Behind the Scenes: Directing AI Video Nexus - Dino Demolition Squad

Look at these images that accompany this guide:

The mindset: direct it, don't roll the dice

AI video stops feeling like a slot machine once you split the work the way a film crew does: someone finds the frame, someone stages the scene, someone gets the performance. I use Seedance for all three jobs, but never in the same pass. Each technique below fixes one problem at a time, so a good result in one area never gets thrown away while I chase another.

The common thread is to keep what works and regenerate only what doesn't. A good composition, a good camera angle and a good performance are three separate wins. I lock each one in as an image or a reference before asking the model for the next.

Technique 1: Paint it out, let Seedance fill it back in

When a generation has the right composition and camera angle but something wrong inside it, I keep the frame and repair only the wrong part. In one shot the soldiers were running the wrong way and were too big for the scene. Instead of regenerating the whole shot and losing the framing, I took that frame into an image editor and painted the offending soldiers out in solid white.

The edited frame becomes the start frame. The rest of the picture stays untouched: the light, the lens, the environment, the dinosaur. Then I give Seedance a reference image of the soldiers I want and describe in the prompt what they should do, for example sprinting left to right through the foreground, all the way past the camera. The model treats the white area as a gap and fills it with the new action.

The steps:

1. Pick the frame with the camera angle and composition you want. Frames that came from a generation are best, because the lighting and lens are already consistent. 2. Paint over the wrong elements in flat white. Cover the whole offending area, not just its outline, and let the white overlap the surroundings a little so no remnants show. 3. Attach the cleaned frame as the start frame and name it exactly that in the prompt. 4. Attach reference images for the elements that fill the gap (soldier lineups, creature sheets), each with its role stated: appearance only, never composition.

5. Describe the direction, speed and behavior of the new elements in plain physical terms.
6. Say explicitly that the white shapes are placeholder masks and are never rendered. Without that line, the white can show up in the video as white blobs.

Technique 2: Graybox in Maya to lock composition and camera

For shots where the staging or the camera move is complicated, I build a quick graybox in Maya first. It is nothing more than plain gray shapes standing in for the characters, creatures and landmarks, plus a camera. It takes minutes, and it answers the questions a prompt can't answer reliably: where everything stands relative to the camera, how big it is, and how the camera moves through the space.

There are two ways I use the graybox. For composition, I render a still of the graybox from the camera position I want, then use it as a reference for staging only, so the model gets the layout and scale without any of the gray look. For camera movement, I render the graybox move itself and give it to Seedance as a video reference that is explicitly for camera and staging, never for appearance. Motion carries over from a video reference much better than it does from text.

Tips that make the graybox work:

Keep the shapes simple and clearly different from each other, such as a tall box for the creature and small boxes for soldiers, so the model reads scale and

placement easily. Put a single character in a single flat color, and never rely on gray to say anything about the final look. State the final appearance through the character references. Say in the prompt what the graybox is for ("staging and camera only") and that its gray shapes are never rendered.

Keep the move simple. A handheld push-in or a slow pull-back reads clearly even in a rough blockout.

Technique 3:Blast the whole scene, then harvest the angles

For dialogue scenes, I don't try to get a perfect take in one pass. I write one long prompt, 20 to 30 seconds, that covers the whole scene with every angle: wide, over-the-shoulder, singles, close-ups. Then I run it repeatedly and judge only one thing, whether the angles and the shot order are right. I ignore the performances at this stage, because they are rarely right on the first run.

When I get a run with the right coverage, I cut it apart. From each shot in that run I grab one clean frame, and each frame becomes the start frame for its own generation. A single wide frame gives me the room and the blocking. A single close-up frame gives me the lens, the framing and the face.

The rest of the work is performance. I upscale each frame, then generate each shot individually from its frame until the acting lands. Because the angle and framing are fixed by the image, the only thing changing between attempts is the performance, so every retry is working on one problem.

Why this works:

Angles and acting are decided separately, so a failure in one never costs the other. Shots from different runs still match, because they all come from frames of the same coverage pass.

Each shot is short, and short generations hold a performance better than long ones. The frames stay consistent in lighting, wardrobe and set, so the cut together reads as one scene.

Technique 4: Direct like a director, and be specific

Seedance is remarkably good at understanding intent, especially in acting. The more specific the direction, the more precisely it follows. I write prompts the way I'd brief an actor and a camera operator: what the character wants, what they are feeling, what the camera is doing and why.

Three layers of direction can be stacked in one prompt:

> **Camera:** name the move and give it a physical cause. A handheld shot works better described as an operator walking forward on foot, bobbing about two or three times a second and bumped by passing soldiers, than as "very shaky." Cause and effect is easier for the model to follow than an adjective.
>
> **Blocking:** say who stands where, which way they face and which way they move, using screen left and right. Give timing in seconds, such as "first word at about 0.5s." **Emotion and performance:** describe the feeling and the physical signs of it, and write every instruction in positive terms. Say what the face and body do, not what they don't do.

Two habits help the model get acting right:

1. **Use a performance reference.** When a take has the exact movement I want, I feed that video back in as a reference for the shot and the voice, and describe only what continues or changes.
2. **Time the beats.** Give each beat a time window and say what finishes by when, such as "the breath runs out completely at about 7.0s." A beat with no end time tends to get cut off when the shot ends.

Case study: Fec and the explosions

The best example of how much nuance the model can pick up is the close-up of Fec, the gaunt imperial officer, watching explosions go off in the distance. The joke depends entirely on his reaction: he is never surprised or alarmed. He knows what is

happening and has already accepted that his life is over.

In the prompt I described his state of mind through small, physical behavior:

> His eyes search the distance, taking in all the ways his life is falling apart. He purses his lips and gives tiny nods to himself, as if accepting that nothing will ever be the same.
>
> Each explosion flash gets its own small reaction, a slow blink for the first two, then a single flat head tick on the third. The tick had to look deadpan, almost like counting. The shot ends on his breath: a slow inhale with his shoulders rising, then one long, tired sigh, like a man who is just over it.

The performance came from a handful of iterations. Early versions read as sarcastic or knowing, so I took the nod out of those words and described it as a single flat head movement with a completely blank face. That version got the laugh. It shows how precisely the model responds when the direction is concrete: tiny changes in wording changed the entire emotional read of the shot.

Quick tips and pitfalls

These are the habits that saved me the most time across the project.

Name every reference's role. For each image or video, say whether it is for appearance only or for composition and staging. Without this the model copies the wrong thing, such as a backdrop or a lineup layout.

Describe placeholders as placeholders. White paint-outs and gray blockout shapes must be named as masks that are never rendered.

Describe the wanted result, not the unwanted one. Positive, physical descriptions beat "don't" instructions, which the model often ignores.

Give motion a physical cause. An operator walking on foot and bumped by soldiers produces better handheld than the word "shaky."

Spell words normally. Stretched spellings such as extra vowels in a shouted word confuse the model. Describe the length instead: held out much longer than

the others. **Watch the reference count.** Seedance caps total references (the Dreamina FAQ lists 12 files), so merge or drop the least useful ones.

Keep the prompt tight. Shorter prompts with one clear job per shot outperformed long ones. For a single four-second shot, a few thousand characters is plenty.

Change one thing per iteration. Edit the latest working version and change only what failed, so you know what fixed it.