

Workflow “AI Filmmaking is finally here!”

This guide shows you how to use midjourney to generate consistent characters and environments, how to bring them to life with a simple click, and how to give your characters AI-generated voices and facial animations.

This guide is based on the script of my YouTube video “[Generate entire movies with AI!](#)”, so you can follow along.



Some basic knowledge for the following tools is required:

- Video Editor (After Effects, Premiere, Resolve...)
- Stable Diffusion WebUI

Tools used for this workflow:

- midjourney: <https://www.midjourney.com/>
- elevenlabs: <https://elevenlabs.io/>
- Stable Diffusion WebUI: <https://github.com/AUTOMATIC1111/stable-diffusion-webui>
- Adobe Creative Cloud: <https://www.adobe.com/>
- DaVinci Resolve: <https://www.blackmagicdesign.com/de/products/davinciresolve>
- runway GEN-1: <https://research.runwayml.com/gen1>
- Wav2Lip: https://colab.research.google.com/github/justinjohn0306/Wav2Lip/blob/master/Wav2Lip_simplified_v5.ipynb#scrollTo=Qgo-oal3JU2u

Step 1: Consistent directable characters & sets in midjourney [00:44](#)

Keeping characters consistent is a constant challenge with AI images. There are tools to train models for the faces of specific characters like [DreamBooth](#), and tools to replace the faces of characters after generation, like [Roop](#). But I would like to try to achieve this with Midjourney only!

1 So the first step is to create a character sheet for all my characters, let's start with a full body T-pose. Enter keywords like **T-pose, symmetrical, full body, arms outstretched, front view** and it should work. If it doesn't, you can use an additional image prompt with any image of a character in a T-pose. I used a screenshot of this 3d model that I found online. Upload the screenshot to Discord and copy the URL. Paste it into your prompt, add the rest of the character descriptions, and set the image weight to about 0.2. Now your character's pose should be more consistent. If there are still some parts cut off, you can bring them into the image using Midjourneys panning tool.



2 When you are happy with your character, copy the image into a document and put the link next to it. This is just for organization so we don't have to scroll through endless threads to find the right image and link later.

3 Next, we want to create a close-up of the face. Take a screenshot of the character's face and copy it into Discord using the link as the image prompt. I've also adjusted the prompt to generate a character's face, and increased the image weight to make the face as close to the one in the full body pose as possible. When you're done, copy the image and link into your character sheet.

4. Now let's create the environments for our characters. [02:11](#) It's usually best to create environments without people in them, because otherwise they might blend with our characters later. To keep the look of the environments consistent, I first created a main reference image and to create new views from it, I used the pan or zoom tool. You can also create entirely new shots in the same location by using a similar prompt structure and using the original image as the image prompt, with the image weighting set to 0.2 to match only the style, but not the composition.

5 Now the fun part is to create the final composition of your shot, by merging our environments and characters together! I use the prompt format:

[image of characters face] [image of environment] --ar 16:9

Or this one:

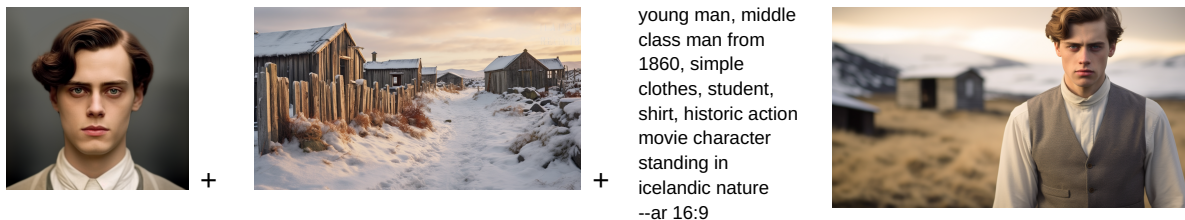
Mickmumpitz

“Generate entire movies with AI”
Exclusive Patreon Workflow Sheet #2 - v01

[image of characters face] [image of environment] original prompt to generate this character, additional description --ar 16:9

For example:

<https://s.mj.run/f6Jlmv7WwrQ> <https://s.mj.run/wyMhT8wU9gE> young man, middle class man from 1860, simple clothes, student, shirt, historic action movie character standing in icelandic nature --ar 16:9



And that usually works very well. The characters can still vary slightly, but I just repeat the prompt once or twice and there's usually an image that's perfect. Using the close-up of the character's face usually produces a close-up of the figure in a neutral pose. You can use the zoom out or pan function to change the framing of your shot.

But of course we can also direct our characters! We do that by slightly changing the prompt.

He looks so grim, let's make him happy!

Mickmumpitz

"Generate entire movies with AI"
Exclusive Patreon Workflow Sheet #2 - v01

[image of characters face] [image of environment] original prompt to generate this character, **happy, joyful**, --ar 16:9



...or scared!

[image of characters face] [image of environment] original prompt to generate this character, **scared, arms raised in the air, running, panic** --ar 16:9



But we are not limited to emotions! Let's make him play soccer! First let's look for a reference image of a person in a cool soccer pose. <https://www.posemaniacs.com/> is a great resource for this! Just look for a pose that you like, use the 3D viewer to rotate it and take a screenshot. Upload that screenshot to discord. Now let's build the prompt:

I'll start with a link to my character's T-pose, followed by a medium shot that I've already generated by blending together the environment and the face. I add the screenshot of the soccer pose as the third link, and another image of the environment as the fourth. I also add "**Playing soccer**" to the text part of the prompt.

[image: T-Pose] [image: medium shot] [image: pose] [image: setting]
young man playing soccer, middle class man from 1860, simple clothes, student, shirt, historic action movie character, icelandic nature --ar 16:9



[image: T-Pose]



[image: medium shot]



[image: pose]



[image: setting]

Mickmumpitz

“Generate entire movies with AI”
Exclusive Patreon Workflow Sheet #2 - v01



Now we have our character running around, but to see the full image I use the zoom out feature and there it is! By creatively blending images in this way, you can create consistent characters for pretty much any situation in your movie. **Just remember:** If your characters look too different, you can always combine it with the original close-up image of your character's face.



Step 2: Make our characters act! [02:58](#)

The main ingredient for this workflow, the one tool that made it all possible, is [runway gen-2's](#) new feature to generate video out of image prompts. So let's take this shot for example. I got it by blending these images.

Then I went to Runways Gen-2 and dragged my image into the image prompt window and hit generate. Now at this point an additional text prompt doesn't really work, so just leave it blank. I usually run the same prompt a few times because the motion it generates is pretty random. Sometimes the first try is amazing and matches the voice-over perfectly, sometimes it takes many tries. And sometimes the faces come out pretty broken, but that's okay, we can fix that later. Just focus on finding the right movement for your shot!

Also check out [Pika Labs](#), another AI video generator that is now in open beta! With this tool, the prompt doesn't destroy the image and can have a useful impact on the video, which is amazing. The quality is a bit worse than GEN-2, but you can try it for free on their Discord channel, similar to how you would use Midjourney!

Step 3: Make them talk! [05:17](#)

Once you've found your perfect shot, it's time to add the lip sync. I used [Elevenlabs](#) to generate my voices, but you can of course use any other tool or even your own voice! Import the voice file and generated video into Premiere, but you can also use another editing program like DaVinci resolve, which is free to use, and match the audio to your video. Export the audio at the same length as the clip as a wave or mp3 file.

Then go to the [Wav2Lip google Colab](#), a tool that generates really good lip movement by uploading a video and a voice audio file.

Click Copy to Drive and run the setup. This may take about a minute. Then scroll all the way down to "LipSync to your video file" and click the small play button in the second step. You are now ready to upload your video file. Take the next step and upload your audio file as an .mp3 or .wav file. Make sure it's just the voice, though. I accidentally uploaded the voice with the music underneath and that's what happened.

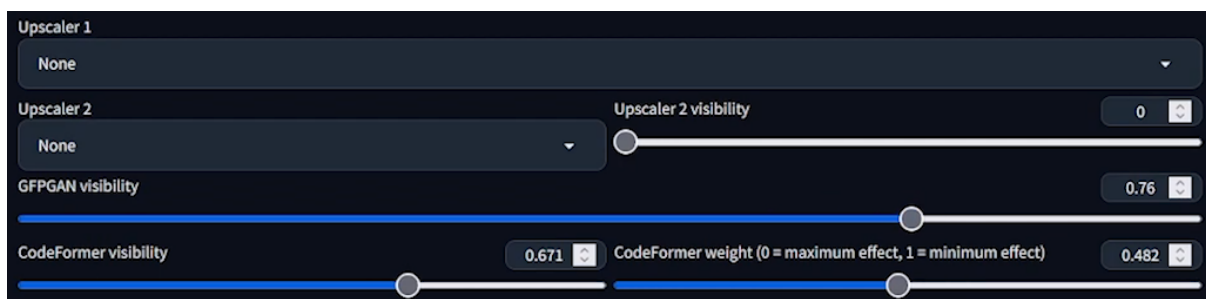
So once you've made sure you have the right clip and audio, start the last step and after another minute or so you'll have a result that looks like this: Amazing lip sync, but really poor quality and the face is still broken. So let's fix that!

Step 4: Fix the face! [06:37](#)

Import the original Gen-2 output and the Wav2Lip output into a video editor of your choice. Put the Wav2Lip file over the original video and create a rough mask around the mouth. Use some keyframes to make it move with the face, and blur it so no seam is visible. Export the video as a .png sequence.

Now go to the Extras tab in the Automatic 1111 Stable Diffusion WebUI and upload an image, preferably one from the middle of the sequence with the mouth open. Now we have two different tools that will be used to fix the broken AI faces. GFPGAN and CodeFormer.

They produce very different results, so play around with the settings and mix them until you find a look you like. I usually set the GFPGAN pretty high and use a little sprinkle of Code Former on top of it with reduced intensity.



When you find a look you like, go to the Batch tab and enter the path to your .png image sequence and an output folder. Click "Generate" and the whole sequence should render pretty quickly. This is the result:

Before



After



Now, if your result comes out flickery or you want to enhance your result even further, here are some quick tips and tricks!

Tips & Tricks [07:58](#) [+Additional Tricks]

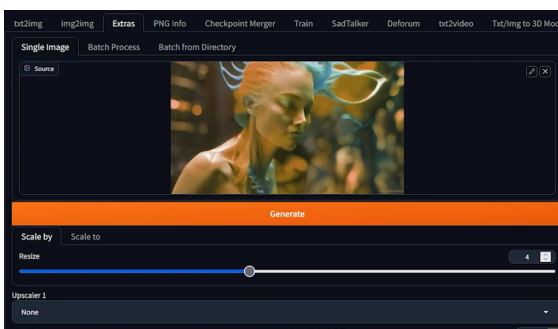
- 1 Limitations

There are some limitations with this technique. For example, it does not work with the side profiles of the characters, so keep the characters looking somewhat towards the camera. Also the faces should be as human-like as possible.

- 2 Hue Shift, Color Shift

Wav2Lip will not work if your character is too contrasty or too dark. If you still want to use this shot, increase the brightness before uploading it to Wav2Lip. You can bring it down again later.

The same applies to characters with non-human skin tones: Wav2Lip is really good at recognizing faces with non-human skin tones, but CodeFormer and GFPGAN won't work. But you can use a hue shift effect to make the skin more orange / human-like, before importing the sequence into stable diffusion to fix the faces. Later, you can use another hue shift effect with the opposite strength to restore the original colors.



- **3 Make the shot longer**

Gen-2 generates only 4 seconds of video. If your audio is longer, simply re-upload the last frame of your video into Gen-2 and stitch it together.

- **4 Prompt in runway**

I've said that using a runway prompt dramatically changes or even breaks the image, and while that's true, sometimes you can create really cool variations of shots. For example, I created this close-up of this old lady this way. You can use the free preview feature to check the image before generating it, and it's worth a try!

- **5 Reframe your shot**

Another way to reframe your shot is to use an image editor on your midjourney image. Use the zoom out feature or panning tool in midjourney to generate a wide shot of your scene and crop the image before you upload it to Gen-2.

- **6 Fix Face before Wav2Lip**

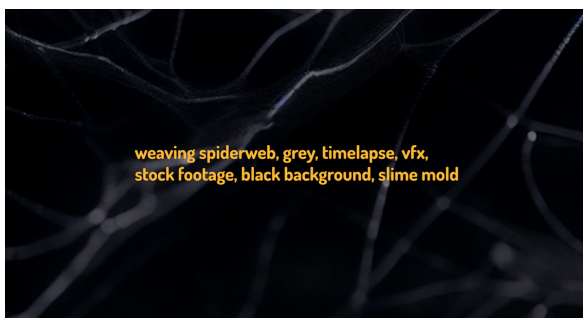
If your Gen-2 generated face is too broken, Wav2Lip may not be able to detect it. In this case, you can use the GFPGAN technique to correct the face before uploading the video to Wav2Lip and then it should work.

- **7 deflicker runway // wav2lip output**

Sometimes the clips created by gen -2 can flicker a little. The same can happen in the last step when we fix the face. Especially for characters with wrinkles or glasses or both, the GANs can get a little confused about where exactly to place the eyes. You can use DaVinci Resolves' Deflicker tool set to Fluoro Light or the RepairImage node to make the footage look much smoother.

- **Stock Footage vfx**

To bring your shot to the next level, you can also add stock footage. But not the normal kind of stock footage, NO! AI stock footage generated with Gen-2!
For this shot, for example, I wanted to add nasty abstract spider webs, so I used this prompt in Runway to generate them. With the black background, you can use the Add blending mode and play around with the opacity.
In this test shot, I used Gen-2 to create some flower-colored lens flares. I think something in the foreground really hides some of the imperfections and makes everything blend together.



weaving spiderweb, grey, timelapse, vfx,
stock footage, black background, slime mold



Mickmumpitz

“Generate entire movies with AI”
Exclusive Patreon Workflow Sheet #2 - v01

I hope you enjoyed this tutorial and are inspired to create your own AI movies with this workflow. If you do, please send me a link or tag me in your work. I want to create a little community playlist with all your projects soon.

And please let me know if this guide is helpful to you and give me feedback so I can improve it for the next workflow sheets! ❤️