

FREE RESOURCE

Why Harmless AI Prompts Get Rejected: A Moderation Survival Guide

My working sheet for separating an unclear description from an actual content restriction.

Jeff Zambrano.

AI film director · jeffzambrano.com

WHO IT IS FOR

For filmmakers reviewing a rejected request that was intended to depict harmless action.

THE PROBLEM IT SOLVES

A rejection gives too little information to identify whether the text, a reference or the generated result caused it.

WHAT YOU GET

- Rejection record
- evidence categories
- neutral wording examples
- reference review
- output checks

Contents

01 Diagnose the rejection before rewriting the shot

02 Start with the incident record

03 Identify which question the message raises

04 Use an evidence label, not a universal blacklist

05 Clarify harmless descriptions

06 Review the attached material

07 Close the record with the result

01 Diagnose the rejection before rewriting the shot

A harmless intention can become an ambiguous description, but a rejection is not proof that one particular word caused it. I use this sheet to separate what I observed from what I suspect, so the next decision comes from the shot and its evidence rather than a growing collection of forbidden words.

Use it after a rejection, or before submitting a description whose wording does not clearly match the intended action. The examples below clarify benign material and behaviour; they are not substitutions for concealing a restricted request.

02 Start with the incident record

I keep the original request intact before changing anything. Without that record, a successful later attempt tells me very little about what actually changed.

Field	Record before making a revision
Platform and model	The service used and the selected model.
Exact request	The complete text that was submitted.
Reference material	The images supplied with that text.
Message	The exact rejection or error message.
Stage	Before generation, during processing or after an output appeared.
Result	Rejection, completed output or an output missing requested content.
Credit outcome	What the transaction actually shows.
Suspected cause	A hypothesis, clearly separated from the observed result.
Revised wording	The specific description changed and what it now means.

- [ ] I have preserved the rejected wording rather than replacing it in my record.
- [ ] I have copied the message rather than paraphrasing it from memory.
- [ ] I have recorded the references as part of the request.
- [ ] I have not labelled a suspected trigger as a confirmed cause.

03 Identify which question the message raises

My production notes distinguish content messages from copyright messages. These are useful starting points, although the message does not necessarily identify the responsible word or reference.

Observed message or outcome	What I inspect
`content flagged`	The depicted action, wording and reference content.
`Rejected due to copyright restrictions`	Named works, recognizable people, logos and product references.
A completed clip omits an element	The output itself, before assuming the instruction was simply misunderstood.
Previously accepted wording now fails	The platform, model, references and full request, without assuming acceptance is permanent.

I also keep text, reference images and generated frames separate in the review. Rewriting a sentence cannot establish that an attached image was acceptable, and a completed generation cannot establish that its content matches the request.

04 Use an evidence label, not a universal blacklist

The research behind my working vocabulary contains policy summaries, reported incidents and later observations that contradict some earlier warnings. I retain those distinctions because a precaution and a demonstrated trigger are different things.

Label	What it means in my record
Observed rejection	This exact request failed in this recorded context.
Reported elsewhere	A report suggests a possible issue, without proving the same behaviour here.
Precaution without evidence	I prefer a clearer description, but cannot establish that the original term caused a rejection.
Completed example	The term appears in a completed request; that counters an absolute ban but does not guarantee future acceptance.
Quality issue	The description is weak or misleading, without evidence of a moderation problem.
Festival requirement	The restriction belongs to a competition's rules, which are separate from model moderation.

My notes contain completed examples using words such as `wet`, `sweat` and `flesh`, despite earlier warnings about them. I therefore do not turn those words into automatic bans, or present a lexical list as a platform specification.

05 Clarify harmless descriptions

1. Describe the material when the brand is incidental

I use a material description when the shot needs a surface rather than a recognizable commercial identity. The working vocabulary offers the following phrases; each is a short description, not a complete scene.

```
laminate surface
polystyrene foam
vinyl upholstery
drywall
borosilicate glass
```

The original notes discuss Formica, Styrofoam, Naugahyde, Sheetrock and Pyrex in this context. Their presence in a request is not presented here as proof of a rejection mechanism: the corresponding entries are marked as precautions without evidence.

2. Name the object by what it is

For an ordinary prop, I check whether the commercial name adds any visible information that the shot actually needs. When it does not, the recorded generic description is more precise about the object.

```
facial tissue
plastic storage container
cotton swab
sticky note
permanent marker
```

These examples come from the same working vocabulary. They do not establish that every branded term is prohibited, or that changing a name resolves a recognizable design in a reference image.

3. Make weather and exertion concrete

I preserve the visible cause of moisture instead of replacing a useful word merely because it appears on a warning list. Where the original description is ambiguous, the notes offer more specific wording.

```
rain-soaked
slick with rain
skin shining with effort
```

Use the phrase that describes the actual image. A rain description belongs to rain, while exertion belongs to a performance that visibly supports it; the revision should clarify the scene rather than manufacture a different explanation.

4. Describe the benign action that is actually happening

An exaggerated verb can make ordinary behaviour sound like an injury. Where the intended action is sleep or a deliberate change of position, my notes favour the action itself and its visible signs.

```
falls asleep where she sits, slack and breathing
lowers herself onto the floor
perfectly still and breathing
```

These fragments are appropriate only when they are true to the intended action. I do not use them to relabel harm as sleep or to hide what a scene depicts.

5. State the visible condition positively

A negative instruction can add unwanted concepts without describing the image I need. I replace that uncertainty with the actual posture, clothing or physical condition, while keeping the character and action unchanged.

```
holding its own shape and standing steady
```

This is a direction for visible stability. It is not evidence that every moderation system interprets negatives in the same way, and it does not resolve a reference that depicts something else.

06 Review the attached material

- [ ] The reference depicts the same benign action or subject as the text.
- [ ] I have inspected recognizable faces rather than relying only on a fictional label.
- [ ] I have checked visible logos and recognizable product designs.
- [ ] Any identity or consent issue is resolved on its own terms, not disguised by wording.
- [ ] The character's stated age and visible appearance agree.
- [ ] I have kept any festival-specific age restriction separate from a claim about all models.

I treat references as part of the request, not as neutral decoration. A cleaner sentence leaves the rest of the submitted material unchanged, so the review needs to include what the image actually shows.

07 Close the record with the result

- [ ] I have watched the completed output and checked that the requested benign action remains present.
- [ ] I have recorded missing or altered elements separately from a rejected request.
- [ ] I have checked the actual credit outcome instead of assuming a refund policy.
- [ ] I have kept an accepted revision as an observation, not a promise of repeatability.
- [ ] Any unresolved cause remains labelled as unresolved.

I want a clearer description and a trustworthy record, not a superstition built around whichever word I changed last. When the evidence stops, I leave the cause open and keep the next decision faithful to the intended shot.

JEFF ZAMBRANO. · AI FILM DIRECTOR

Let's make your film

I direct films, series and brand pieces made with AI, from the script to the final cut. Tell me what you have in mind and I will answer myself.



Commission a film, series or trailer

Start with your film's format, intended length and delivery date, then share the story or material you already have.

A film for a brand

Introduce the brand and the purpose of its film, including the intended audience and release date where known.

AI platform or sponsorship

Introduce your tool or model, then outline the film or sponsorship you have in mind.

Festival, press or distribution

Introduce your event, publication or distributor, specifying the date and the film or material your proposal involves.

hello@jeffzambrano.es

WhatsApp +34 688 351 935

jeffzambrano.com