

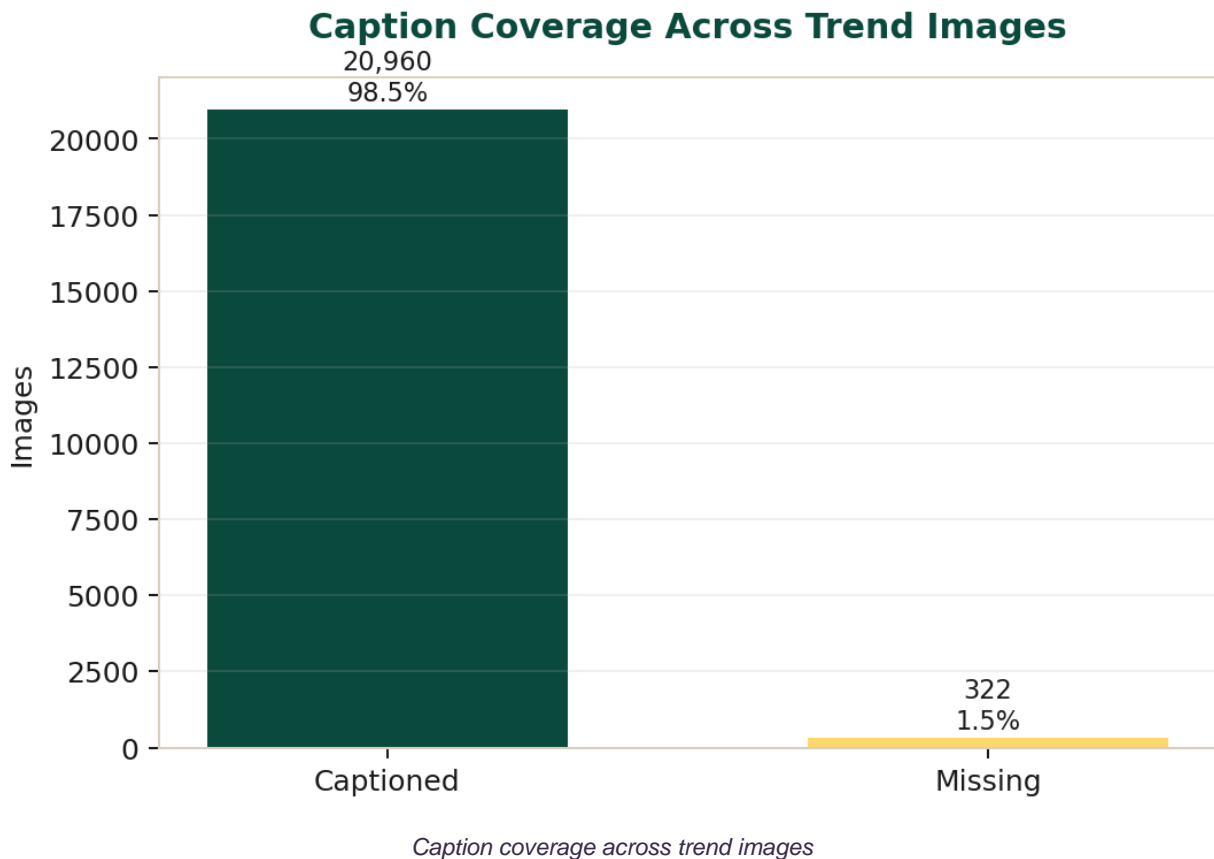
# Civitai Caption Analysis

## Xuthal Labs · Natural-Language Layer · April 2026

Observation window: 2026-02-25 to 2026-04-15 · 20,960 captions analyzed · charts generated from Postgres caption data

**Are you worried about the Civitai split?** We analyzed **20,960** natural-language trend captions across **21,282** tracked images to see how the front page is being described, what kinds of scenes dominate the corpus, and what the caption layer can tell us that the WD14 tags cannot.

## Xuthal Labs Brief



**Observation window:** 2026-02-25 to 2026-04-15

**Captioned images:** 20,960

**Missing captions:** 322

**Primary source:** `trend_images.caption` in Postgres

**Important caveat:** these captions are machine-written scene descriptions. They are useful for recurring language and scene structure, but they carry the habits and redundancies of the vision model that wrote them.

## What the Captions See in the Dark

The WD14 tag layer shows the bones of the front page. The caption layer shows the drapery over those bones — the weather, the light, the mood, the way the machine insists on narrating what it sees.

And the first thing it reveals is that the front page is not being described as flat output.

It is being described as **atmosphere**.

Across the caption corpus, the most common language is full of light, background, composition, softness, depth, color, expression, sky, glow, and the general sense that every image wishes to be a scene rather than a mere subject. The captions keep reaching for the same ideas:

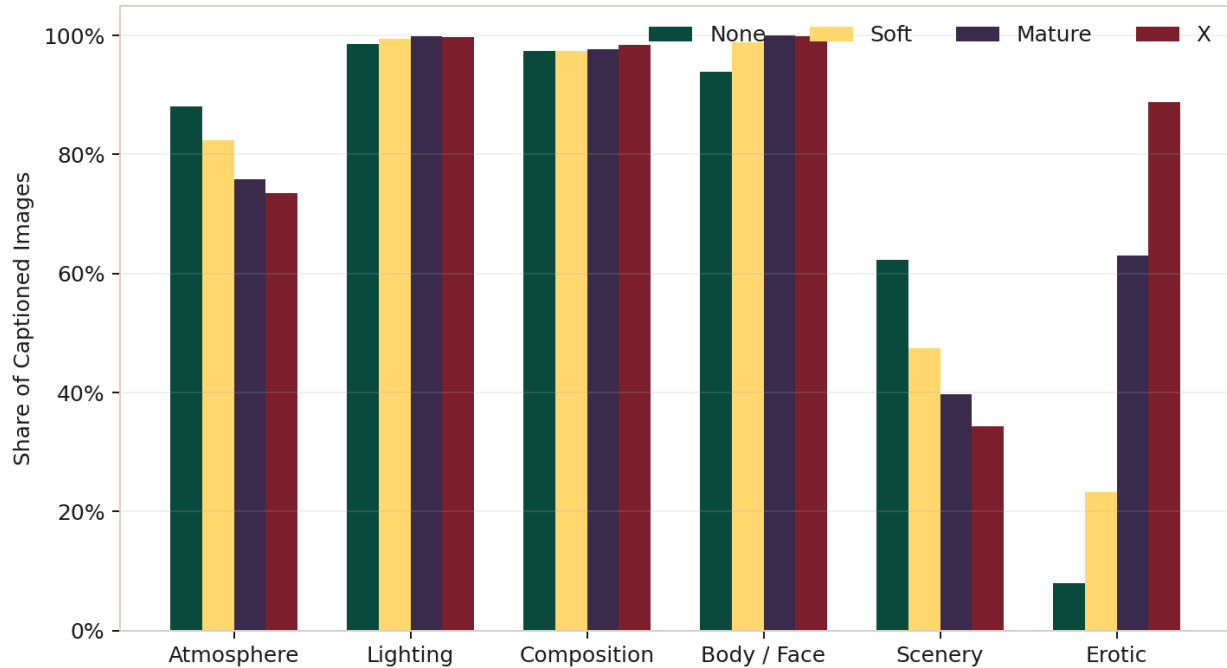
- dark backgrounds and luminous edges
- dramatic or soft lighting
- centered compositions and focal points
- expressive eyes and visible skin
- sky, water, landscape, and environmental context
- mood, atmosphere, and the promise of some private narrative just outside the frame

This is true even before we split the data by rating.

The machine sees the front page as a gallery of **posed intensities** — beauty under pressure, flesh under light, scenery around desire.

## The First Strong Finding: The Caption Layer Is Atmospheric Almost Everywhere

## Caption Theme Prevalence by Content Rating



*Caption theme prevalence by content rating*

The strongest caption themes by content level are remarkably consistent.

### None

Among safe-rated images with captions:

- 98.6% mention lighting or glow
- 97.4% mention background or composition
- 93.9% mention body or face details
- 88.0% mention atmosphere or mood
- 62.3% mention scenery or nature
- 57.0% mention fantasy or occult language

### X

Among explicit images with captions:

- 99.9% mention body or face details
- 99.8% mention lighting or glow
- 98.4% mention background or composition

- 88.8% mention erotic or intimate framing
- 73.6% mention atmosphere or mood

## Mature

Among `Mature` images with captions:

- 100.0% mention body or face details
- 99.8% mention lighting or glow
- 97.8% mention background or composition
- 75.9% mention atmosphere or mood
- 63.1% mention erotic or intimate framing

## Soft

Among `Soft` images with captions:

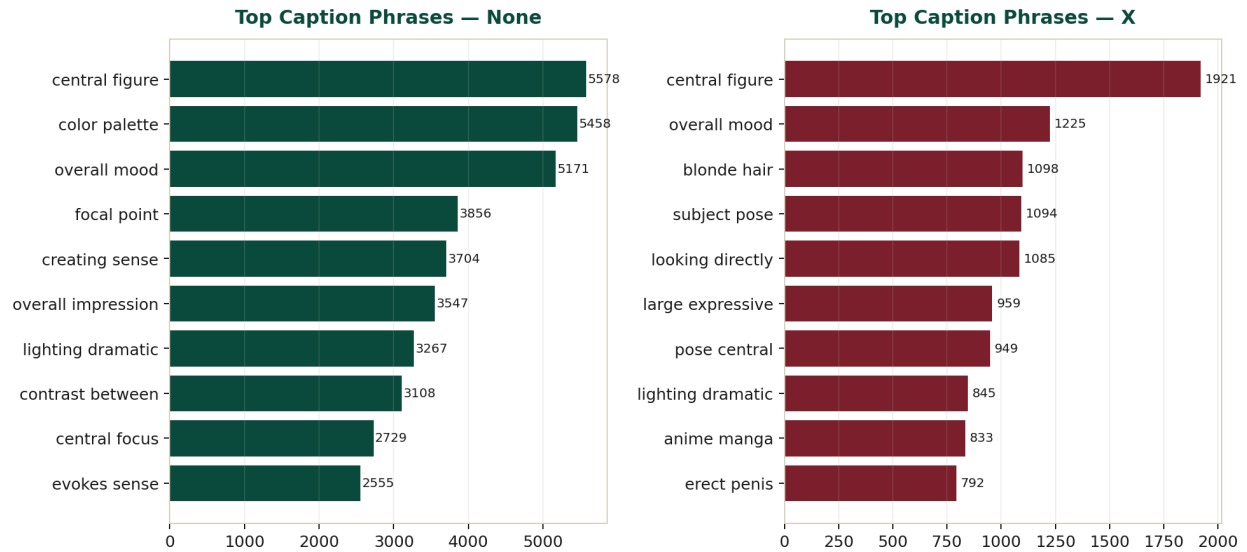
- 99.5% mention lighting or glow
- 98.9% mention body or face details
- 97.4% mention background or composition
- 82.3% mention atmosphere or mood

The important point is not merely that the erotic side is erotic. The important point is that **all four ratings are being described as staged environments under intentional light.**

This matches the older Howard principle exactly: beauty is stronger when it is placed in a chamber, a corridor, a ruined court, a night sky, a blue glow, a candlelit room. The caption model keeps trying to turn everything into a scene because the front page keeps rewarding scenes.

## The Second Strong Finding: The Caption Model Has a House Style

## Recurring Natural-Language Motifs in the Caption Layer



*Recurring natural-language motifs in the caption layer*

The caption layer is useful, but it is not neutral.

Across safe images, the most common caption bigrams include:

- central figure
- color palette
- overall mood
- focal point
- creating sense
- overall impression
- lighting dramatic
- contrast between
- central focus
- warm golden

Across explicit images, the most common caption bigrams include:

- central figure
- overall mood
- blonde hair
- subject pose
- looking directly

- large expressive
- lighting dramatic
- erect penis
- right hand
- legs spread

Across *Mature* and *Soft*, the same structural phrasing repeats:

- central figure
- overall mood
- subject pose
- looking directly
- anime manga
- setting background
- large expressive

That tells us something important: the caption model is not just recording the image. It is filtering the image through a **cinematic descriptive template**.

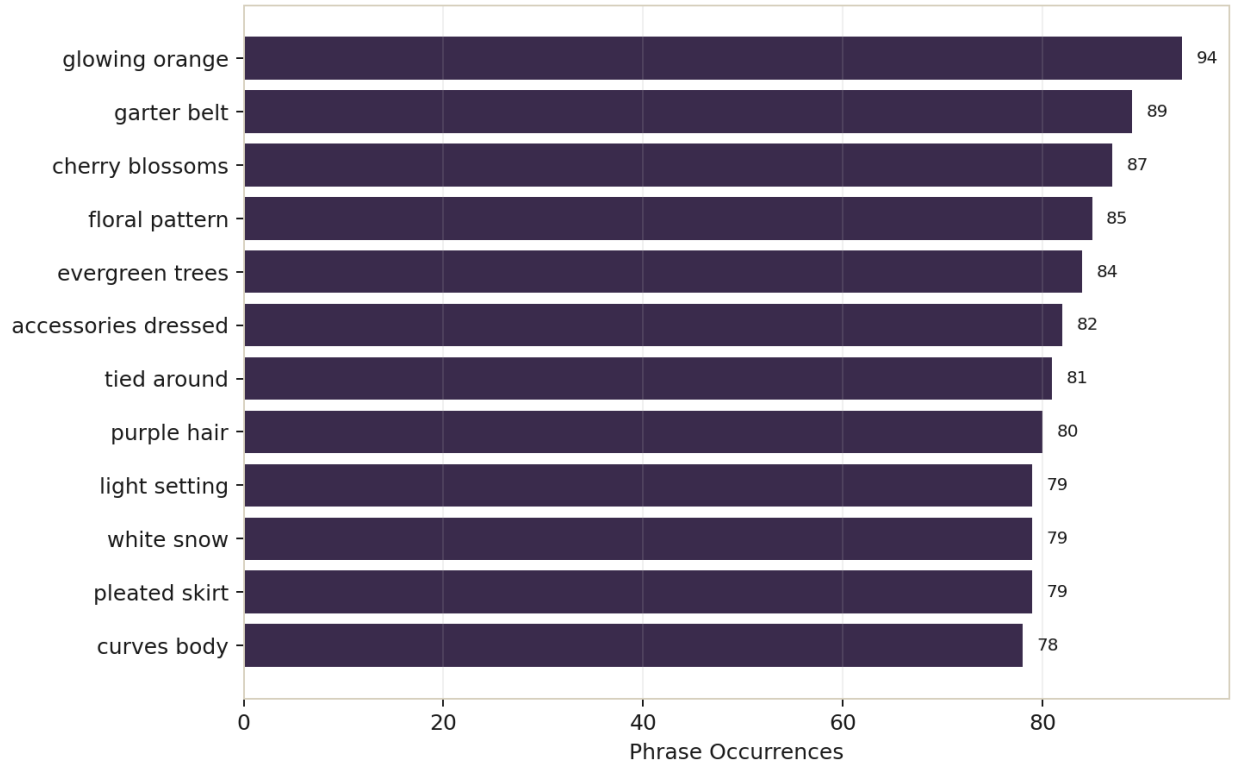
It loves:

- a declared focal point
- a named mood
- a visible palette
- dramatic or soft lighting
- a centered subject
- a contrast between foreground and background

This makes the caption layer extremely useful for understanding **how images are being framed**, but weaker as a literal record of what is uniquely in any single image.

In other words: the captions are excellent for scene language and poor for fine-grained uniqueness.

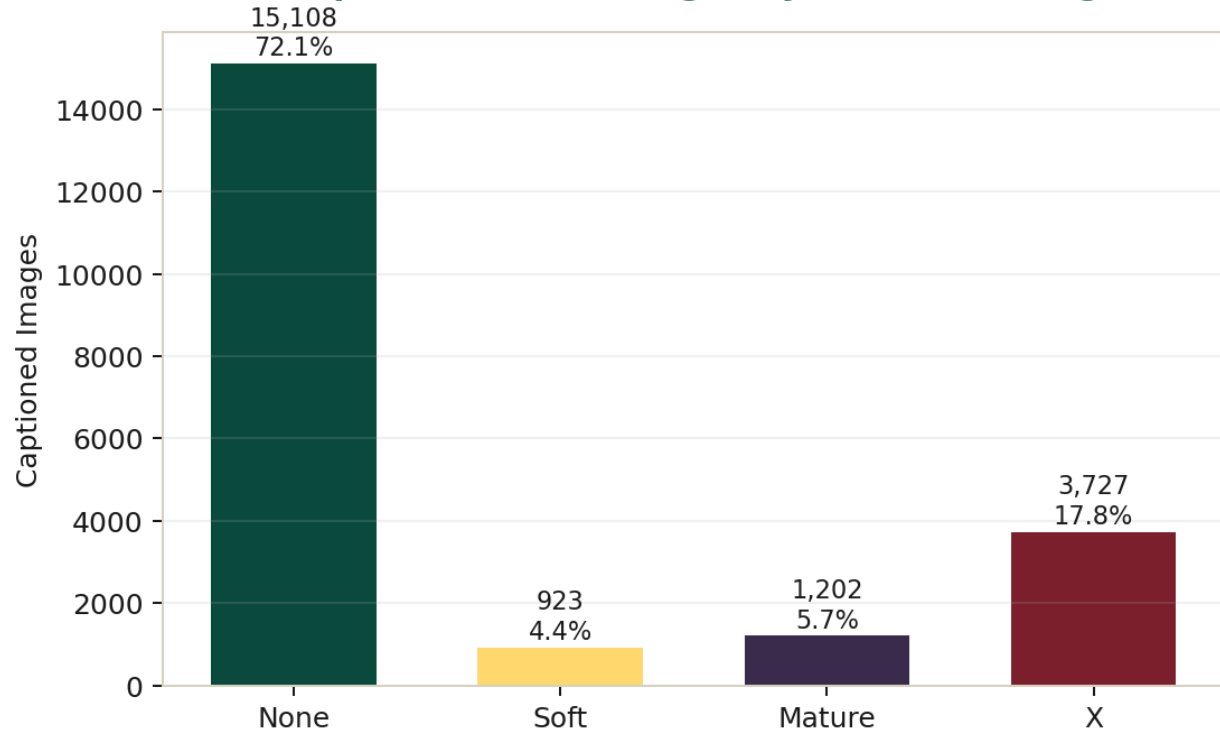
### Top Evolving Caption Phrases by Total Frequency



*Top evolving caption phrases by total frequency*

## What the Safe Side Sounds Like

## Captioned Trend Images by Content Rating



*Captioned trend images by content rating*

The safe side of the front page is described in a language of:

- sky
- glowing light
- color palette
- mood
- landscape
- focal point
- overall impression
- flowing hair
- dramatic lighting
- otherworldly scenery

This is not the language of catalog utility. It is the language of spectacle, atmosphere, and dreamlike placement.

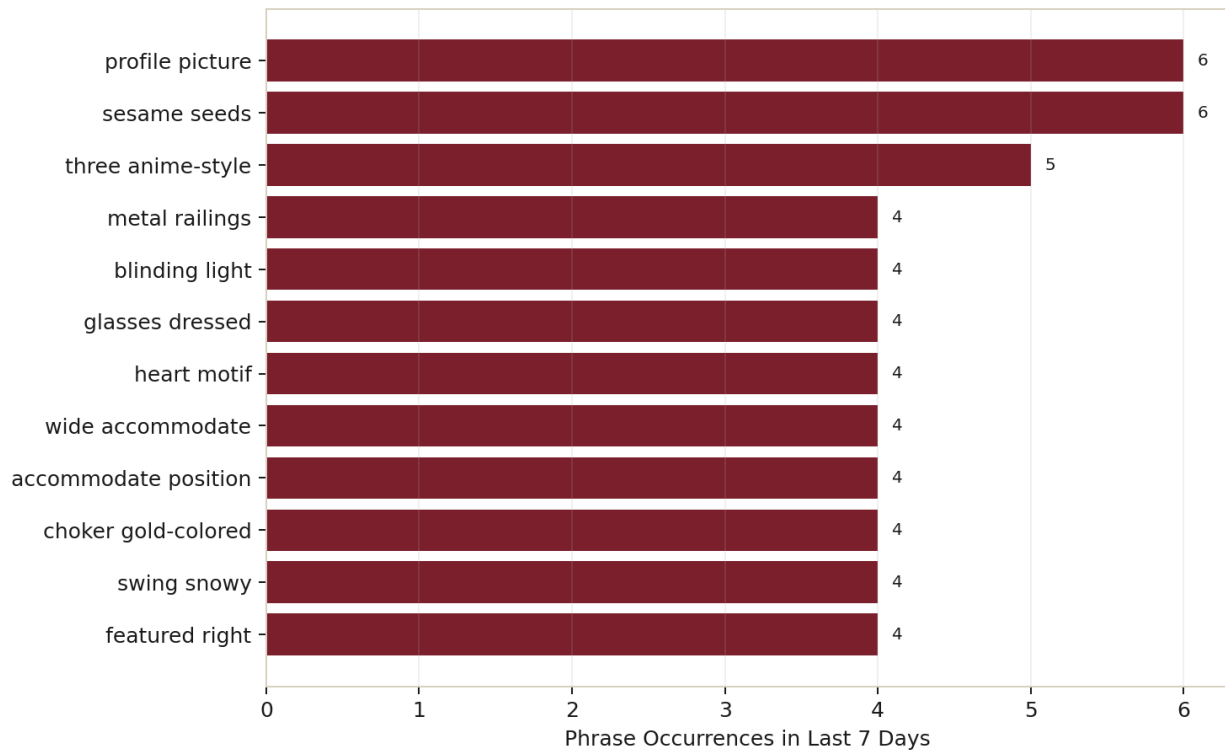
The safe front page, as the caption model sees it, is not merely a row of assets. It is a procession of luminous tableaux: neon landscapes, moonlit scenes, cosmic skies, rivers, ice, ruined architecture, and solitary figures arranged against those backdrops.

That matters for us because it strengthens the same argument the WD14 layer already made:

**the safe side is not plain. It is scenic.**

If `civital.com` is going to become the cleaner front door, it is likely to reward work that feels like an intentional world, not just work that happens to be non-explicit.

### Top Emerging Caption Phrases by Recent Frequency



*Top emerging caption phrases by recent frequency*

## What the Explicit Side Sounds Like

The explicit side is described differently, but not as differently as a shallow read might assume.

Yes, the erotic terms are much stronger there. The captions are far more likely to talk about:

- visible skin
- body positioning
- pose
- direct eye contact
- intimacy
- explicit anatomy

But the caption layer still leans hard on atmosphere:

- soft lighting
- dark backgrounds
- composition
- expression
- dramatic contrast
- cinematic framing

This is the useful distinction.

The explicit side is not winning because it is naked. It is winning because it is **naked under authored light**.

The language of the captions keeps reaching for softness, mood, dramatic composition, and expressive staging. Even the explicit front page is being narrated as a scene, not just as exposure.

That suggests the adult corridor after the split will continue rewarding not only niche specificity, but also **presentation quality**.

## What the Middle Looks Like in Prose

The *Mature* and *Soft* layers are the most painterly in language.

The caption model describes them with:

- upper-body compositions
- dramatic or soft light
- lingerie and attire details
- direct gaze
- long flowing hair
- anime or manga framing
- skin highlights and careful pose language

The middle ratings sound like the last velvet curtain before the door opens. They are less explicit than *x*, but they still borrow the same language of staging, seduction, and careful display.

That is one reason the middle may keep shrinking after the split. If the safe side rewards scenic clarity and the explicit side rewards specialized erotic clarity, the middle loses some of its reason to exist.

# What the Caption Layer Adds Beyond Tags

The tags told us that the platform is separating into two visual grammars.

The captions add something more delicate.

They show that **both grammars are cinematic**.

The safe side is cinematic through scenery, palette, and mood.

The erotic side is cinematic through pose, expression, lighting, and intimacy framing.

The caption layer also shows just how strong the platform's appetite is for:

- darkness punctured by glow
- soft or dramatic light against darker backgrounds
- centered or carefully framed subjects
- scenes that imply a world larger than the crop

That is deeply aligned with our brand. It is green-corridor logic. The machine keeps describing the front page as though it is full of chambers, thresholds, stained skies, glowing objects, and bodies made significant by where the light falls.

## What This Means for Xuthal Labs

### Safe corridor

The safe corridor should continue to benefit from:

- world morphs
- scenic fantasy
- cosmic or environmental concepts
- material transformations with strong lighting behavior
- posts and releases that emphasize atmosphere, palette, and place

The caption layer says that safe content succeeds when it feels like **a world you could walk into**.

### Adult corridor

The adult corridor should continue to benefit from:

- sharper pose and intimacy framing
- explicit work with strong composition and lighting
- niche-specific releases that still look authored rather than casual
- copy that emphasizes mood and staging, not only explicitness

The caption layer says erotic work succeeds when it feels **lit, placed, and intentional**.

## Messaging implication

When we write about our own releases, we should not flatten them into utility language.

The front page itself is being narrated in cinematic terms. Our marketing should meet it there:

- light behavior
- atmosphere
- silhouette
- staging
- scenery
- pressure between beauty and danger

In other words: write like the corridors matter, because they do.

## Limitations

The caption layer is useful, but it has obvious bias.

The vision model overuses a house style:

- central figure
- overall mood
- color palette
- focal point
- dramatic lighting

So the captions are best read as a **descriptive weather system**, not as perfect metadata.

They tell us how the corpus is being interpreted at scale. They do not tell us every unique fact inside every frame.

## Bottom Line

The caption layer confirms what the tags suggested, but with more silk on the bones.

Civital's front page is being described as a procession of staged intensities: scenic, luminous, atmospheric, body-aware, and almost compulsively cinematic.

The safe side sounds like sky, glow, palette, and world.

The explicit side sounds like skin, pose, expression, and intimacy under controlled light.

That means the split is unlikely to produce one beautiful side and one crude side.

It is more likely to produce **two different kinds of theater**.

And that is useful for us, because Xuthal has always known how to build a stage.