Mochi 1: A new SOTA in open-source video generation models

Mochi 1 preview is an open state-of-the-art video generation model with high-fidelity motion and strong prompt adherence. Our new model dramatically closes the gap between closed and open video generation systems. We're releasing the model under a permissive Apache 2.0 license. Download the weights now or try this preview version for free on our playground.

Introduction

We are thrilled to announce a research preview of Mochi 1, our latest open-source video generation model. Mochi 1 demonstrates dramatic improvements in quality of motion as well as extremely strong prompt adherence. Licensed under the Apache 2.0 license, a preview of Mochi 1 is freely available for personal and commercial use.

In addition to the model release, we're excited to unveil our hosted playground, where you can try Mochi 1 for free today at genmo.ai/play. The weights and

We're also pleased to share that Genmo has raised a $28.4 million Series A funding round led by NEA led by Rick Yang with participation from The House Fund, Gold House Ventures, WndrCo, Eastlink Capital Partners, and Essence VC, as well as angel investors Abhay Parasnis (CEO of Typespace), Amjad Masad (CEO of Replit), Sabrina Hahn, Bonita Stewart, and Michele Catasta.

At Genmo, our mission is to unlock the right brain of artificial general intelligence. Mochi 1 is the first step toward building world simulators that can imagine anything, whether possible or impossible.

Our team includes core members of projects like DDPM (Denoising Diffusion Probabilistic Models), DreamFusion, and Emu Video. Genmo is advised by leading technical experts, including Ion Stoica (Executive Chairman and co-founder of Databricks and Anyscale), Pieter Abbeel (co-founder of Covariant and early team at OpenAI), and Joey Gonzalez (pioneer in language model systems and co-founder of Turi).

Evaluations

Today, there is an enormous gap between video generation models and reality. Motion quality and prompt adherence are two of the most critical capabilities that are still missing from video generation models.

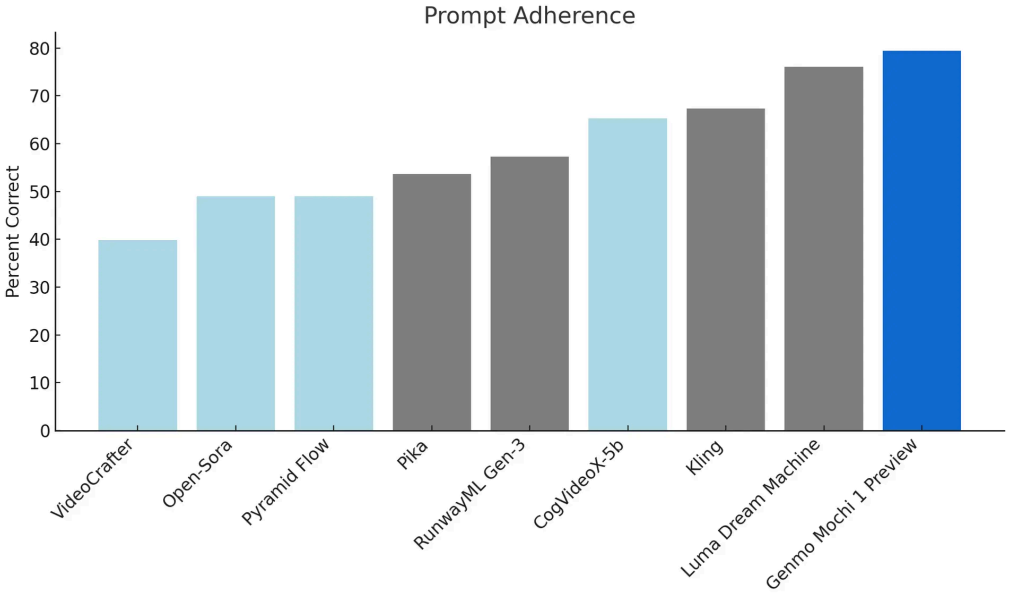
Mochi 1 sets a new best-in-class standard for open-source video generation. It also performs very competitively with the leading closed models. Specifically, our 480p preview has strong:

- Prompt Adherence: Demonstrates exceptional alignment with textual prompts, ensuring that generated videos accurately reflect the given instructions. This allows users detailed control over characters, settings and actions. We benchmark prompt adherence with an automated metric using a vision language model as a judge following the protocol in OpenAI DALL-E 3. We evaluate generated videos using Gemini-1.5-Pro-002.
- Motion Quality: Mochi 1 generates smooth videos at 30 frames per second for durations up to 5.4 seconds, with high temporal coherence and realistic motion

uncanny valley. Raters were instructed to focus on motion rather than frame-level aesthetics (criteria include interestingness of the motion, physical plausibility, and fluidity). Elo scores are computed following the LMSYS Chatbot Arena protocol.

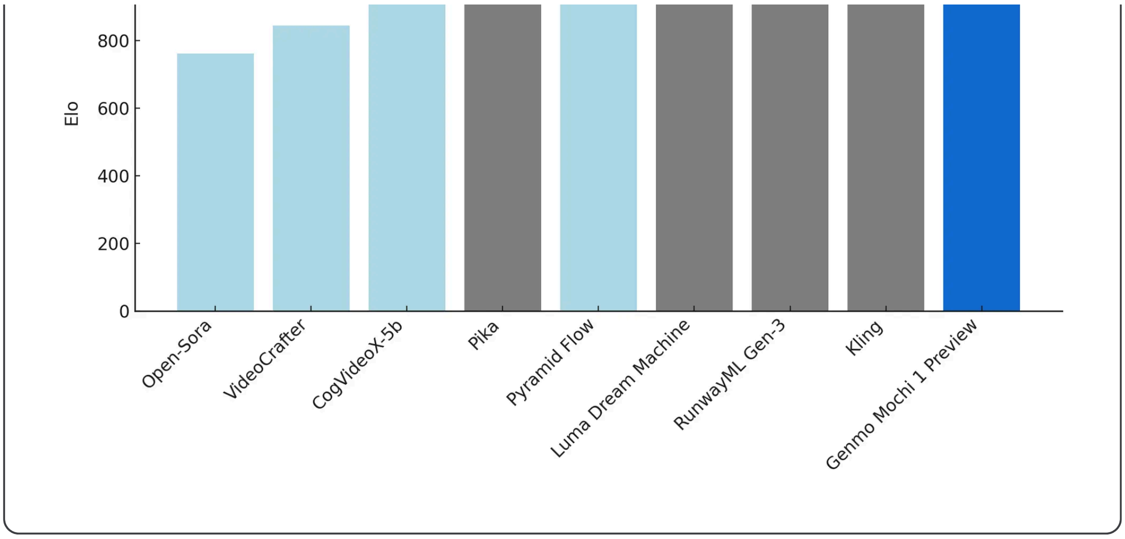
Prompt Adherence

Measures how accurately generated videos follow the provided textual instructions, ensuring high fidelity to user intent.



Elo Score

Evaluates both motion smoothness and spatial realism, ensuring that generated videos are fluid and visually captivating.



Limitations

Under the research preview, Mochi 1 is a living and evolving checkpoint. There are a few known limitations. The initial release generates videos at 480p today. In some edge cases with extreme motion, minor warping and distortions can also occur. Mochi 1 is also optimized for photorealistic styles so does not perform well with animated content. We also anticipate that the community will fine-tune the model to suit various aesthetic preferences. Additionally, we have implemented robust safety moderation protocols in the playground to ensure that all video generations remain safe and aligned with ethical guidelines.

Model Architecture

Mochi 1 represents a significant advancement in open-source video generation, featuring a 10 billion parameter diffusion model built on our novel Asymmetric Diffusion Transformer (AsymmDiT) architecture. Trained entirely from scratch, it is the largest video generative model ever openly released. And best of all, it’s a simple, hackable architecture.

Efficiency is critical to ensure the community can run our models. Alongside Mochi, we are open-sourcing our video VAE. Our VAE causally compresses

An AsymmDiT efficiently processes user prompts alongside compressed video tokens by streamlining text processing and focusing neural network capacity on visual reasoning. AsymmDiT jointly attends to text and visual tokens with multi-modal self-attention and learns separate MLP layers for each modality, similar to Stable Diffusion 3. However, our visual stream has nearly 4 times as many parameters as the text stream via a larger hidden dimension. To unify the modalities in self-attention, we use non-square QKV and output projection layers. This asymmetric design reduces inference memory requirements.

Many modern diffusion models use multiple pretrained language models to represent user prompts. In contrast, Mochi 1 simply encodes prompts with a single T5-XXL language model.

Mochi 1 jointly reasons over a context window of 44,520 video tokens with full 3D attention. To localize each token, we extend learnable rotary positional embeddings (RoPE) to 3-dimensions. The network end-to-end learns mixing frequencies for space and time axes.

Mochi benefits from some of the latest improvements in language model scaling including SwiGLU feedforward layers, query-key normalization for enhanced stability, and sandwich normalization to control internal activations.

A technical paper will follow with additional details to encourage progress in video generation.

Open-Source Release

We are releasing Mochi 1 under the Apache 2.0 license. It’s critical that there is an open research ecosystem around video generation. We believe that open-source models drive progress and democratize access to state-of-the-art AI capabilities.

Try Mochi 1 today

—all for free.

Developer Resources

We have partnered with leading platforms to make Mochi 1 easily accessible:

- Open weights: Download the weights from huggingface.co/genmo or via magnet link.
- GitHub Repository: Access the source code at github.com/genmoai/models.
- APIs partners: Integrate Mochi 1 into your applications seamlessly using APIs from our partners.

Applications

Our research preview of Mochi 1 opens up new possibilities across various domains:

- Research and Development: Advance the field of video generation and explore new methodologies.
- Product Development: Build innovative applications in entertainment, advertising, education, and more.
- Creative Expression: Empower artists and creators to bring their visions to life with AI-generated videos.
- Robotics: Generate synthetic data for training AI models in robotics, autonomous vehicles and virtual environments.

What is coming next?

Today, we are releasing the Mochi 1 preview, showcasing the capabilities of our 480p base model. But this is just the beginning. Before the end of the year, we will release the full version of Mochi 1, which includes Mochi 1 HD. Mochi 1 HD will support 720p video generation with enhanced fidelity and even smoother motion, addressing edge cases such as warping in complex scenes.

models to give our users even more precise control over their outputs.

Future vision

The Mochi 1 preview has limitations including a 480p resolution for computational efficiency on end-user devices. Looking forward, we will continue to advance the SOTA in video generation with support for high-resolution, long video generation as well as image-to-video synthesis.

Join us

Mochi 1 represents a significant leap forward in open-source video generation. We invite you to join us at the frontier of the right brain of intelligence. We are hiring strong researchers and engineers to join our team: https://genmo.ai/careers.

Try Mochi 1 today at genmo.ai/play and be part of the future of video generation.

COMPANY

OPEN SOURCE

PRODUCT

CONNECT

All Systems Operational

Terms of Service

Privacy Policy