

THE MAKING OF

Fan Cam

The making of a crowd with twelve fans hidden in it

A broadcast camera at a night football match. One generated photograph of a floodlit stand holds hundreds of spectators; twelve were planted in it, one seat at a time. Hold the shot on a fan and the big screen cuts to them as they notice themselves and react. Scenario made the stand, the fans, their close-ups, their reactions and the crowd sound, all through Scenario MCP.

One photograph of a stand



The finished plate at 6336 × 2688 pixels: hundreds of generated spectators, twelve of them planted. On the page, the camera zooms in until one pixel of the plate fills one pixel of the screen.

1 · THE STAND

Every face distinct

The widest, largest photo the model could make, with no type anywhere



A detail of the plate at full size.

Asked for the widest and largest photoreal crowd, Scenario's model recommendation ranked Gemini 3.1 Flash first, as the only candidate with both native 4K and a true 21:9 ratio. One run returned two takes; the open stand under floodlights was kept.

> Night football match, live television. A long telephoto broadcast lens looks straight across the stadium at the opposite stand, so the packed seating tiers fill the whole 21:9 frame edge to edge: about twelve rows of spectators from bottom to top, several hundred people, each face and outfit clearly distinct. … Strictly no text anywhere: no letters, numbers, logos, signs, banners, flags with writing, sponsor boards or seat numbers.

At full size a face is about 60 pixels tall. A face detector found 422 faces in the take, which made choosing seats and placing crops precise.

2 · PLANTING

One seat at a time

The edit sees a 512-pixel crop of the stand, never the whole of it



The seat



The edit, at twice the size



What changed



Pasted back into the stand

The plate holds hundreds of people and is far larger than an edit's output, so each fan was planted on a crop centred just below the seat's face. Gemini 3.1 Flash edited it, the crop as its only reference, two variants per seat.

> Replace only the person seated in the exact centre of the image with a new spectator: an elderly East Asian woman in her late seventies … Change nothing else: every other spectator, seat and object stays exactly as it is, in the same place.

Only the change goes back: each edit is registered to its crop (all came back within 0.1 pixel), matched in colour on the unchanged surround and blended through a mask of what changed near the seat. Masks covered 12 to 37 percent of a crop; the rest of the stand never moved, and the plate rebuilds byte for byte.

2 · PLANTING

Twelve fans

Each with something that reads from across the pitch


The Superfan


The Knitter


The Nap


The Hoodie


War Paint


The Snack


Fifty Seasons


The Reader


The Dancer


The Manager


The Grump


The Twins

A striped foam top hat, a long striped scarf, a bobble hat on a sleeping head, a teal hoodie, face paint, a hot dog, a tweed flat cap, a green paperback, a bucket hat, a pointing arm, a tartan blanket, two identical puffer jackets. Three fans are pairs: a sleeping man and his wife, an elderly couple and the twins. Of the two variants, the second was kept for the couple, the reader and the twins.

3 · CLEAN-UP

Two fixes before the close-ups



Default threshold: an 8.6 percent mask, no coat, no arm



Threshold 20, every change kept, convex hull: 36.1 percent



Generated lettering



Edited to plain fabric

The manager wears a dark grey overcoat in a dark crowd, so it differs little from what it replaced: with the default threshold, the mask dropped his coat and his pointing arm. A threshold of 20, every changed region over 150 pixels and their convex hull brought both back.

OCR over the finished plate found strings such as “To Fruita” along a scarf. Three spots of generated lettering, on a hoodie, a sleeve and that scarf, were edited to plain fabric and pasted back the same way.

4 · CLOSE-UPS

The broadcast close-up

Re-shot the way a long lens across the pitch would see her



The knitter's close-up, 2752 × 1536 pixels.

Each planting edit holds its fan at twice the plate's resolution, so it became the reference for a second Gemini 3.1 Flash edit: a 16:9 close-up at 2K with the same person, clothes and props, the neighbours slightly out of focus.

> Create the live television fan-cam close-up of the elderly woman knitting in the centre of it, as a long telephoto broadcast camera would frame her: a 16:9 medium close-up from the chest up …

Hers came back with a floodlight behind her. From the opposite stand, a camera sees more rows behind a fan seated mid-stand, so most later prompts keep the background to spectators: “no sky and no lights in frame”.

4 · CLOSE-UPS

Three takes for the twins

A close-up has to stay in the same stand, at the same hour



Take 1: empty seats behind them



Take 2: packed, but daylight and a seat number



Take 3: night, every seat taken. Kept.

Asked for the close-up alone, the twins came back in front of empty seats. Asked for a packed stand, they came back in daylight with a legible seat number. The third prompt states night, floodlights, every seat taken and no numbers, with two variants; the first was kept. A close-up costs 20 credits; the twins' three runs cost 80.

5 · REACTIONS

Notice, then react

Eight seconds with sound, three beats, from the close-up



0.8 s: knitting



3.2 s: she sees herself on the big screen



6.4 s: the scarf goes up

Kling V3 I2V Pro animated each close-up into 8 seconds at 1928 × 1072, with its own sound: crowd ambience, laughter and cheering. Every prompt has three beats: what the fan does while oblivious, the moment they see themselves on the stadium's giant screen, and the reaction, with what the neighbours do.

> Live TV fan-cam shot at a night football match, long telephoto lens. The elderly woman in the lilac coat knits calmly, absorbed in her striped scarf. She glances up at the stadium's giant screen and sees herself: eyebrows lift, then a delighted grin. She proudly holds the long teal and saffron scarf up with both hands like a trophy and does a little shimmy in her seat while her neighbours laugh, point and cheer. One slow gentle push-in. Crowd ambience and cheering, no speech.

5 · REACTIONS

One continuous take

What the frame checks caught



Gemini Omni 1.1 Flash cut to a second camera angle



The reader's first take: a screen floating in the crowd



The retry, with the screen off-camera

Scenario's recommendation for human performance ranked Gemini Omni 1.1 Flash first and Kling V3 I2V Pro fourth, so both ran on the knitter. Omni (160 credits) played the arc but cut to a second angle; Kling (126 credits) held one shot, kept her identity and made all twelve.

Later prompts ask for "one continuous take" and put camera cut and scene change in the negative prompt. Every clip was checked two frames a second with OCR and by eye. The reader's first take drew a translucent giant screen behind her for about 1.5 seconds; the retry puts the screen behind the camera.

5 · REACTIONS

Twelve reactions


The Superfan


The Knitter


The Nap


The Hoodie


War Paint


The Snack


Fifty Seasons


The Reader


The Dancer


The Manager


The Grump


The Twins

Thirteen Kling V3 I2V Pro runs at 126 credits, one of them the rejected reader take, and the Gemini Omni comparison: 1,798 credits for the reactions.

6 · ASSEMBLY

Cuts, reel and sound

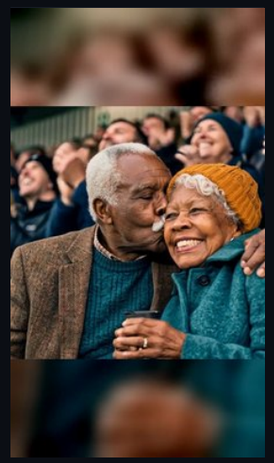
Two formats, one loudness, every graphic drawn in post



16:9



9:16 around one fan



A pair on a blurred fill



The reel's lower third, drawn with the page's fonts

Each clip is cut twice: 1280 × 720, and 720 × 1280 around a single fan. A 9:16 crop of a pair loses one of them, so pairs sit as a square on a blurred, darkened fill of the same frame. Every clip is levelled to the same loudness.

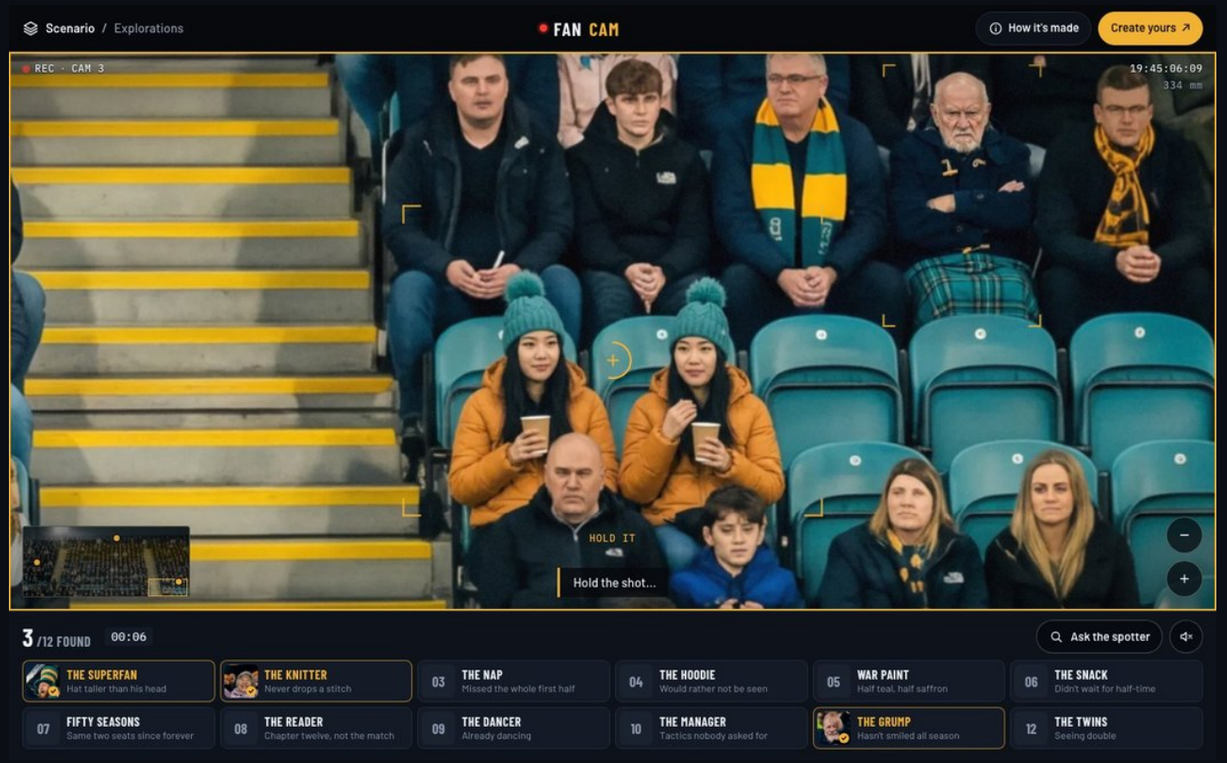
The reel strings three seconds of each reaction between two slates. Its lower thirds and REPLAY tally are drawn with Pillow in the page's fonts, so no type in it is generated.

Sonilo V1.1 Text to SFX made the crowd: a 40-second stadium bed, its last six seconds crossfaded into its start so it loops without a seam, and two cheers that play on every cut, for 21 credits.

7 · THE GAME

Hold the shot

The page as the visitor plays it



Holding the shot on the twins, three fans found



On a phone, in 9:16

The stand is a tile pyramid: a 1584-pixel overview first, then 512-pixel tiles at half and full size, drawn on a 2D canvas only while the camera moves.

The shot locks when a fan sits near the centre and fills 30 percent of the frame for 0.75 seconds. The focus box follows the fan, like a camera's face detection.

Then the camera glides in, the monitor flashes, a cheer plays and the clip takes over. Find all twelve to roll the reel or download it.

Make your own

The route, in six steps

1. Generate one very wide, very large crowd photograph with every face distinct and strictly no text.
2. Plant each fan with an image edit on a square crop around one seat, then paste back only what changed.
3. Scan the plate for generated lettering and edit it out the same way.
4. Re-shoot each fan as a 16:9 broadcast close-up from the planting edit, in the same packed stand at night.
5. Animate each close-up into a three-beat reaction with sound, in one continuous take, and check every clip frame by frame.
6. Cut 16:9 and 9:16 versions, draw the graphics in post, and hide the fans behind a camera that locks on when the shot is held.

The whole build used 2,526 credits on Scenario; the jobs behind what ships cost 2,200. Plant only yourself, people who agreed, or invented fans.

Play Fan Cam

The agent skill for this route (ZIP)

Scenario MCP

Scenario skills on GitHub