Open source · runs locally

A faceless Short from a script, with open-source tools

Write a script, voice it with Piper text-to-speech, lay it over your own footage or generated B-roll, and add captions with Whisper. Everything runs on your computer.

VideoToShorts · https://videotoshortsconverter.com/workflows/faceless-short-from-script-local · checked November 19, 2026

COST
Free

TIME
About 30 minutes per Short

LEVEL
Intermediate (command line)

HARDWARE
Any recent computer for voice and captions; a strong GPU only if you generate B-roll locally.

What you need

Piper: Local text-to-speech voices

Whisper: Captions timed to the voice

FFmpeg: Combine voice, footage and captions

Wan 2.2: Alibaba model, open weights

LTX-2.5: Lightricks model, open weights

Best for: Explainers and listicles, Channels that don't show a face, Testing ideas quickly.

Steps

1 Write a short script

Aim for 120 to 150 words for about a minute. Put the hook in the first sentence and keep one idea per sentence. Save it as script.txt.

2 Install Piper and download a voice

Listen to the voice samples first; each voice has its own license on its model card.

```
pip install piper-tts
python3 -m piper.download_voices en_US-lessac-medium
```

3 Generate the voiceover

--input-file reads the script; --sentence-silence adds a short pause between sentences.

```
python3 -m piper -m en_US-lessac-medium --input-file script.txt --sentence-silence 0.3
-f voice.wav
```

4 Get background footage

Use your own clips or stock you have rights to, or generate 5-second B-roll with an open model (see our Wan 2.2 in ComfyUI workflow). A vertical 1080x1920 clip works best.

5 Combine footage and voice

-stream_loop -1 repeats the background until the voice ends; -shortest stops at the end of the voice.

```
ffmpeg -stream_loop -1 -i broll.mp4 -i voice.wav -map 0:v -map 1:a -vf
"scale=1080:1920:force_original_aspect_ratio=increase,crop=1080:1920" -c:v libx264 -
pix_fmt yuv420p -c:a aac -shortest draft.mp4
```

6 Caption it

Transcribe the draft so captions match the voice exactly, then burn them in.

```
whisper draft.mp4 --model small --language en --word_timestamps True --
max_line_width 24 --max_line_count 1 --output_format srt
ffmpeg -i draft.mp4 -vf
"subtitles=draft.srt:force_style='FontName=Arial,FontSize=16,Bold=1,Outline=3,Alignment=2,MarginV=
-c:a copy final.mp4
```

Questions

Can I use the Piper voices commercially?

It depends on the voice. Piper itself is open source, and each voice model lists its own license and training data on its model card. Check it before monetizing.

Does YouTube allow AI voiceovers?

Platforms have their own rules on synthetic media and disclosure. Check the current policy of each platform you post to.

Sources

- Piper CLI documentation
- Whisper CLI options
- FFmpeg documentation (stream_loop, scale, crop)

Commands checked against each tool's docs on November 19, 2026.