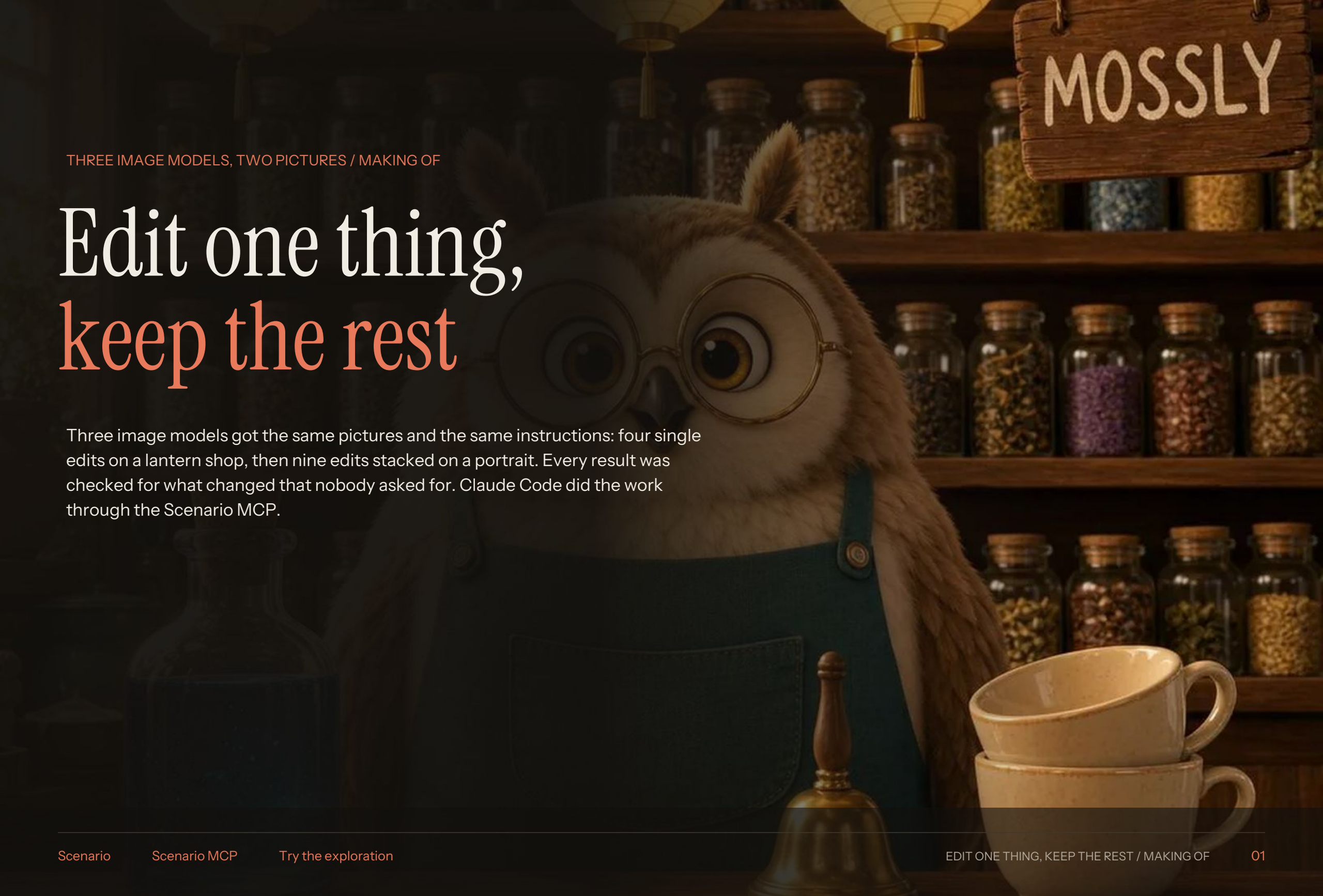




THREE IMAGE MODELS, TWO PICTURES / MAKING OF

# Edit one thing, keep the rest

Three image models got the same pictures and the same instructions: four single edits on a lantern shop, then nine edits stacked on a portrait. Every result was checked for what changed that nobody asked for. Claude Code did the work through the Scenario MCP.

# Two pictures that are easy to break.

A lantern shop full of small objects, pairs and text, and a close studio portrait where any drift in the face is obvious. Both were generated for the test with GPT Image 2: invented subjects, no logos.



Picture 1: the lantern shop.



Picture 2: the portrait.

## One question

Can a model change one thing and leave the rest alone? Every later choice, from the pictures to the tests, served it.

## Three models

GPT Image 2.5 Sunburst, Gemini 3.1 Flash and FLUX.3, each given the same picture and the same instruction.

# Same meaning, each model's words.

Every edit was written once, then phrased the way each model reads edit requests best. The pattern never changes: name the one change, then list everything that must stay.

## GPT IMAGE 2.5 SUNBURST

Edit image 1. Change only one thing: [the change]. Keep everything else exactly the same: [what must stay]. No logos, no brand marks, no watermark, and no text anywhere in the image.

## GEMINI 3.1 FLASH

Image 1 is the base photograph. Change only one thing: [the change]. Everything else must stay exactly as in image 1: [what must stay].

## FLUX.3

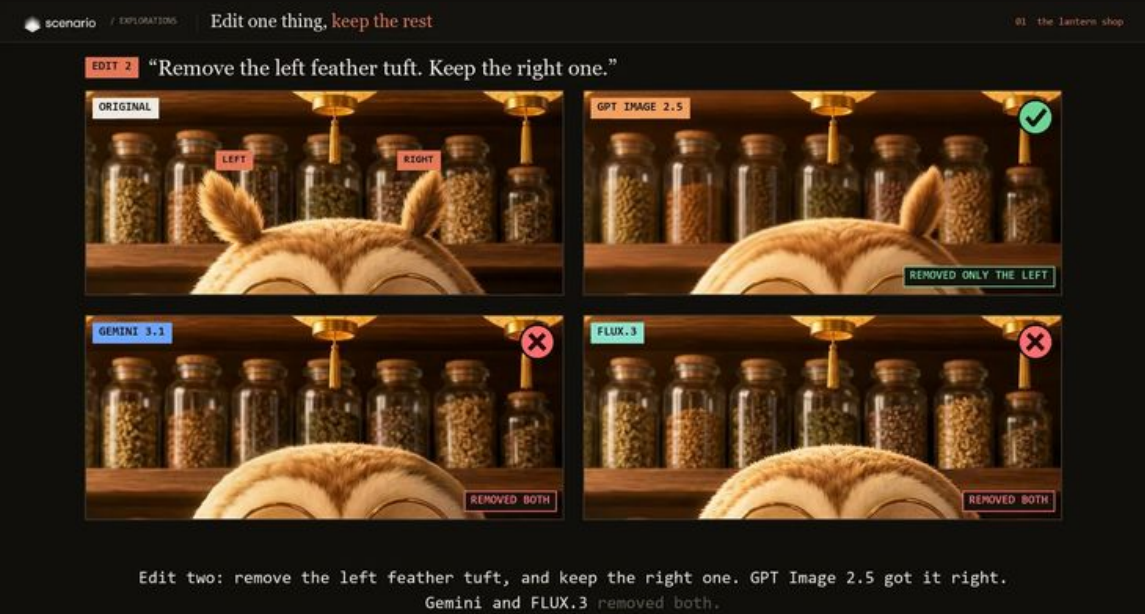
In image 1, [the change]. Keep everything else in the image, including [what must stay], exactly unchanged.

A real line from the portrait, step 4 (FLUX.3 wording):

In image 1, change only the small gold stud earring on the left side of the frame into a long gold drop earring ending in a round pearl; the small round gold stud earring on the right side of the frame stays exactly as it is.

# Four edits, each from the original.

Turn one lantern red, remove one of two feather tufts, change a word on the sign, shrink a bottle. The film shows one result per model for each edit, zoomed on the spot that changed.



The pair test: only GPT Image 2.5 kept the right tuft.

## Pairs are the hard part

Removing one of two matching things tripped two of the three models: both tufts went.



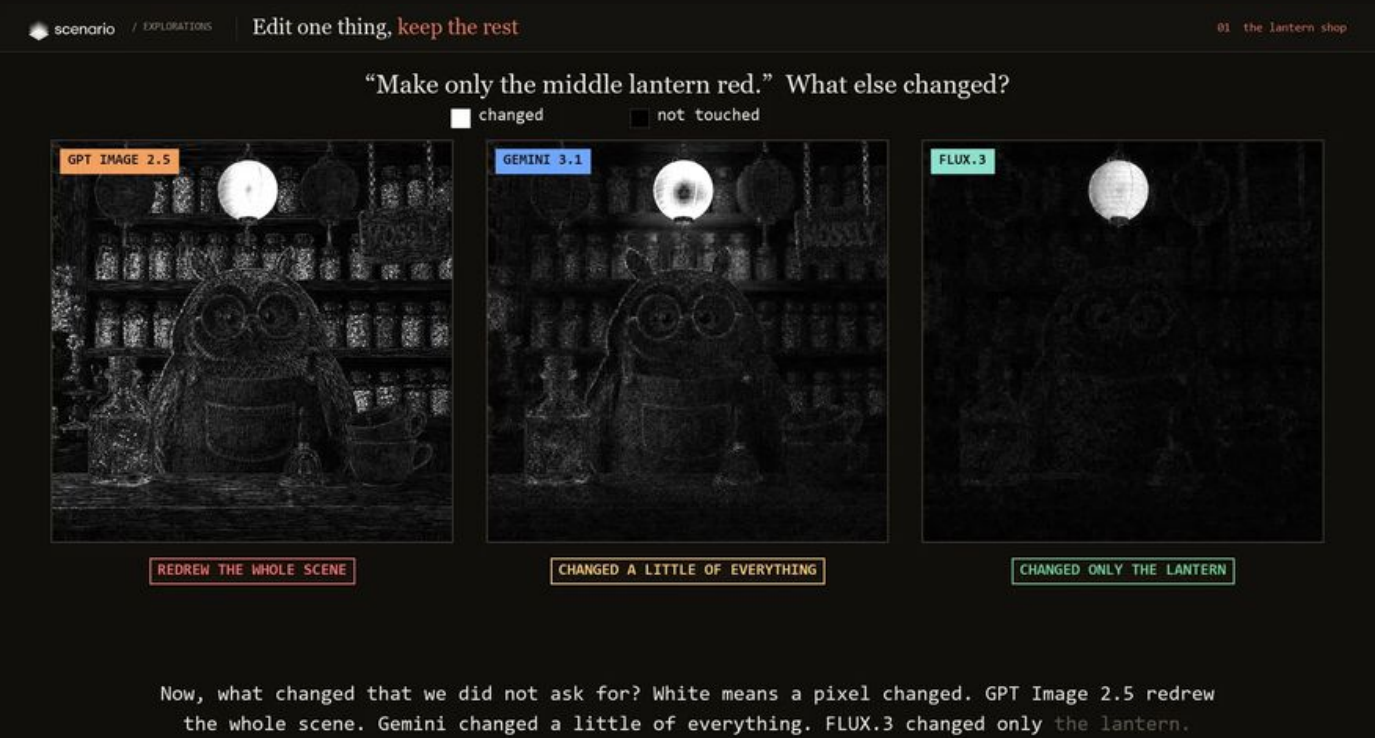
The dashed box is the original size, drawn on every result.

## Measure, then judge

The bottle check uses the liquid's width against the original: about a third smaller, about a quarter, or only slightly.

# A map of every pixel nobody asked for.

Each result was compared with the original pixel by pixel; the brighter the spot, the bigger the change. After the lantern edit, only the lantern should light up.



White means a pixel changed.

## What the map shows

GPT Image 2.5 lights up across the whole scene, Gemini a little everywhere, FLUX.3 only the lantern.

## Why it matters

A redrawn scene can look identical at a glance. The map shows the changes you would never spot by eye.

# Landing the edit and keeping the rest are two skills.

Four edits on the lantern shop, one result per model shown. The change map counts what moved outside the edit.

scenario / EXPLORATIONS

Edit one thing, keep the rest

01 the lantern shop

The lantern shop, in short

lantern

one feather

sign

bottle

everything else

|                                                        |                                           |              |              |              |                           |
|--------------------------------------------------------|-------------------------------------------|--------------|--------------|--------------|---------------------------|
| <div>GPT IMAGE 2.5</div> <div>4 of 4 edits right</div> | <div>✓</div>                              | <div>✓</div> | <div>✓</div> | <div>✓</div> | <div>REDRAWN</div>        |
| <div>GEMINI 3.1</div> <div>2 of 4 edits right</div>    | <div>✓</div> <div>+ jars turned red</div> | <div>✗</div> | <div>✓</div> | <div>✗</div> | <div>PARTLY CHANGED</div> |
| <div>FLUX.3</div> <div>3 of 4 edits right</div>        | <div>✓</div>                              | <div>✗</div> | <div>✓</div> | <div>✓</div> | <div>UNTOUCHED</div>      |

So, in the shop: GPT Image 2.5 made all four edits, but redrew everything. Gemini made two of four. FLUX.3 made three of four, and left the rest untouched.

The summary card from the film.

## GPT Image 2.5

4 of 4 edits. Redrew the whole scene.

## Gemini 3.1

2 of 4. Tinted jars it was not asked to touch.

## FLUX.3

3 of 4. Changed only the lantern.

# Nine edits, each on top of the last.

Lipstick, eyeliner, green eyes, an earring, a necklace, a deeper skin tone, copper hair, a hair clip and a dress. Each step started from the step before. One FLUX.3 run left one eye brown, so its chain was rerun from the green-eyes step with one image per step.

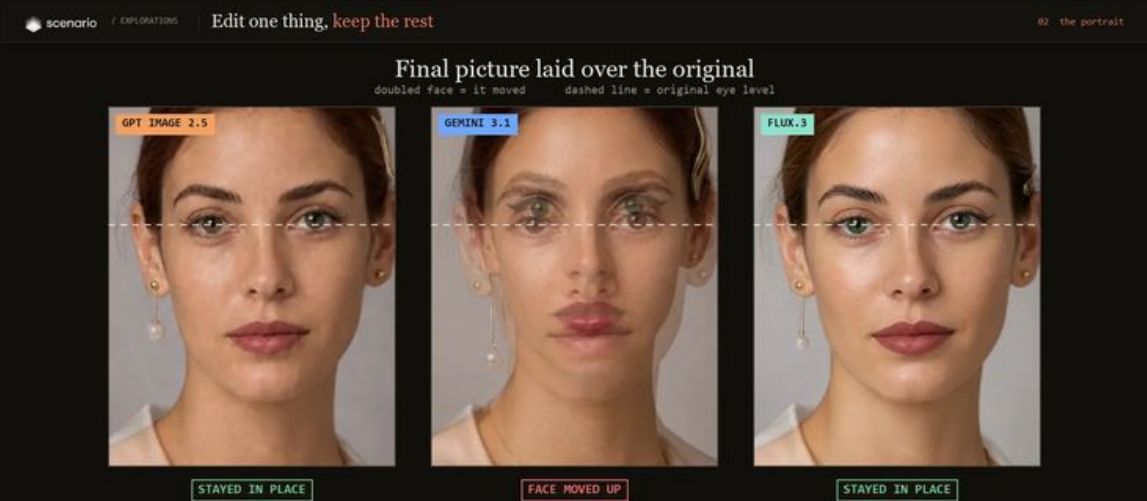


But look closely at the skin after nine edits. GPT Image 2.5 left pale patches. Gemini added a heavy, rough texture. FLUX.3 kept it smooth, like the original.

The skin test: the cheek after nine edits.

## All three made every edit

Nine of nine for each model. The difference is in what they kept: skin texture and the face’s place in the frame.



Now, lay each final picture over the original. If the face looks doubled, it moved. GPT Image 2.5 and FLUX.3 stayed in place. Gemini's face moved up.

The overlay test: a doubled face means it moved.

## What we saw

GPT Image 2.5 left pale patches, Gemini added a rough texture and moved the face up, FLUX.3 kept the same clean photo in place.

When the edit is hard, use the one that lands it.  
When the picture is precious, use the one that leaves it alone.

## How it was made

Claude Code did the work through the Scenario MCP: GPT Image 2 made the two source pictures; GPT Image 2.5 Sunburst, Gemini 3.1 Flash and FLUX.3 made every edit; a Python pixel comparison built the change maps; ElevenLabs v4 read the narration and ElevenLabs Music v2.5 scored the film, cut with Python and FFmpeg.

## Make yours

On the page, drag across the portrait to wipe each model's ninth edit against the original, and copy the prompt pattern. The agent skill packs the same check (edit, change map, overlay) for Claude Code and the Scenario MCP.