Ten months after first tease, OpenAI launches Sora video generation publicly

[embedded content]

A music video by Canadian art collective Vallée Duhamel made with Sora-generated video. “[We] just shoot stuff and then use Sora to combine it with a more interesting, more surreal vision.”

During a livestream on Monday—during Day 3 of OpenAI’s “12 days of OpenAi”—Sora’s developers showcased a new “Explore” interface that allows people to browse through videos generated by others to get prompting ideas. OpenAI says that anyone can enjoy viewing the “Explore” feed for free, but generating videos requires a subscription.

They also showed off a new feature called “Storyboard” that allows users to direct a video with multiple actions in a frame-by-frame manner.

Safety measures and limitations

In addition to the release, OpenAI also publish Sora’s System Card for the first time. It includes technical details about how the model works and safety testing the company undertook prior to this release.

“Whereas LLMs have text tokens, Sora has visual patches,” OpenAI writes, describing the new training chunks as “an effective representation for models of visual data… At a high level, we turn videos into patches by first compressing videos

into a lower-dimensional latent space, and subsequently decomposing the representation into spacetime patches."

Sora also makes use of a "recaptioning technique"—similar to that seen in the company's DALL-E 3 image generation, to "generate highly descriptive captions for the visual training data." That, in turn, lets Sora "follow the user's text instructions in the generated video more faithfully," OpenAI writes.

Sora-generated video provided by OpenAI, from the prompt: "Loop: a golden retriever puppy wearing a superhero outfit complete with a mask and cape stands perched on the top of the empire state building in winter, overlooking the nyc it protects at night. the back of the pup is visible to the camera; his attention faced to nyc"

Sora-generated video provided by OpenAI, from the prompt: "Loop: a golden retriever puppy wearing a superhero outfit complete with a mask and cape stands perched on the top of the empire state building in winter, overlooking the nyc it protects at night. the back of the pup is visible to the camera; his attention faced to nyc"

OpenAI implemented several safety measures in the release. The platform embeds C2PA metadata in all generated videos for identification and origin verification. Videos display visible watermarks by default, and OpenAI developed an internal search tool to verify Sora-generated content.

The company acknowledged technical limitations in the current release. "This early version of Sora will make mistakes, it's not perfect," said one developer during the livestream launch. The model reportedly struggles with physics simulations and complex actions over extended durations.

In the past, we've seen that these types of limitations are based on what example videos were used to train AI models. This current generation of AI video-synthesis models has

difficulty generating truly new things, since the underlying architecture excels at transforming existing concepts into new presentations, but so far typically fails at true originality. Still, it’s early in AI video generation, and the technology is improving all the time.

https://arstechnica.com/ai/2024/12/ten-months-after-first-tease-openai-launches-sora-video-generation-publicly/