The Second Workshop on Evaluation for Generative Foundation Models (EvGenFM)

CVPR 2026 | June 3–4, 2026 | Denver, CO

The landscape of generative AI has transformed dramatically since our inaugural workshop in 2024. Image generation has reached photorealistic fidelity with models such as FLUX, Imagen 3, and Stable Diffusion 3; long-form video generation has emerged as a new frontier with systems including Sora, Runway Gen-3, and Seedance; multimodal large language models (MLLMs), GPT-4o, Gemini 1.5/2.0, and LLaVA-NeXT have blurred the boundary between perception and generation; and 3D/4D content synthesis has matured toward production-ready systems. Concurrently, reinforcement learning-based post-training (RLHF, GRPO, DPO) has become a standard tool for aligning generative models to human preferences, introducing entirely new evaluation challenges around reward hacking, distributional shift, and alignment verification.

Yet evaluation has struggled to keep pace. Automatic metrics frequently fail to capture nuanced human preferences, benchmarks saturate faster than they can be replaced, and the emergence of instruction-following, long-context reasoning, and cross-modal generation demands fundamentally new evaluation paradigms.

The 2nd Workshop on Evaluation for Generative Foundation Models at CVPR 2026 aims to build a forum to discuss ongoing efforts in industry and academia, share best practices, and engage the community in working towards more reliable and scalable approaches for GenFM evaluation. We invite original research papers on the evaluation of Generative Foundation Models (GenFMs) across all modalities, including but not limited to text, images, video, 3D models, speech, and music. We encourage submissions in the following areas:

- Presentation of new datasets and benchmarks tailored for comprehensive GenFM evaluation.
- Human-in-the-Loop methodologies that incorporate human judgment and preferences.
- LLM and MLLM-as-judge paradigms: their reliability, calibration, and biases when evaluating generative outputs.
- Demonstrations of evaluation methodology across diverse generative tasks and public datasets across modalities (e.g., T2I-CompBench, GenAI-Bench, EvalCrafter, VBench, VideoScore, MusicCaps, Objaverse-XL, T3Bench, etc.).
- Comprehensive analyses of existing evaluation techniques and benchmarks (e.g., HELM, HEIM, GenAI-Bench, VBench, WildBench, Arena-Hard, Chatbot Arena, MT-Bench).
- Evaluation of models trained with Reinforcement Learning (GRPO, DAPO, GSPO), including reward model evaluation and reward hacking detection and mitigation.
- Evaluation of long-form and long-context generation: coherence, temporal consistency, and narrative quality in video and multi-turn settings.
- Discussion of unique challenges of multimodal GenFM evaluation, including 3D/4D generative models.
- Quantification of emergent capabilities and the interplay between modalities within GenFMs.
- Exploration of techniques to leverage automated evaluation for model refinement.

Accepted papers will be included in the CVPR proceedings, on IEEE Xplore, and on the CVF website.

Paper Submission Deadline: March 21, 2026, 23:59 Pacific Time
Notification to Authors: March 25, 2026, 23:59 Pacific Time
Submission Format: CVPR 2026 paper format
Workshop Website: https://evgenfm2026.github.io/
Submissions: cmt