

Hero Frame and Video Generation

Generate a still of the influencer holding the product first, then hand that still to the video model as the start frame.

01 When to use this

Use this every time you generate AI video with a product in it. Going straight from text to video is where products melt, teleport, and change shape. The hero frame is the single biggest driver of whether the product survives the clip.

02 The prompt

STAGE 1, the hero frame.

Using `<PRODUCT_IMAGE_PATH>` and `<INFLUENCER_IMAGE_PATH>`, generate a single still image:

`<INFLUENCER_DESCRIPTION>` holding `<PRODUCT_NAME>` in `<PLAUSIBLE_SETTING>`, e.g. a bathroom for a face mask>.

Rules for the hero frame:

- The product must be clearly visible, correctly oriented, and readable
- Upload the product as a reusable element or prop so the model keeps it consistent rather than inventing it
- The setting must be somewhere this product is actually used
- Natural lighting, phone-camera framing, not studio

Show me the hero frame before you go any further. This one image decides the product fidelity of everything downstream.

STAGE 2, the video.

Once I approve the hero frame, generate video with:

- Start frame: the approved hero frame
- Script: `<SCRIPT_TEXT>`
- Duration: `<8 or 10>` seconds, chosen by the script word count rule
- Aspect ratio: `<9:16>`, fixed
- Sound: on
- Candidates: 3 per script

Concurrency: submit a maximum of 3 jobs at a time. The platform queues and caps around 8 concurrent, so firing everything at once fails. Wait for a batch to finish before submitting the next.

If a CLI or MCP connector exists for this platform, use that instead of a raw API key.

03 What it is actually doing

Text to video with a real product in it does not work reliably. The model has never seen your product and will invent something roughly product shaped, then change its mind halfway through the clip.

So you split it in two. First you generate a still image, the hero frame, of the influencer holding the product somewhere the product actually gets used. A face mask goes in a bathroom. You look at that one image and judge it carefully, because it decides the product fidelity of everything that follows.

Then you hand that still to the video model as the start frame. The model is no longer inventing the product. It is animating a picture that already has the product in it, correct and visible.

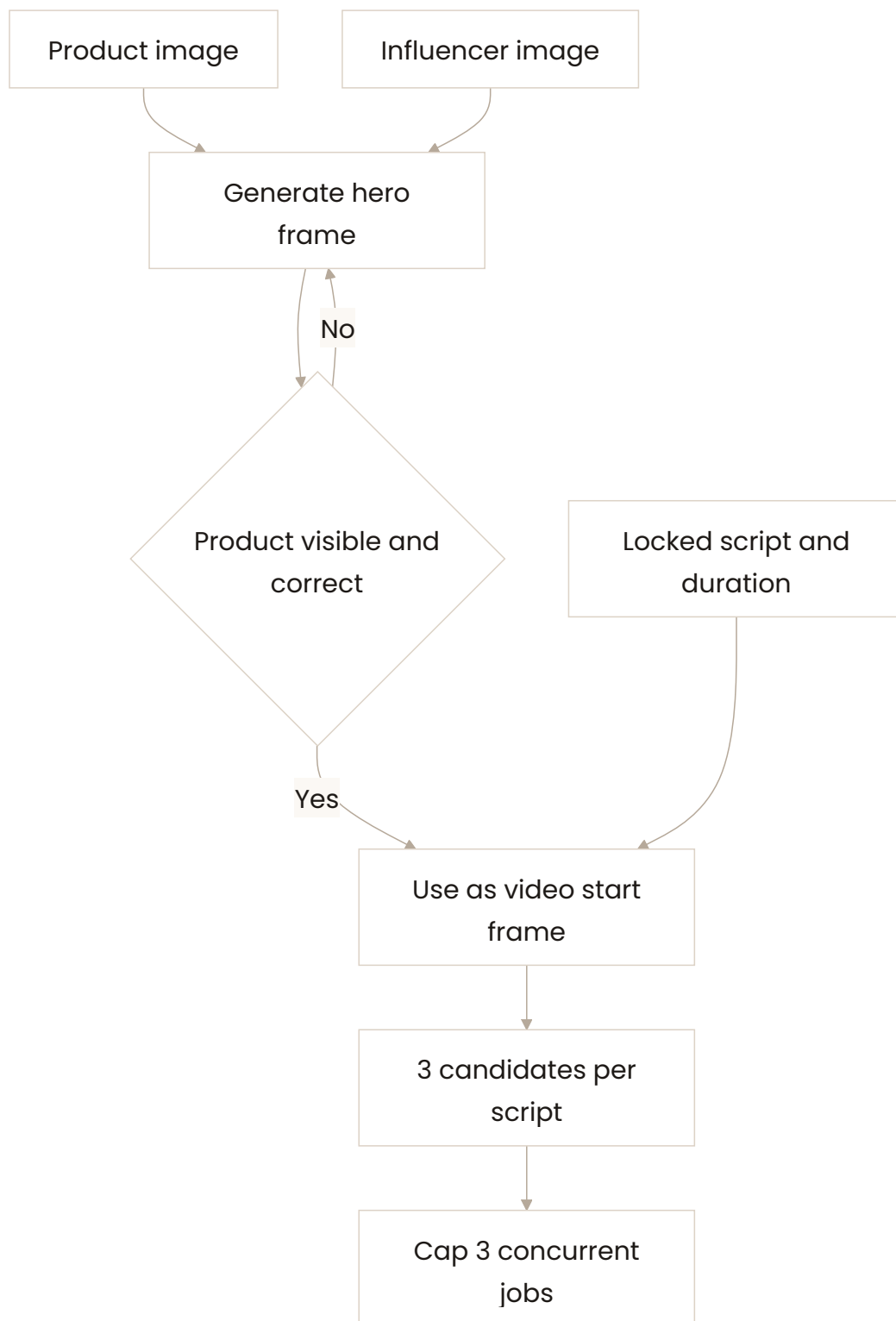
Nick also generated the influencer herself the same way. He took a reference image, regenerated it different enough to be a new person, and moved her indoors, because she was going to be used in a product shoot. Another useful move: upload the product as a reusable element or prop, so the platform holds onto it instead of guessing again each time.

Two practical settings. Fix your duration and aspect ratio, and turn sound on. Generate three candidates per script, because most generations are unusable and you are buying attempts.

And watch concurrency. These jobs queue. The platform Nick used capped around eight concurrent, so firing twelve at once failed outright. Cap your parallel submissions at three.

On access, prefer a CLI or an MCP connector over raw API keys when the vendor offers one. A login flow beats copying keys around, and there is no key rotation to manage later.

04 How it flows



05 Step by step

01 **Generate the influencer first if you need one**

Take a reference image, regenerate it different enough to be a new person, and move the setting indoors if she will be used in a product shoot.

02 **Upload the product as an element**

Register the product as a reusable element or prop on the platform. That keeps it consistent instead of reinvented on every generation.

03 **Pick a plausible setting**

Put the product where it is really used. A face mask belongs in a bathroom. A wrong setting reads as fake before anyone hears a word.

04 **Generate and judge the hero frame**

One still image. Check the product is visible, correctly oriented, and readable. Do not move on until this frame is right.

05 **Feed it as the start frame**

Hand the approved hero frame to the video model as the start frame. This is what stops the product from melting mid clip.

06 **Fix duration, ratio, and sound**

Set the duration from the script word count rule, lock the aspect ratio, and turn sound on. Leaving these loose creates variation you did not ask for.

07 **Generate three candidates per script**

Most generations are unusable. Three attempts per script is the working ratio, and it is cheap next to a real shoot.

08

Cap parallel jobs at three

Jobs queue and the platform caps around eight concurrent. Nick fired twelve and they failed. Submit three, wait, submit the next three.

06 Words explained

Hero frame

The still image you generate first, showing the influencer holding the product. It becomes the video's opening frame.

Start frame

The image the video model animates outward from. Give it one and it stops inventing the scene from scratch.

Product fidelity

How closely the product in the video matches the real product. Low fidelity means the label warps or the shape changes.

Element or prop

A product image saved on the platform for reuse, so every generation refers to the same object instead of guessing.

Concurrency cap

The most jobs a platform will run at once. Send more than the cap and the extras fail rather than waiting politely.

MCP connector

A ready-made bridge between Claude and an outside tool. You log in once and Claude can use the tool directly, with no API key to manage.

07 Watch out for

-
- Skipping the hero frame lets the product vanish, warp, or teleport mid clip. This is the most common video failure there is.
-
- Nick fired 12 jobs at once against a platform capped around 8 concurrent and they failed. Cap parallel submissions at 3.
-
- A setting the product would never be used in reads as fake instantly, no matter how good the generation is.
-
- Not registering the product as a reusable element means the model re-guesses the packaging on every single generation.
-
- Raw API keys are worse than a CLI or MCP login when one is offered, because you take on key handling and rotation for no benefit.
-

08 How it connects

UGC Script Generation

The script sets the duration you request here, and the hero frame supplies the visuals it will be spoken over.

Cross-Multiply Autonomous Video Spec

This is the unit of work the fleet repeats. Cross-multiplying influencers and products means generating one hero frame per pairing.

UGC Video Ad Pipeline

This note is steps one, two, and four of that chain, including the concurrency cap that broke Nick's first parallel run.

Compositing vs Direct Generation Models

Video is the far end of direct generation: high cost, low hit rate, and a ceiling nothing procedural can reach.

Credit. Framework and approach by Nick Saraev, from the hero frame and video generation segment of the UGC video build in his Zero to One course on Claude Code for marketing. Written up for the Kape vault by BrewedOps.