baselines, including the open-source REGNN (Reversible graph neural network-based model) [25] and PerReactor (GAN-based model) [27] frameworks.

3 CHALLENGE DATASET AND LABELS

Dataset. The REACT 2026 challenge builds upon the Multi-modal Multiple Appropriate Reaction in Social Dyads (MARS) dataset, which was originally introduced in REACT 2025 and is specifically designed for Multiple Appropriate Facial Reaction Generation (MAFRG) tasks. Compared to previous editions, REACT 2026 extends the dataset with individual-level personality annotations and

neurophysiological signals (EEG), enabling the study of personalised AFR generation driven by both observable behaviour and latent internal states. The dataset comprises 137 human–human dyadic interaction recordings involving 23 speakers and 137 listeners, resulting in 270 multi-modal recordings and 3,105 interaction sessions. Each recording contains synchronised audio, facial video, and EEG signals captured during remote conversations conducted via Microsoft Teams. Sessions cover five structured conversational topics, including cultural discussions, movie sharing, policy debates, quizzes and games, and scenario-based interviews, ensuring a controlled semantic context while maintaining naturalistic interaction dynamics. During data collection, speakers and listeners were located in separate rooms and interacted remotely to reduce physical co-presence bias while preserving conversational realism. Two volunteers supervised the recording process to ensure adherence to the experimental protocol and to handle potential interruptions. The recordings range from approximately 20 to 35 minutes, capturing diverse behavioural patterns across participants and interaction contexts. The introduction of EEG signals and personality labels introduces several new research challenges compared to previous REACT datasets. First, the dataset exhibits substantial inter-subject variability in AFR style, reflecting differences in personality traits and internal states. Second, multi-modal synchronisation between audio-visual behaviours and EEG signals raises challenges related to temporal alignment and noise robustness. Third, the presence of multiple AFRs for the same speaker behaviour introduces inherent ambiguity, requiring models to learn probabilistic rather than deterministic mappings.

Ground-truth labels. Ground-truth AFRs are defined at the session level. For a given speaker behaviour expressed within a specific conversational context, all real AFRs expressed by human listeners recorded under the same session are considered valid appropriate responses. This formulation reflects the one-to-many nature of human reactions and encourages models to capture behavioural diversity rather than reproducing a single deterministic target. Each participant also completed a self-reported Big-Five personality questionnaire, which provides trait-level annotations enabling personalised reaction generation. The addition of personality information allows researchers to investigate how internal dispositions modulate behavioural responses, bridging observable multi-modal signals with latent user characteristics.

Ethical consideration. FR: Ethical approval for the MARS dataset was obtained from the University of Leicester. All participants provided informed consent prior to data acquisition and were informed about the recording procedures and use of wearable devices, including MUSE-2 EEG sensors. Participants received financial compensation (£20 per hour) for their participation. During the challenge, access to the dataset will be granted under an End-User License Agreement (EULA) to ensure responsible use and data protection.

4 EVALUATION PROCESS

Challenge participants will be given access to the training and validation sets to develop their ML models, with a challenge guideline and a baseline paper released to provide more details. Then, they will be asked to submit the developed generic MAFRG and personalised MAFRG models and weights, while the organisers will evaluate their models on the test set (the test set will not be released to the challenge participants). Each participating group will be allowed to submit their models and results for the test set up to five times. The submitted models will be automatically evaluated and ranked using two sets of objective measures, i.e., the **appropriateness** and the **diversity** of the generated facial reactions, as defined and described in [9].

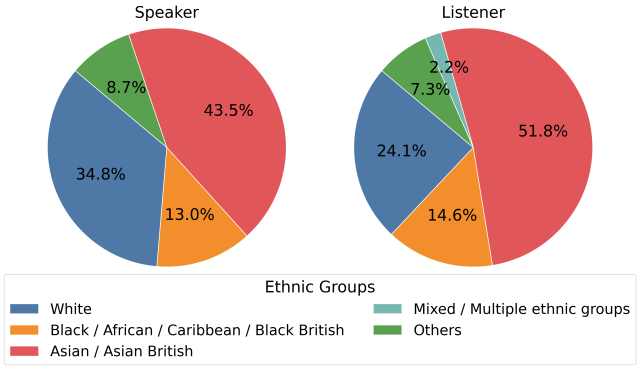

Figure 2: Statistics of participants' ethnic groups.

5 ADMINISTRATIVE DETAILS

A challenge website will be set up with a commitment to be maintained for at least the next three years (REACT 2026, REACT 2027, and REACT 2028 challenges). A GitHub repository [1] will be continuously established along with the official website (the previous REACT 2023, REACT 2024, and REACT 2025 challenge websites are [2], [3], and [4]) to guide and support the challenge participants. The challenge participants will be invited to submit challenge-style papers describing their ML solutions and results on the challenge dataset - these will be peer-reviewed and once accepted, will appear in the Challenge Proceedings. Dr Siyang Song (Email: s.song@exeter.ac.uk) and Dr Micol Spitale (Email: micol.spitale@polimi.it) will organise, publicise (e.g., in social media), review, and judge the REACT 2026 Challenge submissions in discussion with, and under the guidance of the rest of the challenge organisers. The Programme Committee (PC) will be composed of the organisers as well as other academics/researchers from various institutions that will be invited upon the acceptance of this challenge proposal. We aim to continuously organise this challenge series for at least the next 3 years.

6 ORGANISING TEAM

The challenge organising team consists of five top research groups (and a spin-out company) in the fields of Human Behaviour Analysis, Affective Computing and Social Signal Processing, from six countries, with > 100,000 citations to their works. Additionally, the Cambridge, Barcelona, Nottingham and Grenoble teams have an

[1] https://github.com/reactmultimodalchallenge/
[2] https://sites.google.com/cam.ac.uk/react2023/home
[3] https://sites.google.com/cam.ac.uk/react2024/home
[4] https://sites.google.com/view/react2025

Role	Gender			Age Group				Highest Degree*				Mother Language		Total
	Male	Female	Others	≤20	21-30	31-40	≥ 41	B	U	M	D	English	Others	
Speaker	10	13	0	2	18	2	1	2	4	14	2	11	12	23
Listener	71	65	1	18	81	26	12	26	29	58	24	47	90	137

Table 1: Demographics of participants. *: The highest levels of educational attainment are categorised as B: Below Undergraduate, U: Undergraduate Degree, M: Master's Degree, D: Doctoral Degree.

Topics	Time		
	Min	Max	Average
Cultural differences	195s	530s	339s
Movie scene sharing	143s	573s	321s
Policy changes	153s	726s	336s
Quizzes and games	191s	415s	311s
Scenario-based interviews	161s	518s	305s
Overall interactions	19m'50s	40m'49s	27m'16s

Table 2: Interaction statistics of different topics/sessions and overall interaction recordings computed on the entire MARS dataset. 's' denotes seconds, while 'm' denotes minutes.

unprecedented experience in organizing computational challenges since 2011. For example, the Barcelona team has organised 20 workshops and associated challenges, in conjunction with top conferences including ACM MM, CVPR, ICCV, ECCV, NeurIPS, ICMI, IJCNN, FG, and ICPR – their ChaLearn LAP Challenge series, with many high-impact papers, has contributed to significant advances in the field of visual human behaviour analysis and generation.

REFERENCES

[1] Quang Tien Dam, Tri Tung Nguyen Nguyen, Dinh Tuan Tran, and Joo-Ho Lee. 2024. Finite Scalar Quantization as Facial Tokenizer for Dyadic Reaction Generation. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE, 1–5.

[2] Cafaro et al. 2017. The NoXi database: multimodal recordings of mediated novice-expert interactions. In *ICMI 2017*.

[3] Evonne Ng et al. 2022. Learning to listen: Modeling non-deterministic dyadic facial motion. In *IEEE/CVF CVPR 2022*. 20395–20405.

[4] Fabien Ringeval et al. 2013. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In *IEEE FG 2013*.

[5] Haoyu Song et al. 2019. Exploiting persona information for diverse generation of conversational responses. *arXiv preprint arXiv:1905.12188* (2019).

[6] Palmero et al. 2021. Context-aware personality inference in dyadic scenarios: Introducing the udiva dataset. In *IEEE/CVF 2021*.

[7] Shuna et al. 2019. The affective facial recognition task: The influence of cognitive styles and exposure times. *Journal of Visual Communication and Image Representation* 65 (2019), 102674.

[8] Song et al. 2022. Learning Person-specific Cognition from Facial Reactions for Automatic Personality Recognition. *IEEE Transactions on Affective Computing* (2022).

[9] Song et al. 2023. Multiple Appropriate Facial Reaction Generation in Dyadic Interaction Settings: What, Why and How? *https://arxiv.org/abs/2302.06514* (2023).

[10] Yuchi et al. 2017. Dyadgan: Generating facial expressions in dyadic interactions. In *IEEE CVPR Workshops 2017*. 11–18.

[11] Yoon et al. 2022. The GENEA Challenge 2022: A large evaluation of data-driven co-speech gesture generation. *arXiv preprint arXiv:2208.10441* (2022).

[12] Ximi Hoque, Adamay Mann, Gulshan Sharma, and Abhinav Dhall. 2023. BEAMER: Behavioral Encoder to Generate Multiple Appropriate Facial Reactions. In *Proceedings of the ACM International Conference on Multimedia*. 9536–9540.

[13] Zhenjie Liu, Cong Liang, Jiahe Wang, Haofan Zhang, Yadong Liu, Caichao Zhang, Jialin Gui, and Shangfei Wang. 2024. One-to-Many Appropriate Reaction Mapping Modeling with Discrete Latent Variable. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE, 1–5.

[14] Cheng Luo, Siyang Song, Weicheng Xie, Micol Spitale, Zongyuan Ge, Linlin Shen, and Hatice Gunes. 2024. ReactFace: Online Multiple Appropriate Facial Reaction Generation in Dyadic Interactions. *IEEE Transactions on Visualization and Computer Graphics* (2024).

[15] Cheng Luo, Siyang Song, Siyuan Yan, Zhen Yu, and Zongyuan Ge. 2025. ReactDiff: Fundamental Multiple Appropriate Facial Reaction Diffusion Model. In *Proceedings of the 33rd ACM International Conference on Multimedia*. 5607–5616.

[16] Qirong Mao, Qiwei Wu, Na Liu, Yakui Ding, and Lijian Gao. 2025. Scattering-Conditioned Diffusion Models for Multiple Appropriate Facial Reaction Generation. In *Proceedings of the 33rd ACM International Conference on Multimedia*. 13985–13991.

[17] Fabian Mentzer, David Minnen, Eirikur Agustsson, and Michael Tschannen. 2024. Finite Scalar Quantization: VQ-VAE Made Simple. In *The Twelfth International Conference on Learning Representations*.

[18] Dang-Khanh Nguyen, Prabesh Paudel, Seung-Won Kim, Ji-Eun Shin, Soo-Hyung Kim, and Hyung-Jeong Yang. 2024. Multiple Facial Reaction Generation Using Gaussian Mixture of Models and Multimodal Bottleneck Transformer. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE, 1–5.

[19] Minh-Duc Nguyen, Hyung-Jeong Yang, Ngoc-Huynh Ho, Soo-Hyung Kim, Seungwon Kim, and Ji-Eun Shin. 2024. Vector Quantized Diffusion Models for Multiple Appropriate Reactions Generation. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE, 1–5.

[20] Minh-Duc Nguyen, Hyung-Jeong Yang, Soo-Hyung Kim, Ji-Eun Shin, and Seung-Won Kim. 2025. Latent Behavior Diffusion for Sequential Reaction Generation in Dyadic Setting. In *International Conference on Pattern Recognition*. Springer, 233–248.

[21] Siyang Song, Micol Spitale, Xiangyu Kong, Hengde Zhu, Cheng Luo, Cristina Palmero, German Barquero, Sergio Escalera, Michel Valstar, Mohamed Daoudi, et al. 2025. React 2025: the third multiple appropriate facial reaction generation challenge. In *Proceedings of the 33rd ACM International Conference on Multimedia*. 13979–13984.

[22] Siyang Song, Micol Spitale, Cheng Luo, Germán Barquero, Cristina Palmero, Sergio Escalera, Michel Valstar, Tobias Baur, Fabien Ringeval, Elisabeth André, et al. 2023. React2023: The first multiple appropriate facial reaction generation challenge. In *Proceedings of the 31st ACM International Conference on Multimedia*. 9620–9624.

[23] Siyang Song, Micol Spitale, Cheng Luo, Cristina Palmero, German Barquero, Hengde Zhu, Sergio Escalera, Michel Valstar, Tobias Baur, Fabien Ringeval, et al. 2024. React 2024: the second multiple appropriate facial reaction generation challenge. *arXiv preprint arXiv:2401.05166* (2024).

[24] Weicheng Xie, Chunlin Yan, Siyang Song, Zitong Yu, Linlin Shen, and Laizhong Cui. 2025. Smooth Online Multiple Appropriate Facial Reaction Generation. In *Proceedings of the 33rd ACM International Conference on Multimedia*. 5804–5813.

[25] Tong Xu, Micol Spitale, Hao Tang, Lu Liu, Hatice Gunes, and Siyang Song. 2023. Reversible Graph Neural Network-based Reaction Distribution Learning for Multiple Appropriate Facial Reactions Generation. *arXiv preprint arXiv:2305.15270* (2023).

[26] Jun Yu, Ji Zhao, Guochen Xie, Fengxin Chen, Ye Yu, Liang Peng, Minglei Li, and Zonghong Dai. 2023. Leveraging the Latent Diffusion Models for Offline Facial Multiple Appropriate Reactions Generation. In *Proceedings of the ACM International Conference on Multimedia*. 9561–9565.

[27] Hengde Zhu, Xiangyu Kong, Weicheng Xie, Xin Huang, Xilin He, Lu Liu, Linlin Shen, Zhang Wei, Hatice Gunes, and Siyang Song. 2025. PerReactor: Offline Personalised Multiple Appropriate Facial Reaction Generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

[28] Hengde Zhu, Xiangyu Kong, Weicheng Xie, Xin Huang, Linlin Shen, Lu Liu, Hatice Gunes, and Siyang Song. 2024. Perfrdiff: Personalised weight editing for multiple appropriate facial reaction generation. In *Proceedings of the 32nd ACM International Conference on Multimedia*. 9495–9504.