

# SORA: AI Video Generation

Ryan Bowering, Matteo Calviello

# Introduction

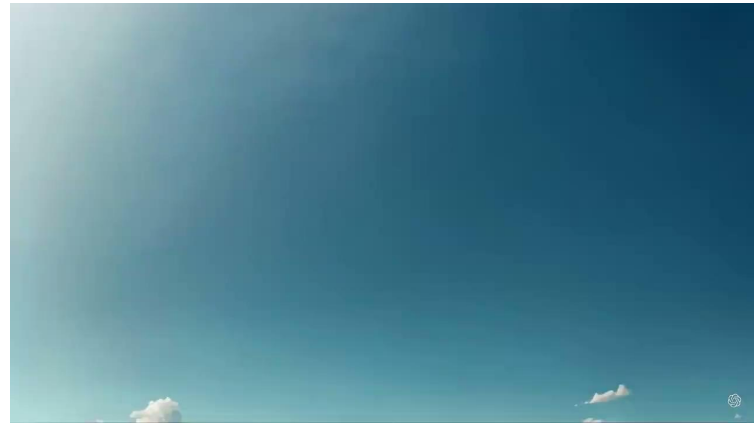
Text-to-video generative AI model

Released by OpenAI in February 2nd 2024

Ability to produce up to 1-minute long high quality coherent videos

Trained on internet-scale data

Uses visual patches - most effective representations for models of visual data



# Highlights and Limitations

## Highlights:

- Improve simulation abilities
- Boost creativity
- Drive educational innovations
- Enhance accessibility

## Limitations:

- Hard to capture facial expressions
- Hard to capture complex actions
- Biases
- Cannot prevent harmful/ false output



Videos up to 1 Minute Long!



# Text to Video

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



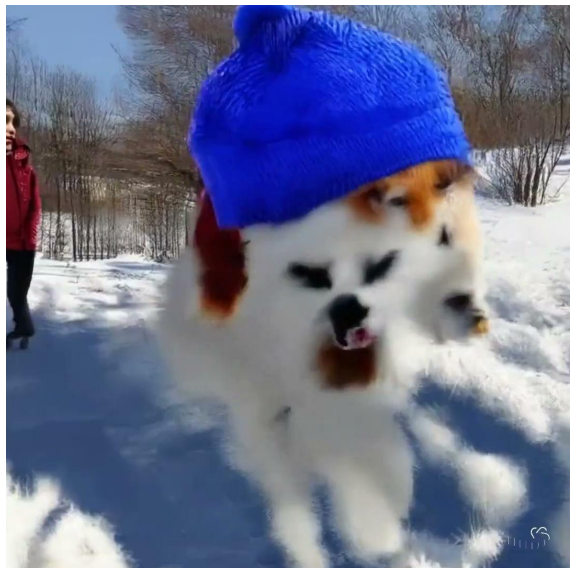
# Image to Video



# Video to Video



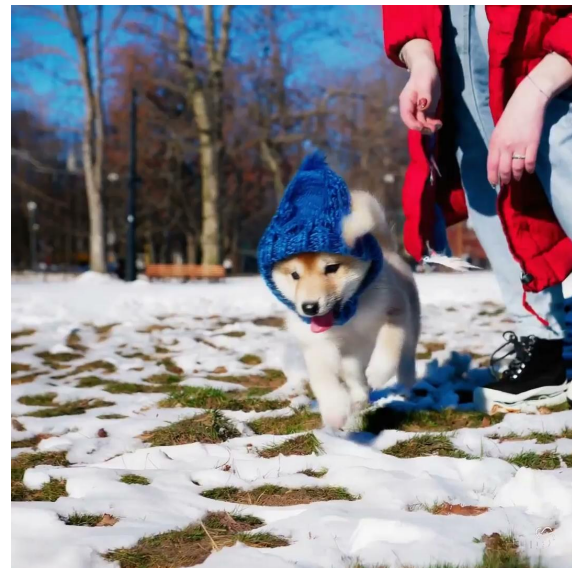
# Scaling Quality



Base Compute



4x Compute



32x Compute



# Scaling Resolution



# Image Generation: why use DALL-E when Sora does it better



Close-up portrait shot of a woman in autumn, extreme detail, shallow depth of field



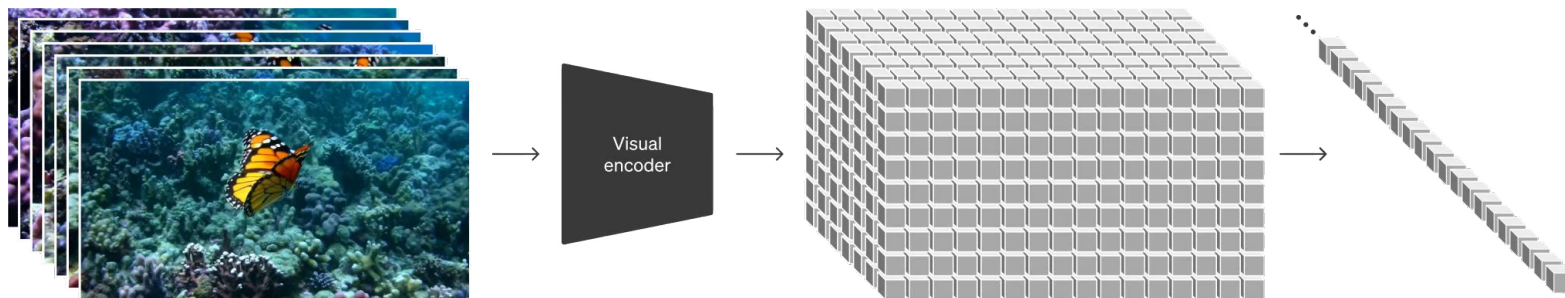
A snowy mountain village with cozy cabins and a northern lights display, high detail and photorealistic dslr, 50mm f/1.2



Digital art of a young tiger under an apple tree in a matte painting style with gorgeous details

# Technical: Video to Patch Transformation

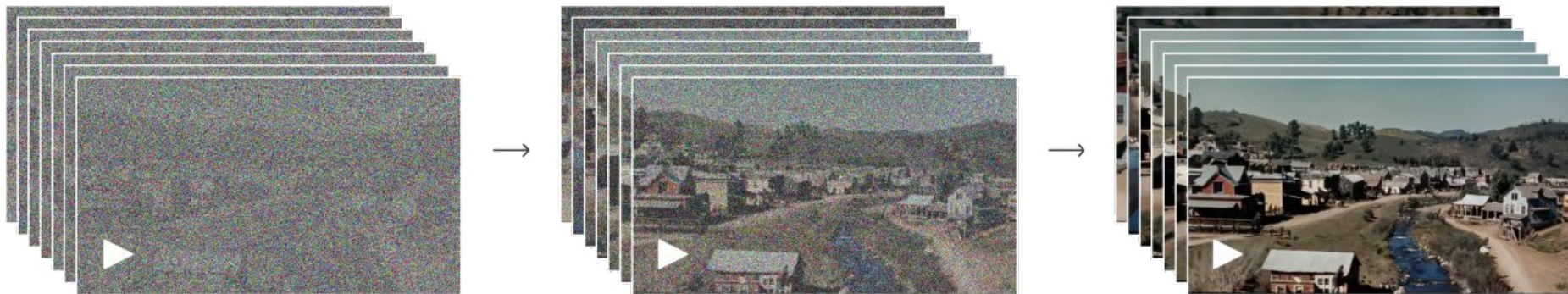
- Video Compression Network: compress visual data into a lower-dimensional latent space, temporally and spatially.
- Spacetime Latent Patches: serve as transformer tokens for training the generative model.
- Patch-based Representation: enables Sora to train on videos and images of variable characteristics.



# Technical: Transformers for Video Generation

Diffusion Transformers:

- Given input noisy patches and conditioning information, Sora predicts the original "clean" patches.
- Sora's effective scalability allows it to efficiently utilize increasing computational resources during training, resulting in higher-quality video samples.



# Lack of Information

- No official paper, just a technical report article
- Closed source
- OpenAI keeps a lot of details secret

# Ethical Considerations

- 1) Spread of false information:
  - a) Social Media
  - b) Elections
  - c) Ads
- 2) Bias:
  - a) Stereotypes
  - b) Under or Overrepresentation of certain groups

Questions?