



SCENARIO EXPLORATIONS / MAKING OF

A Cloud Called Moss

An invented picture book: twelve spreads in six illustration styles, typeset in English, French and Spanish and read aloud in all three. Made by an AI agent through the Scenario MCP.

260

Scenario jobs

3,828.5

CU in all

72

pictures kept

6

styles

3

languages

[Open the exploration](#)

One book, three characters, three languages

The story is invented: Lina, seven, finds a lost little cloud called Moss in a dry hill village and walks him home to the sky while Grandma Oda waits for rain. Twelve double-page spreads, each with one quiet area kept free for the words. The cloud's name is localised for meaning and gender: Moss in English, Mousse in French, Pelusa in Spanish.



Lina: turnaround and four expressions



Moss, drawn from Lina's sheet



Grandma Oda, drawn from Lina's sheet

Sheets first

Every later picture takes these three sheets as references, so the raincoat, the three red buttons, the red boots and the cloud's face hold across 72 pictures.

Words never drawn

No prompt asks for text. The words are typeset in the page in each spread's quiet area, so a new language is a typesetting pass and not a new picture.

The same character prompt, three image models

The prompt asked for a yellow raincoat with three round red buttons and two patch pockets. The models were picked with the Scenario MCP's recommend tool and a newest-first search, then compared on the same sheet.



GPT Image 2.5 Sunburst: kept, 9 of 10

Why Sunburst

It kept the three buttons and the coat colour. Seedream drew four buttons and let the coat drift to yellow-green.



Seedream 5.0 Pro: runner-up, 7 of 10

One refusal

Seedream's first try at 3072 x 2048 was over its pixel cap and was refused; the retry ran at 2400 x 1600.



Gemini 3.1: 6.5 of 10

Cost of the step

6 calls, 108 CU, one of them refused.

Twelve spreads in watercolour and ink

One watercolour style tile set the look. Each spread was generated at 3072 x 1536 with the three sheets and the tile as references, then checked like a physics test: feet on the ground, hands really holding, depth order, nothing important on the centre fold.



The style tile

Continuity is checked

The first spread 12 broke continuity with spread 1 and its roofs filled the text area. The second take used spread 1 as a location reference and passed.



Spread 12, first take: a square roofed well



Kept: the round well with its iron arch from spread 1

Whole pictures, no patches

Where a spread had a fault (a shadow floating in the air, a line down the fold, a neck drawn badly, a head turned back too far, a boot pointing backwards), the picture was drawn again as a whole instead of patched.

Framing written in words

Spread 3 has three characters in a small room. The first take put Lina's head in the wall kept for the text. The framing was then written into the prompt: the camera standing well back, Lina about forty percent of the picture height, her head below the middle of the picture.



First take: Lina's head in the text wall



Blender 5.2 layout render, tried as a reference



Kept: framed by the prompt, wall clear

What the render did

Given as a reference, the Blender layout carried the room plan, but the model zoomed in about 15 percent, so it did not register to the pixel. It stayed a test.

Cost of the step

3 image calls, 54 CU, and one Blender render (Cycles, CPU) at 0 CU.

A layout-held edit is not a new medium

The first route redrew each key spread with its layout held ("keep everything else exactly"). The linocut read as a new book; gouache, pencil, cut paper and clay kept the watercolour underneath. The looser route drew a new tile per medium from the character sheets only, then drew each spread again with the lead as a layout sketch, or with the layout in words.



Gouache, first route: the watercolour survives



Gouache, looser route: kept



Clay and felt, looser route: kept

Clay shoot-out

On spread 7: Seedream 5.0 Pro kept (8.5, a photographed miniature), GPT Image 2.5 Sunburst runner-up (7.5, reads as a rendered illustration), Gemini 3.1 (7, a seam down the sky), Teletype V2 (5, identity lost).

One picture, not two

"Double-page spread" invited two separately painted halves with a seam. Prompts now open with "One single wide panoramic picture (2:1 landscape)".

Twelve spreads, six styles each

The nine other spreads were drawn in the five other styles by the same routes: the linocut redrawn over the lead with GPT Image 2.5 Sunburst, the four other media drawn again with Seedream 5.0 Pro. 45 new pictures, 72 in the book.



Spread 12 in the six styles: watercolour, gouache, linocut (top); pencil, cut paper, clay (bottom)

67 takes for 45 pictures

45 passed (862 CU). 22 failed a check and were drawn again (433 CU). All nine linocuts passed first time.

Checked against the story

Each take was read against its page's text and its medium at full size, not only for anatomy.

Why 22 takes were drawn again

Most faults were story faults the pictures could not know about. Each was fixed by correcting the words or the references and drawing the whole picture again.



Spread 1: Moss in the sky before he arrives

Story

Moss on spread 1 before he arrives (4 takes: his sheet came out of the references). The river too full on spread 5 while the text says the stones stood tall and dry (4).



Spread 8: three boots beside her

Props and physics

Boots on spread 8: two pairs, three boots, one boot, or eyes open while asleep (5). Rain against the wind on spread 9 (2). A well that read as an oven (1).



Spread 8 kept: asleep, one pair of boots

Medium

The picture drifting back to the watercolour or to soft gradients (5). A pencil page drawn as an open sketchbook with a fold (1).

The words are set, never painted

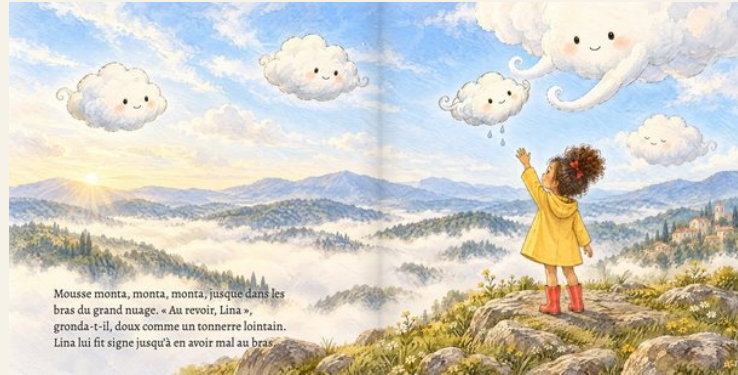
Each spread and style keeps a quiet area, written into the prompt (for example "The upper left is open sky."). The page sets the text there in Alegreya, with a pale panel where a picture is busy. French and Spanish run 15 to 25 percent longer than English and fit the same areas.



Spread 11, English

No Scenario call

Typesetting costs nothing: the same 72 pictures serve every language.



French: guillemets, Mousse



Spanish: angle quotes, Pelusa

Dialogue punctuation

Curly quotes in English, guillemets with spaces in French, angle quotes in Spanish, and no dashes.

Four voices, three languages, two ears

Four voices read the same page (spread 11) in the three languages. Each take was transcribed by Scenario Speech to Text (Whisper Large-v3) and scored against the book text, with pace, pauses and pitch range measured.



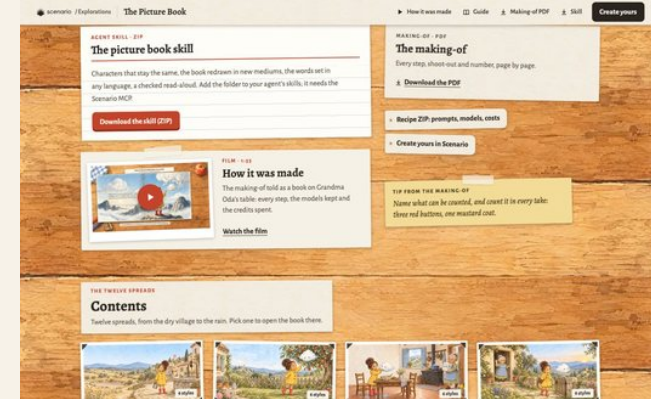
The page at 1440 px: spread 6, the style rail, the page tabs, the CU gauge, Read to me

The shoot-out

Gemini 3.8 Flash TTS (Sulafat) kept: 0 errors in 92 words, the widest pitch range, 2 CU a take. ElevenLabs 3 (Matilda) runner-up: 2 errors in French, 34 CU a take. Seed Audio 1.0 Multilingual: 7 errors. Minimax Speech 2.8 HD: 9 errors.

The check

36 spread takes and 3 titles, each language joined and transcribed by two independent transcribers (Speech to Text and Gemini 3.5 Transcribe). A word counts as wrong only where both disagree with the text. One French word in 1,110 was flagged (one transcriber wrote a homophone, the other heard another word); that take was voiced again as a precaution.



Take it home: the skill, the making-of, and the film card that opens the how-to

On the page

"Read to me" reads the title, then each spread, turning the pages and following the text language. Files trimmed and levelled to about -18 LUFS. Step cost: 69 calls, 373 CU.

The Rain Gauge: how it was made, in two minutes

The narrated how-to is a dummy book on Grandma Oda's table, one spread per step, with a garden rain gauge that fills with the book's CU as Moss rains into it. Every number on screen comes from the job ledger.



Step 3, the kitchen: what was asked, the references, kept and runner-up, what failed

Pictures

Table, rain gauge, Moss and blank paper by GPT Image 2.5
Sunburst with the watercolour tile as reference. All four passed at the first take.

Voice and sound

Narration by shoot-out: Minimax Speech 2.8 HD kept for its calm pace, Gemini 3.8 Flash TTS runner-up. Music by ElevenLabs Music Advanced v2.5 over MiniMax Music 3.0; sound by ElevenLabs Sound Effects 2 over Sonilo V1.1, on measurements.



The read-aloud step: the gauge at 3,454 CU for the book (3,453.5, rounded)

Cost of the film

24 jobs, 375 CU, no failure. Titles and clean versions, 1920 x 1080, 30 fps, mixed to -16 LUFS.

Every model, and how to make your own

260 Scenario jobs: the book 236 (232 successful at 3,386.5 CU, 4 failed at 67 CU) and the film 24 (375 CU), 3,828.5 CU in all. The recipe ZIP on the page holds every job with its prompt, parameters and CU.

Model	Calls	CU	Used for
Seedream 5.0 Pro	94	1,849.5	images, shoot-out
GPT Image 2.5 Sunburst	73	1,220	film, images, shoot-out
Speech to Text (Whisper Large-v3)	21	145	film, read-aloud, transcription check
ElevenLabs 3	4	136	film, read-aloud
Minimax Speech 2.8 HD	7	105	film, read-aloud
ElevenLabs Sound Effects 2	3	90	film
Gemini 3.8 Flash TTS	41	82	film, read-aloud
Gemini 3.1	3	60	shoot-out
ElevenLabs Music Advanced v2.5	1	60	film
Seed Audio 1.0 Multilingual	4	31	film, read-aloud
MiniMax Music 3.0	1	30	film
Gemini 3.5 Transcribe	4	12	read-aloud, transcription check
Telestyle V2 Style Transfer	1	5	shoot-out
Sonilo V1.1 Text to SFX	3	3	film

The skill

The picture book skill on the page gives an agent the workflow: sheets first, a tile per medium, spreads drawn again per style, words typeset, a read-aloud checked by two transcribers. It needs the Scenario MCP, which picks the models per job.

Try it

Turn the pages, switch any spread between the six styles and the three languages, compare a style with the watercolour lead, or press Read to me.

Open the exploration