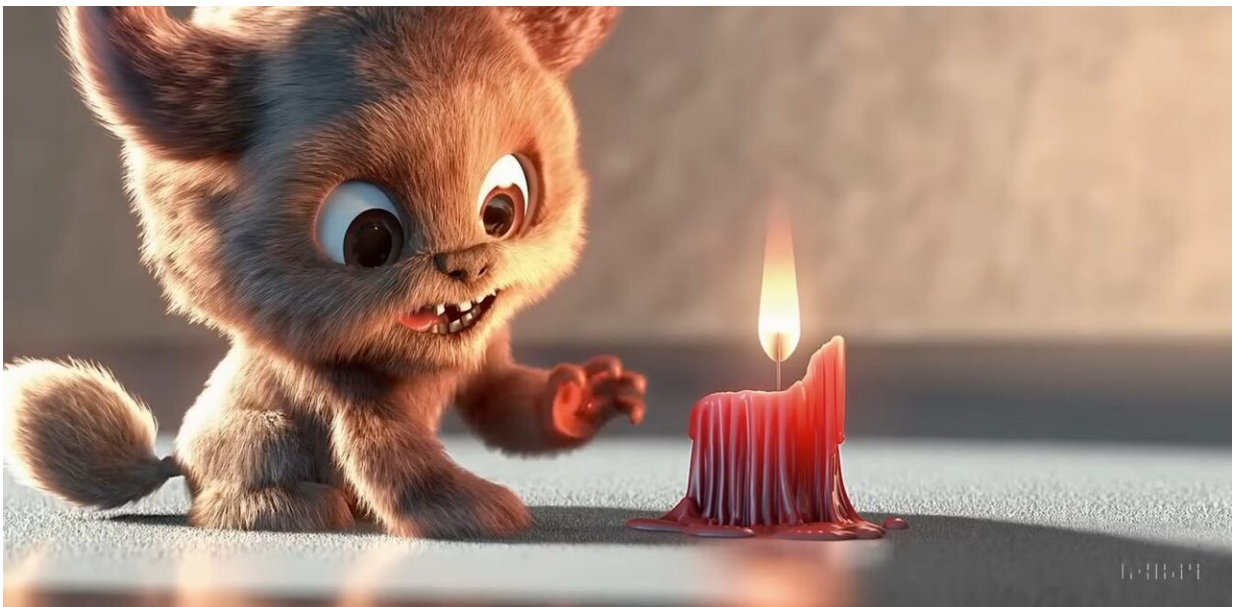# What is Sora? A new generative AI tool could transform video production and amplify disinformation risks

February 20 2024, by Vahid Pooryousef and Lonni Besançon

Credit: OpenAI

Late last week, OpenAI announced a new generative AI system named Sora, which produces short videos from text prompts. While Sora is not yet available to the public, the high quality of the sample outputs published so far has provoked both excited and concerned reactions.

The sample videos published by OpenAI, which the company says were

created directly by Sora without modification, show outputs from prompts like "photorealistic closeup video of two pirate ships battling each other as they sail inside a cup of coffee" and "historical footage of California during the gold rush."

At first glance, it is often hard to tell they are generated by AI, due to the high quality of the videos, textures, dynamics of scenes, camera movements, and a good level of consistency.

OpenAI chief executive Sam Altman also posted some videos to X (formerly Twitter) generated in response to user-suggested prompts, to demonstrate Sora's capabilities.

How does Sora work?

Sora combines features of text and image generating tools in what is called a "diffusion transformer model".

Transformers are a type of neural network first introduced by Google in 2017. They are best known for their use in large language models such as ChatGPT and Google Gemini.

Diffusion models, on the other hand, are the foundation of many AI image generators. They work by starting with random noise and iterating towards a "clean" image that fits an input prompt.

A video can be made from a sequence of such images. However, in a video, coherence and consistency between frames are essential.

Sora uses the transformer architecture to handle how frames relate to one another. While transformers were initially designed to find patterns in tokens representing text, Sora instead uses tokens representing small patches of space and time.

Leading the pack

Sora is not the first text-to-video model. Earlier models include Emu by Meta, Gen-2 by Runway, Stable Video Diffusion by Stability AI, and recently Lumiere by Google.

Lumiere, released just a few weeks ago, claimed to produce better video than its predecessors. But Sora appears to be more powerful than Lumiere in at least some respects.

Sora can generate videos with a resolution of up to 1920 × 1080 pixels, and in a variety of aspect ratios, while Lumiere is limited to 512 × 512 pixels. Lumiere's videos are around five seconds long, while Sora makes videos up to 60 seconds.

Lumiere cannot make videos composed of multiple shots, while Sora can. Sora, like other models, is also reportedly capable of video-editing tasks such as creating videos from images or other videos, combining elements from different videos, and extending videos in time.

Both models generate broadly realistic videos, but may suffer from hallucinations. Lumiere's videos may be more easily recognized as AI-generated. Sora's videos look more dynamic, having more interactions between elements.
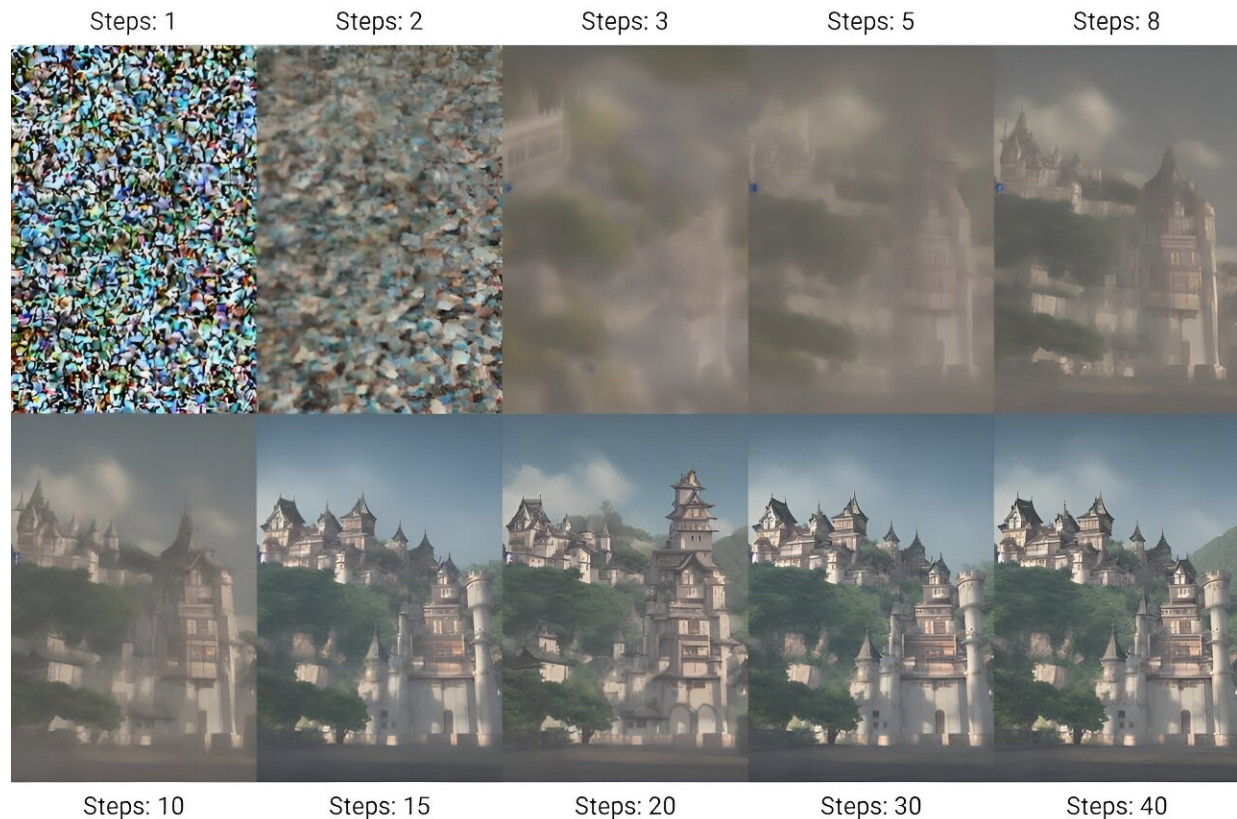
Diffusion models (in this case Stable Diffusion) generate images from noise over many iterations. Credit: Stable Diffusion / Benlisquare / Wikimedia, CC BY-SA

However, in many of the example videos inconsistencies become apparent on close inspection.

Promising applications

Video content is currently produced either by filming the real world or by using special effects, both of which can be costly and time consuming. If Sora becomes available at a reasonable price, people may start using it as a prototyping software to visualize ideas at a much lower cost.

Based on what we know of Sora's capabilities it could even be used to create short videos for some applications in entertainment, advertising and education.

OpenAI's technical paper about Sora is titled "Video generation models as world simulators." The paper argues that bigger versions of video generators like Sora may be "capable simulators of the physical and digital world, and the objects, animals and people that live within them."

If this is correct, future versions may have scientific applications for physical, chemical, and even societal experiments. For example, one might be able to test the impact of tsunamis of different sizes on different kinds of infrastructure—and on the physical and mental health of the people nearby.

Achieving this level of simulation is highly challenging, and some experts say a system like Sora is fundamentally incapable of doing it.

A complete simulator would need to calculate physical and chemical reactions at the most detailed levels of the universe. However, simulating a rough approximation of the world and making realistic videos to human eyes might be within reach in the coming years.

Risks and ethical concerns

The main concerns around tools like Sora revolve around their societal and ethical impact. In a world already plagued by disinformation, tools like Sora may make things worse.

It's easy to see how the ability to generate realistic video of any scene you can describe could be used to spread convincing fake news or throw doubt on real footage. It may endanger public health measures, be used to influence elections, or even burden the justice system with potential

[fake evidence](link).

Video generators may also enable direct threats to targeted individuals, via deepfakes—particularly [pornographic ones](link). These may have terrible repercussions on the lives of the affected individuals and their families.

Beyond these concerns, there are also questions of copyright and intellectual property. Generative AI tools require vast amounts of data for training, and OpenAI has not revealed where Sora's training data came from.

Large language models and image generators have also been criticized for this reason. In the United States, a [group of famous authors have sued OpenAI](link) over a potential misuse of their materials. The case argues that [large language models](link) and the companies who use them are stealing the authors' work to create new content.

It is not the first time in recent memory that technology has run ahead of the law. For instance, the question of the obligations of social media platforms in moderating content has created heated debate in the past couple of years—much of it revolving around [Section 230 of the US Code](link).

While these concerns are real, based on past experience we would not expect them to stop the development of [video](link)-generating technology. OpenAI [says](link) it is "taking several important safety steps" before making Sora available to the public, including working with experts in "misinformation, hateful content, and bias" and "building tools to help detect misleading content."

This article is republished from [The Conversation](link) under a Creative Commons license. Read the [original article](link).

Provided by The Conversation

Citation: What is Sora? A new generative AI tool could transform video production and amplify disinformation risks (2024, February 20) retrieved 8 May 2024 from https://techxplore.com/news/2024-02-sora-generative-ai-tool-video.html