

Culture: how it works.

Follow one scene through the benchmark.

Context & discovery

Cultural Atlas
Indian tourism references

Culture & heritage

Wikipedia · UNESCO
Sahapedia · IGNCA

Visual & regional evidence

Wikidata · Wikimedia Commons · ASI
Census of India · Ministry of Tribal Affairs

ACTUAL DATASET EXAMPLE · IG_G_001139

Magh Bihu dawn near Nagaon, Assam: a bamboo-and-hay meji bonfire, a bhelaghar hut beside it, and women offering til pitha and coconut laru.

01 · Input evidence

Start with the named sources above. Link each cultural claim to its specific supporting reference.

02 · Structured prompt

Claude Code helps turn the scene into a prompt and required visual checks.

03 · Machine verification

Validation gates + Gemma 4 31B check the claims against their evidence.

04 · Human approval

Experts review cultural accuracy and translations before test-set release.

05 · Model evaluation

Generate images from approved prompts and check each required detail. Planned.

Human check, in this example

Are the fire, hut and offerings correctly identified as Magh Bihu? Does the translation preserve those details? Machine verification alone does not release the scene.

Culture: what we have.

The data behind the example.

237

authored scenes

500

target scenes

10

categories

34

states / UTs

143 verified* 55 need experts 39 not yet verified

*Machine-verified. All 237 current scene records remain drafts.

Pending: expert checks, 1,356 translation rows and 263 more scenes to reach 500. Five-language design: EN / HI / TA / TE / BN.

ONE RECORD IS MORE THAN A PROMPT

Magh Bihu · Assam



Bihu dancers · Assam · [Nayan j Nath](#) · CC BY-SA 4.0 · display cropped

Core checks	Meji fire · bhelaghar hut · offerings
Supporting detail	Harvested field · regional clothing
Current state	Machine-verified → human review

Regional clothing reference; this photo does not depict the Magh Bihu ritual or a scored output.

Culture: two kinds of expectation.

Explicit details. Implicit meaning.

ILLUSTRATIVE MAGH BIHU PROMPT

“Magh Bihu in Assam: villagers gather at dawn around a bamboo-and-hay meji bonfire.”

01 · EXPLICIT EXPECTATIONS

Did it show what we asked?

Details written directly in the prompt.

Example checks: villagers, dawn and the bamboo-and-hay meji bonfire.

VLM-assisted checks

A vision-language model inspects the image against a checklist. People verify mistakes and uncertain judgments.

02 · IMPLICIT EXPECTATIONS

Does the culture make sense?

Expectations inferred from the cultural context, even when unstated.

Example checks: does the ritual fit Magh Bihu? Are clothing and offerings appropriate to this Assamese scene?

Culturally knowledgeable people

Review meaning, regional context and stereotypes against evidence. Record valid local variations.

Paper's human rubric

0 no alignment · **0.5** partial · **1** complete

One alignment rating; below 1, flag **explicit** / **implicit** / **both** and explain why.

Our proposed workflow: VLM assistance + human review of both. CulturalFrames used humans for both; its automated metrics showed weak agreement with people. This is evaluation design, not a completed scoring run.