ModelScope Video Diffusion (MSVD)

✔ Text → Video
✔ Very popular open-source model
✔ Runs on GPU or CPU (slow on CPU)

□ Install

```
pip install diffusers transformers accelerate torch torchvision imageio
```

▶□ Code text-video.py :-

```
from diffusers import DiffusionPipeline
import torch
import imageio

pipe = DiffusionPipeline.from_pretrained(
    "damo-vilab/text-to-video-ms-1.7b",
    torch_dtype=torch.float16
).to("cuda")

prompt = "A monkey riding a horse, cinematic"

result = pipe(prompt, num_frames=24, fps=8)
video = result.frames

imageio.mimsave("msvd_video.mp4", video, fps=8)
```

and run it :-

python text-video.py

IF you get any error then try following code and steps :-

FULL FIXED CODE — ModelScope

The **ModelScope Text-to-Video** model used in HuggingFace Diffusers is:

- **damo-vilab/text-to-video-ms-1.7b**

This is the **official** and **correct** model name for:
✔ Text → Video
✔ Open-source
✔ Supported by Diffusers
✔ FP16 + CUDA compatible

Text-to-Video (works today)

1. Install CUDA PyTorch:

```
pip install torch torchvision --index-url
https://download.pytorch.org/whl/cu121
```

2. Install diffusers dependencies:

```
pip install diffusers transformers accelerate imageio
```

3. Working script text-video1.py :

```
from diffusers import DiffusionPipeline
import torch
import imageio

device = "cuda" if torch.cuda.is_available() else "cpu"
print("Using device:", device)

pipe = DiffusionPipeline.from_pretrained(
    "damo-vilab/text-to-video-ms-1.7b",
    torch_dtype=torch.float16,
    variant="fp16"
).to(device)

prompt = "A monkey riding a horse, cinematic lighting"

result = pipe(prompt, num_frames=24)
frames = result.frames
```

```
imageio.mimsave("output_video.mp4", frames, fps=8)
print("Saved: output_video.mp4")
```

run it :-

python text-video1.py