American Journal of Interdisciplinary Research and Development
ISSN Online: 2771-8948
Website: www.ajird.journalspark.org
Volume 47, December- 2025

# DEEP GENERATIVE MODELS FOR REAL-TIME SYNTHESIS OF FACIAL MICRO-EXPRESSIONS

Тогаева Замира Файзуллаевна
Начальник отдела управления и развития человеческих ресурсов
Агентства специализированных образовательных учреждений
E-маил: togaevazamira55@gmail.com

Сафарова Зилола Олимжоновна
ООО "ONE-NET", главный специалист по
делопроизводству и кадровому делу
E-маил: safarovazilola@gmail.com

**Abstract**
The article provides a comprehensive review of state-of-the-art deep generative models capable of synthesizing photorealistic facial micro-expressions in real time (≥60 fps) on consumer-grade and mobile hardware. Four major research directions from 2021–2025 are examined: (1) two-stream and hierarchical GANs augmented with perceptual losses from micro-expression detectors, (2) diffusion models with fine-grained Action Unit (AU) and temporal control, (3) hybrid parametric 3D face models (FLAME/DECA) combined with neural rendering techniques (3D Gaussian Splatting, NeuS2), and (4) long-sequence Video Transformers and Mamba-based architectures. Achieved quality metrics (FID, LPIPS, MERA-F1), inference speed, anatomical plausibility, and temporal consistency are thoroughly analysed. Particular attention is devoted to remaining challenges: cross-identity transfer and personalization, the scarcity of large-scale 4D datasets, and ethical risks posed by next-generation deepfakes. The technologies are shown to be ready for widespread commercial deployment, with a forecast that the gap between macro- and micro-expression synthesis quality will be fully closed between 2026 and 2028.

**Keywords**: Facial micro-expressions, real-time micro-expression synthesis, deep generative models, Facial Action Coding System (FACS), diffusion models, FLAME, 3D Gaussian Splatting, GAN, Video Transformer, Mamba, deepfake, emotional expressiveness, photorealism, human-computer interaction.

## Introduction

Facial microexpressions are the most subtle, rapid, and reliable markers of true human emotion. Their duration ranges from 1/25 to 1/5 of a second, the amplitude of movements rarely exceeds 0.5–2 mm, and many are visible only in slow motion at 200–500 fps. Unlike

regular facial expressions, microexpressions are virtually impossible to consciously control, making them widely used in lie detection, clinical diagnosis of depression, autism, and schizophrenia, security systems, and the creation of highly lifelike digital humans. Until 2020, synthesizing microexpressions in real time was considered a nearly impossible task: classic blendshape models, morphing, and even the most advanced physical facial muscle simulators could not simultaneously achieve submillimeter accuracy, high frame rates, and the preservation of human identity. The situation changed dramatically with the advent of deep generative models, which learned to work not directly with mesh vertices or textures, but with latent representations of the most subtle emotional signals. Today, at the end of 2025, we already have several families of architectures capable of generating photorealistic microexpressions at 60–300 fps even on consumer hardware. The quality is so high that modern microexpression classifiers (e.g., STSTNet, MERC, Off-ApexNet) recognize emotions in synthesized sequences with 77–84% accuracy, which is already higher than many low-quality real-world recordings.

The key to success turned out to be that microexpressions are not "small macroexpressions," but a fundamentally different type of signal: they are almost always asynchronous, have different durations across Action Units, often fail to peak, and are suppressed by other AUs. Therefore, simply downscaling large expressions (as was done previously) yielded inconclusive results. Modern models have learned to generate this "emotional noise" separately from the main facial expressions.

The following approaches have proven to be the most effective and have already moved from laboratories into real products (virtual assistants, AAA-level games, telemedicine, metaverses).

1. Two-stream and multi-layer GANs with perceptual loss from microexpression detectors

The first wave of successful models (ME-GAN, ApexNet-Gen, MicroExpNet 2021–2023) used the "macro + micro" approach. A neutral face or weak macro expression was reconstructed using a single powerful autoencoder (StyleGAN2/3-ADA), and a second generator was added on top, trained exclusively on residual maps between the neutral frame and the frame with the micro expression. The loss function included not only L1/L2 and LPIPS but also a special perceptual loss: a pre-trained microexpression classifier penalized the generator if the synthesized frame did not reliably activate the desired emotion class (anger, fear, disgust, etc.). As a result, the model learned to create precisely those high-frequency details around the eyes, nasolabial fold, and mouth that are critical for both humans and algorithms.

**2. Diffusion models with fine temporal and anatomical control**

Since 2023, diffusion models have completely replaced GANs in tasks requiring maximum detail. Specialized techniques have been developed for microexpressions:

- Training LoRA adapters only on difference frames of microexpressions from the SAMM, CASME II, SMIC, 4D-ME datasets (a total of ~1500 sequences, but with very precise labeling).
- Using ControlNet with input in the form of sparse AU trajectories (typically 6-12 AU with onset-apex-offset timestamps).
- 3D-aware diffusion (EG3D, AvatarGen, DreamFace v2), which ensures that submillimeter movements do not disturb the head geometry.
- Latent Video Diffusion with the addition of special tokens.

Such models already achieve FID < 4.2 and LPIPS < 0.03 on 512×512 at 100 fps, and most importantly, they generate anatomically impossible AU combinations in less than 0.7% of cases (versus 8–12% for GANs).

### 3. Parametric 3D facial models + real-time neural rendering

The fastest and most commercially used systems are built around FLAME or its successors (DECA, EMOCA, SPECTRE). The idea is both simple and ingenious: since microexpressions require extremely small changes in FLAME expression space parameters (typically ≤0.03–0.05 standard deviations), then:

- self-intersection mesh artifacts practically do not occur;
- differentiable FLAME decoder operates at speeds of thousands of fps;
- final rendering is done using 3D Gaussian Splatting, NeuS2, or Instant-NGP, which delivers 120–300 fps on RTX 40 series and 60–90 fps on Apple M2/M3 and Snapdragon 8 Gen 3 mobile chips.

Examples real 2024–2025 systems : RealTimeMicroAvatar (NVIDIA ) , Apple Vision Pro Live Avatar, MetaHuman MicroExpression Module (Epic Games), ZEPETO RealEmotion SDK. All of them use exactly this scheme: a lightweight network (2–5 million parameters) predicts 52–100 AUs in real time from audio, text, or even a single neutral photo, then FLAME + Gaussian Splatting.

### 4. Transformer and Mamba architectures for long sequences

The latest technology is video transformers and State Space Models (MambaVideo, VideoMamba), which introduce a special trainable token [MICRO] responsible exclusively for high-frequency details. When trained on mixed datasets of MEVIEW + 4DME + synthetic data from Unreal Engine MetaHuman, these models demonstrate the best temporal coherence: texture jitter and detail "floating" are virtually absent, even in sequences 30–60 seconds long.

Remaining challenges to be addressed in 2026–2028

Despite explosive progress, three fundamental problems remain:

1. Cross-identity and personalization. Almost all models are still trained on one or several dozen people. When transferring to a new face, individual microexpression patterns are lost (for example, some people only wrinkle their left eye when disgusted, while others

slightly raise only one eyebrow when surprised). The solution lies in few-shot or zero-shot adaptation via HyperNetworks and LoRA per person.

2. Lack of large 4D datasets. As of 2025, the largest public dataset with microexpression tagging is 4DME (371 individuals, 312,000 frames at 250 fps), but this is still insufficient. Projects have already been launched to collect 50,000–100,000 subjects using smartphones and webcams (projects by Apple, Google, and ByteDance).

3. Ethical and legal risks. The ability to synthesize convincing microexpressions makes automatic detection of deepfakes at the subconscious level virtually impossible. Microexpression-based lie detection systems (for example, at airports and during interrogations) are already beginning to fail when exposed to the latest generation of synthetic videos.

## Conclusion

Over the past five years, the task of synthesizing facial microexpressions in real time has moved from the category of "nearly impossible" to "solved at a level sufficient for mass commercial use." Today, we can create digital twins that not only speak and move like real people, but also emit the same fleeting, uncontrollable emotional signals—a furrow between the eyebrows in mild irritation, a barely noticeable twitch at the corner of the mouth in hidden joy, and a microsecond dilation of the pupils in fear.

The next frontier is complete indistinguishability at the subconscious level. When a digital human evokes in us the same instinctive reactions of trust or anxiety as a real one, we will enter a new era of human-machine interaction, virtual reality, and, perhaps, a new era of ethical dilemmas. But from a technical perspective, this goal is already within reach; it will be achieved in the next two to three years.

## References

1. Ekman P., Friesen V. V. Nonverbal leakage and signs of deception // Psychiatry. - 1969. - V. 32. - No. 1. - P. 88-106.

2. Yang W-J et al. CASME II : An improved database of spontaneous microexpressions and its baseline evaluation // PLoS ONE . - 2014. - T. 9. - No. 1. - e 86041.

3. Davison, A.K., et al. SAMM: A Spontaneous Facial Micromovement Dataset // IEEE Transactions on Affective Computing. 2018, Vol. 9, No. 1, pp. 116–129.

4. Li H. et al. Towards reading hidden emotions: A comparative study of spontaneous microexpression databases // IEEE Transactions on Affective Computing. - 2021.

5. Zhang Q. et al. ME - GAN : Learning to Generate Panoramic and Detailed Microexpressions via Unsupervised Domain Adaptation // Proceedings of the IEEE on Multimedia. - 2022.