

SCENARIO EXPLORATION · HOW IT WAS MADE

The Babel Tavern

A painted tavern that answers in your language

Six characters, eight languages and one small quest. Claude Opus 5.5 made every picture, voice, mouth and song through Scenario MCP in two headless sessions: a pilot that tested the hardest cases first, then the full cast.

Scenario MCP Scenario Skills Enter the tavern Agent skill (ZIP)

THE BRIEF

One prompt, then the full cast

The brief: a fantasy tavern that greets the visitor in their language, where each character holds a clue to a small quest. It reached Claude Opus 5.5 as a single prompt in a headless session with Scenario MCP. Nobody answered questions, so every choice of model, prompt and take was the agent's, and every clip had to pass a check before it counted.

1 **PILOT · 52 MINUTES**

The hardest cases first: a painted set as a moving hub, a raven's beak and a chest's lid as mouths, a pun rebuilt in French and Japanese, one dwarf in two regional Spanishes. 52 jobs, 3,589.5 CU.

2 **FULL CAST · 90 MINUTES**

The session resumed with its own conclusions: six characters in eight languages, six Spanish accents, a ballad sung in four languages, two Seedance 2.5 video finishes and a working page. 207 jobs, 11,128.75 CU.

3 **EVERY CALL THROUGH SCENARIO MCP**

252 generations reached the Scenario MCP server through the Scenario Blender Plugin's own client, with Blender itself never opened; the 7 language calls used the Scenario MCP tool asset_analyze. The ledger keeps each call's model, parameters and credits.



The review board of the full cast: every final clip in viewing order, with its models and a verdict (holds, holds with a workaround, flagged for a native ear).

THE SET

A painting that learns to move

GPT Image 2.5 Sunburst painted the empty tavern at 2560×1440: a carved bar, a stone hearth, a raven's perch and an iron-bound chest, lit by candles (25 CU, 53 s). The first pass hung an emblem on the back-wall banners; to keep every design original, an unmasked edit swapped them for plain tapestries and left the rest of the painting intact. Masked edits of the same banners had failed three times and were refunded.



The kept set, after the banner edit.



Layerize's nine outputs: the rebuilt base plate and eight cutouts.



A frame of the Seedance 2.5 living tavern.

Layers

Seedream 5.0 Pro Layerize split the painting into eight cutouts and a rebuilt base plate, with the floor and wall repainted where objects had been lifted, so a moving layer never reveals a hole (63 CU, 154 s). The cutouts came back without positions; the session placed them by template matching and built a 6-second 2.5D parallax move, plus a 1080×1920 pan for phones.

A living loop

Seedance 2.5 then re-rendered the parallax as a living scene at 1080p: the same camera and framing, a moving fire, flickering candles, with a crackling fire and room tone as sound (818 CU, 324 s). Runway Aleph 2 made a livelier version for review only; Seedance stayed closer to the painting. The loop plays behind the page's title.

THE CAST

Six characters, one voice each

GPT Image 2.5 Sunburst painted every character with the tavern as its reference image, so the cast shares one painted style. The two portraits added for the full cast were painted at 3:4, the lip-sync model's frame, after the pilot found that it cropped the 4:5 ones; the knight wears her hood pushed back so her mouth stays lit.



Brunna, dwarf barkeep



Grizzlewick, goblin merchant



Faelan, elf bard



Dame Wenna, hooded knight



Nib, the raven



Old Hasp, the mimic chest

A voice from a portrait

Seed Audio 1.0 derived each voice from its character's portrait (imageReference): a deep voice for the raven, a female voice for the dwarf barkeep, a mid, nasal one for the goblin. The first approved take became that character's lock.

The same voice in every language

Every other language used the lock as an audio reference (@Audio1). The same call can still drift, so each take went through a pitch gate: 10 of 40 missed it by more than 2.5 semitones and were redone, 5 of them with Seed Audio's pitch setting. Two stay flagged for a native listen: the goblin's Mandarin and the chest's Italian.

LIP SYNC

Mouths that are not mouths

The pilot's first question: can a beak and a chest's lid talk? The same raven portrait and audio went through three lip-sync models, and each verdict rests on a strip of frames at silent gaps against the loudest syllables.



The raven on three models: frames at three pauses (left) and four speech peaks (right).

With its lid open and a tongue out, Old Hasp read to both models as a face with a mouth inside: they animated the tongue and painted human lips between the wooden teeth. A Sunburst edit repainted him with the lid almost shut (21 CU), so the only way to talk is to lift it. H3 now moves the lid like a jaw in all eight languages. Two clips that drew fangs instead of wooden teeth were redone; the German one still sharpens its teeth on the widest openings.

P-Video Avatar · fails

The raven turns out of frame, the beak flattens and the bird changes mid-clip (34 CU).

Sync-3 Avatar · holds

Beak shut at 3 of 3 pauses, open at 3 of 4 speech peaks (380 CU, 330 s).

Minimax H3 Max Lip Sync · holds, kept

Shut at 3 of 3 pauses, open at 4 of 4 peaks, with the clearest hinge, at about half Sync-3's credits and 15 times faster (217 CU, 22 s). It made all 48 character clips and the 4 sung ones.



Open lid: fails on both models. Lid shut: the lid is the jaw.

LITERAL OR LOCALIZED

The joke is rewritten, not translated

Grizzlewick sells “Enchanted daggers, free of charge! Well, the dagger is free. The charge is forty gold.” The joke lives in one English word. The Scenario LLM (Claude Opus 4.8, called through Scenario MCP) returned, for each language, a literal translation, a localized rewrite built on a native pun, back-translations and a note naming the pun.



FRENCH

L'enchantement, par contre, c'est quarante pièces… faut bien que ça m'enchante moi aussi !

enchanter: to cast a spell, and to delight

GERMAN

Der Zauber kostet vierzig Gold. Sonst war alles umsonst.

umsonst: free of charge, and in vain

JAPANESE

まあ、タダの短剣じゃないってことさ！

tada: free of charge, and ordinary

PORTUGUESE

A graça da adaga custa quarenta moedas de ouro.

graça: free, and the fun or magic of a thing

MANDARIN

法力无边，价也无边，四十金币。

wubian: boundless magic, and a boundless price

Seed Audio then spoke each localized line in the goblin's locked voice and H3 animated it. A dubbing model translates on its own (ElevenLabs Dubbing v2, tested on the raven, wrote its own Japanese), so it cannot carry a rewrite: every adapted line went through the LLM first. One more call adapted the five plain lines into seven languages with every quest fact intact (2.75 CU).

One dwarf, six Spanishes, a key that sings



The pilot's comparison: Brunna's English source take, then Spain and Mexico.



Faelan sings the key.

Blue fire rising, moon on the stone,
three little secrets to carry you home.
Oak heart, iron heart, hear what I bring:
open for the one who has learned how to sing.

Six Spanishes

HeyGen Video Translate Precision turned Brunna's English take into six regional Spanishes, keeping her voice and re-syncing her lips (152 to 154 CU, 2.5 to 5 minutes each). The scripts are regional where it shows: Spain says sentaos (vosotros) and guiso, Mexico siéntense and estofado. The accents themselves need a native ear; no machine check here can confirm them.

A key that sings

The Scenario LLM wrote singable, rhymed versions of Faelan's verse in French, Spanish and Japanese. Minimax Music 3.0 sang each one in 44 to 53 seconds with a long intro; H3 takes at most 15 seconds of audio, so each clip sings the last couplet, the key that sings, and the page shows the whole verse. Sung lip sync held on all four.

CHECKS

Every take heard back

The agent could not listen, so it measured. Gemini 3.5 Transcribe heard back every spoken take and scored it against its script, every clip was checked frame by frame at its three loudest syllables and two silent gaps, and every clip was mastered to about -16 LUFS.

40 of the 42 new picks scored 0.87 to 1.00. The two below were read instead: the goblin's Japanese scored 0.57 only because the transcriber wrote homophones (tanken, the dagger, came back as an expedition), and the goblin's Mandarin stays flagged. The same reading caught salt heard as stone in the knight's Japanese, even in a take written in kana: most likely the transcriber, still marked for a native ear.

Transcripts

81 Gemini 3.5 Transcribe back-checks across both sessions, 2 CU and about 12 s each.

Frames

Mouth crops at each clip's three loudest syllables and two silent gaps, found from the audio.

Levels

59 of the 60 clips with sound sit between -15.9 and -16.3 LUFS, true peak at or below -1.5 dBTP.

HOW ONE LINE TRAVELS

The goblin's joke, from an English script to a French goblin telling a French pun, as measured in the pilot:

1 · SCRIPT	2 · ADAPT	3 · VOICE	4 · MASTER	5 · LIPS	6 · SUBTITLES	7 · BACK-CHECK
Written for the tavern	Scenario LLM	Seed Audio 1.0	ffmpeg, -16 LUFS	Minimax H3 Max	Seed Audio timing	Gemini 3.5 Transcribe
by hand	14.7 s	32.7 s	< 2 s	37.2 s	< 5 s	9.3 s
0 CU	2.75 CU	7 CU	0 CU	206 CU	0 CU	2 CU

217.75 CU and about 94 s of model time for one localized line; the lips are 95 % of the credits.

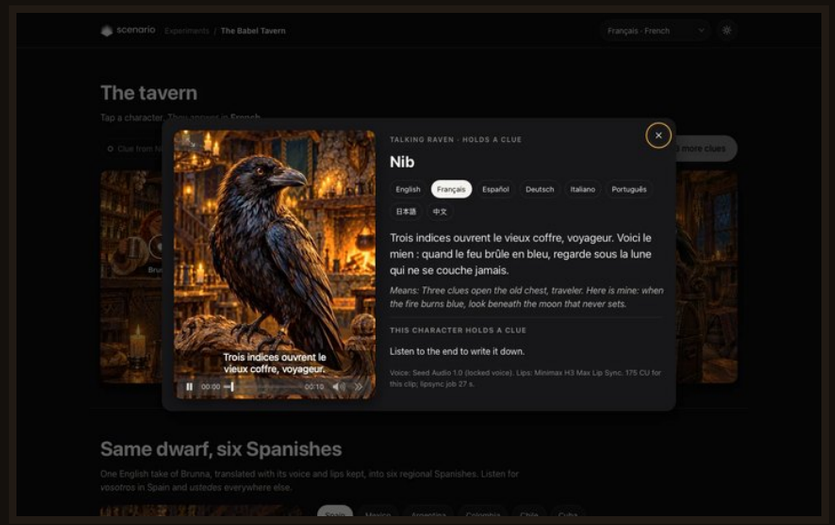
THE PAGE

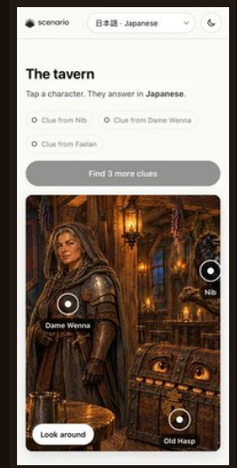
From clips to a quest

GPT Image 2.5 Sunburst painted the panorama at 3840×1280 with all six portraits as references, so each character stands in the room as painted (31 CU). The session built the page around it: the tavern sign greets the visitor in the browser's language, each pin opens a character who answers in the chosen language with subtitles, and three of them hold the clues.

Three true answers open the chest. Seedance 2.5 animated it from two frames, the lid-shut portrait and an open chest painted by Sunburst: the eyes blink and the lid swings open on its hinge, with a creak and a chime (232 CU, 185 s).

Before calling it done, the session drove the page in WebKit at 1440×900 and 390×844, light and dark, clicked through the whole quest and fixed what the screenshots showed: a see-through header, a pin label on the chest's face, a payoff that opened below the fold.







The session's own screenshots: a character sheet on desktop, the panorama pan on a phone.

The open chest Sunburst painted as the payoff's last frame.

THE NUMBERS

What it took

Both sessions, from the job ledger: 259 finished jobs and 14,718.25 CU in about 142 minutes of session time. Six more jobs failed and were refunded.

MODEL	USED FOR	JOBS	CU
Minimax H3 Max Lip Sync	Every mouth, spoken and sung	59	10,269
Seedance 2.5	Living tavern, opening chest	2	1,050
HeyGen Video Translate Precision	Six Spanishes	6	912
Seed Audio 1.0 Multilingual	Every voice, locks and re-rolls	77	408
GPT Image 2.5 Sunburst	Set, portraits, edits, panorama	15	334
Gemini 3.5 Transcribe	Back-checks	81	162
Minimax Music 3.0	Ballad songs	4	120
Seedream 5.0 Pro Layerize	Layers of the set	1	63
Scenario LLM (Claude Opus 4.8)	Adapted lines, puns, rhymes	7	19.25
Compared, not kept	Sync-3, P-Video, ElevenLabs Dubbing v2	4	865
Review versions only	Runway Aleph 2, MiniMax H3 Max, Gemini Omni	3	516
Total, both sessions		**259**	**14,718.25**



Still for a person: a native listen to the Japanese, the Mandarin and the six Spanish accents, the two flagged takes, and whether each voice suits its character, which was measured in pitch and words but never heard. “Babel” is said “babble”, as Seed Audio chose it.

Try the tavern, then make your own cast with the agent skill, Scenario MCP and Scenario Skills.

Download the agent skill (ZIP, 21 KB)