Sora Explained, Kinda.

Roger Lam
Iamroger.com

Sora is OpenAI's SOTA text-to-video model

Sora was released Feb 2024 but is still behind closed doors.

It's quality is stunning - no more Will Smith eating spaghetti.

Videos are up to 1 minute long.

Other text-to-video models include Runway's Gen 2, Pika Labs's Pika, and Google's Lumiere.
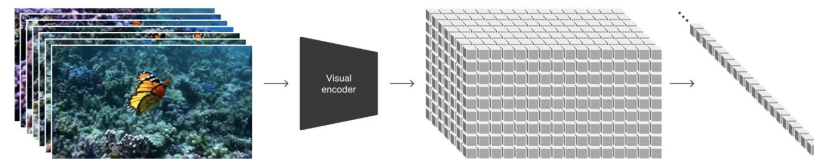

https://openai.com/sora

Like Stable Diffusion, Sora is a Latent Diffusion Model

Videos = Multiple image

Instead of working on individual latent representations, they create patches - sequences of latent representations.
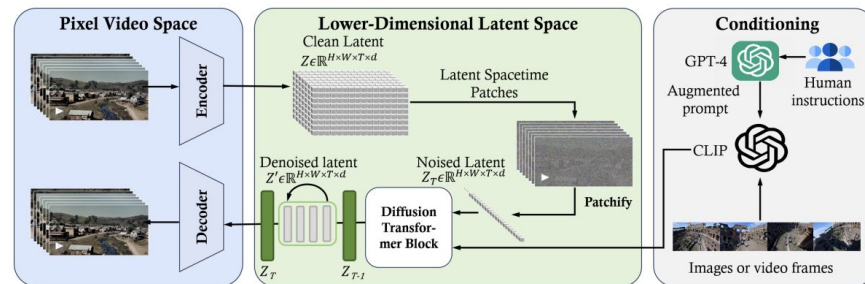
The data accounts for spatial and temporal dimensions.

https://openai.com/research/video-generation
-models-as-world-simulators

Figure 4: **Reverse Engineering**: Overview of `Sora` framework

https://arxiv.org/pdf/2402.17177.pdf

More data and compute

They trained on native resolutions vs 4-sec 256x256 videos.

Re-captioning of videos with a highly descriptive captioner model.

GPT augments short user prompts into longer detailed captions.

an adorable kangaroo wearing a green dress and a sun hat taking a pleasant stroll in Antarctica during a winter storm

Emergent Properties

Similar to the GPT-2 to GPT-3 jump, the videos generated by Sora really WOWed people.

Described at "emergent abilities" - or new things it can do right

3D consistency, long range coherence and consistency, simulating digital worlds, interacting with the world
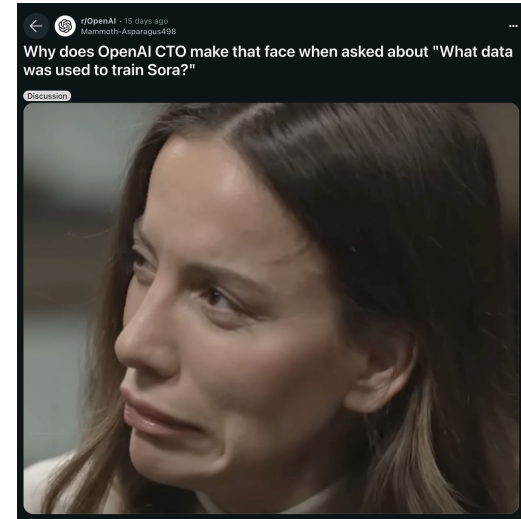
Lots not disclosed

How many parameters is the model? 100x? 1000x?

How long did each video take to generate?

What data did they use to train?

How much would it cost to use? How much did it cost to train?

YouTube Says OpenAI Training Sora With Its Videos Would Break Rules

Neal Mohan, chief executive officer of YouTube *Photographer: Kyle Grillot/Bloomberg*
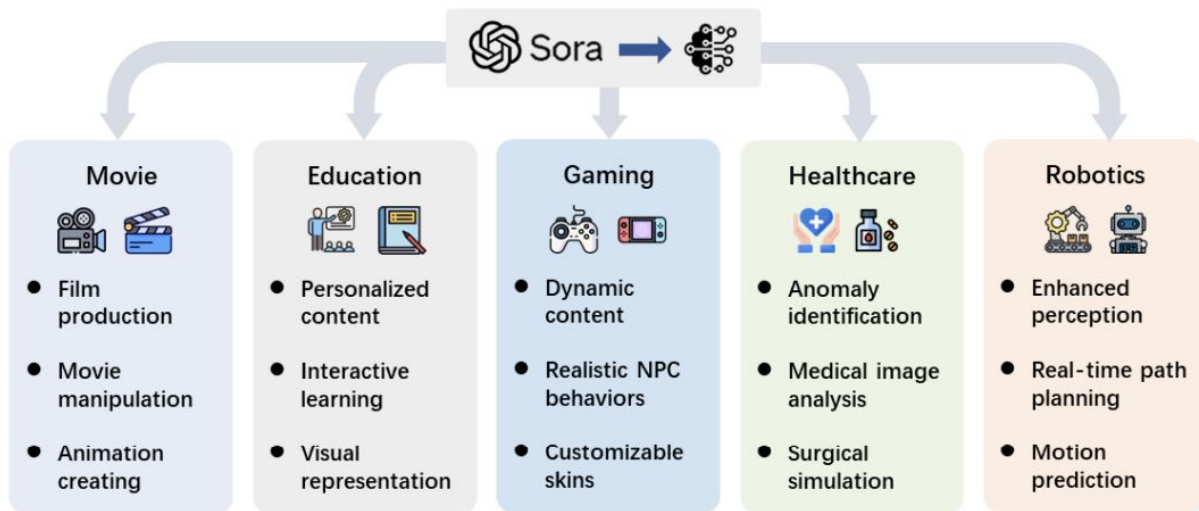
Lots of applications

Figure 18: Applications of Sora.

https://arxiv.org/pdf/2402.17177.pdf

More Resources

Redditor possiblyquestionable explains [spacetime patches](#).

OpenAI's technical page on [Sora](#).

[Speculative paper](#) on Sora and other techniques.