

무료 자료

무해한 AI 프롬프트가 거부되는 이유: 콘텐츠 심사 대응 가이드

불명확한 묘사와 실제 콘텐츠 제한을 구분하기 위해 사용하는 작업 시트.

Jeff Zambrano.

AI 영화감독 · jeffzambrano.com

이런 분께

무해한 행동을 묘사하려 했으나 거부된 요청을 검토하는 영화 제작자.

해결할 문제

거부 메시지만으로는 텍스트, 참조 자료, 생성 결과 중 무엇이 원인인지 파악하기 어렵다.

포함 내용

- 거부 기록
- 근거 분류
- 중립적인 표현 예시
- 참조 자료 검토
- 결과물 점검

목차

01 숏을 다시 쓰기 전에 거부 원인부터 진단하기

02 사례 기록부터 시작하기

03 메시지가 무엇을 확인하라고 하는지 파악하기

04 보편적인 금지어 목록 대신 근거 유형 표시하기

05 무해한 묘사를 명확하게 다듬기

06 첨부 자료 검토하기

07 결과를 기록하며 마무리하기

01 숏을 다시 쓰기 전에 거부 원인부터 진단하기

무해한 의도도 모호한 묘사로 옮겨질 수 있다. 그러나 요청이 거부됐다는 사실만으로 특정 단어가 원인이라고 단정할 수는 없다. 나는 이 시트를 사용해 관찰한 사실과 추측을 구분한다. 그래야 다음 결정을 금지어 목록의 길이가 아니라 숏과 그 근거에 따라 내릴 수 있다.

요청이 거부된 뒤, 또는 표현이 의도한 행동을 명확하게 전달하지 못하는 묘사를 제출하기 전에 사용하세요. 아래 예시는 무해한 소재와 행동을 명확히 설명하기 위한 것입니다. 제한 대상인 요청을 숨기기 위한 대체 표현이 아닙니다.

02 사례 기록부터 시작하기

나는 무엇이든 바꾸기 전에 원래 요청을 그대로 보존한다. 그 기록이 없으면 나중에 생성에 성공해도 실제로 무엇이 달라졌는지 알 수 있는 것이 거의 없다.

항목	수정하기 전에 기록할 내용
플랫폼과 모델	사용한 서비스와 선택한 모델.
정확한 요청	제출한 텍스트 전문.
참조 자료	해당 텍스트와 함께 제공한 이미지.
메시지	거부 또는 오류 메시지 원문.
발생 단계	생성 전, 처리 중, 결과물이 나타난 뒤 중 해당 단계.
결과	거부, 생성 완료, 또는 요청한 내용이 빠진 결과물.
크레딧 처리 결과	실제 거래 내역에 표시된 내용.
추정 원인	관찰한 결과와 명확히 구분한 가설.
수정한 표현	바꾼 구체적인 묘사와 그 표현이 이제 전달하는 의미.

- [ ] 기록에서 거부된 표현을 새 표현으로 대체하지 않고 보존했다.
- [ ] 기억에 의존해 바꾸어 쓰지 않고 메시지를 그대로 복사했다.
- [ ] 참조 자료도 요청의 일부로 기록했다.
- [ ] 의심되는 유발 요인을 확인된 원인으로 표시하지 않았다.

03 메시지가 무엇을 확인하라고 하는지 파악하기

내 제작 노트에서는 콘텐츠 관련 메시지와 저작권 관련 메시지를 구분한다. 이는 유용한 출발점이지만, 메시지가 반드시 문제의 단어나 참조 자료를 특정해 주는 것은 아니다.

관찰한 메시지 또는 결과	점검하는 내용
`content flagged`	묘사된 행동, 표현, 참조 자료의 내용.
`Rejected due to copyright restrictions`	언급된 작품, 알아볼 수 있는 인물, 로고와 제품 참조 자료.
생성된 클립에서 요소가 누락됨	지시를 단순히 오해했다고 가정하기 전에 결과물 자체를 확인한다.
이전에 허용된 표현이 이번에는 거부됨	허용 여부가 영구적이라고 가정하지 않고 플랫폼, 모델, 참조 자료와 요청 전체를 확인한다.

검토할 때는 텍스트, 참조 이미지, 생성된 프레임도 따로 살펴본다. 문장을 다시 썼다는 사실만으로 첨부 이미지에 문제가 없었다고 입증할 수 없으며, 생성이 완료됐다는 사실만으로 결과가 요청에 부합한다고 입증할 수도 없다.

04 보편적인 금지어 목록 대신 근거 유형 표시하기

내 작업용 어휘 자료의 바탕이 된 조사에는 정책 요약, 보고된 사례, 그리고 앞선 경고 일부와 상반되는 후속 관찰이 함께 담겨 있다. 예방 차원의 주의사항과 실제로 입증된 유발 요인은 다르므로, 나는 이 구분을 유지한다.

분류	내 기록에서의 의미
관찰된 거부	기록된 이 조건에서 이 요청이 정확히 거부됐다.
외부 보고	보고된 사례가 문제의 가능성을 시사하지만, 여기서도 같은 반응이 나타난다고 입증하지는 않는다.
근거 없는 예방 조치	더 명확한 묘사를 선호하지만, 원래 용어가 거부를 일으켰다고 입증할 수는 없다.
생성 완료 사례	해당 용어가 생성이 완료된 요청에 등장한다. 절대적인 금지라는 주장에 대한 반례이지만 향후 허용을 보장하지는 않는다.
품질 문제	콘텐츠 심사 문제라는 근거 없이, 묘사가 부실하거나 오해를 부른다.
영화제 요건	모델의 콘텐츠 심사와 별개인 공모전 규정에 따른 제한이다.

내 노트에는 앞서 경고 대상이었던 `wet`, `sweat`, `flesh` 같은 단어를 사용하고도 생성이 완료된 사례가 있다. 따라서 나는 이런 단어를 자동으로 금지하지 않으며, 어휘 목록을 플랫폼의 공식 사양처럼 제시하지도 않는다.

05 무해한 묘사를 명확하게 다듬기

1. 브랜드가 중요하지 않다면 재질 묘사하기

숏에 필요한 것이 알아볼 수 있는 상표가 아니라 표면의 특성이라면 나는 재질을 묘사한다. 작업용 어휘 자료에는 다음 표현이 있다. 각각은 짧은 묘사이며, 완성된 장면이 아니다.

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

원래 노트에서는 이 맥락에서 Formica, Styrofoam, Naugahyde, Sheetrock, Pyrex를 다룬다. 요청에 이 이름이 등장한다는 사실을 거부 메커니즘의 증거로 제시하는 것은 아니다. 해당 항목은 근거 없는 예방 조치로 분류돼 있다.

2. 사물의 실제 종류로 이름 붙이기

일반적인 소품이라면, 나는 상품명이 숏에 꼭 필요한 시각 정보를 더하는지 확인한다. 그렇지 않다면 기록된 일반 명칭이 그 사물을 더 정확히 설명한다.

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

이 예시 역시 같은 작업용 어휘 자료에서 가져왔다. 모든 브랜드 명칭이 금지된다는 근거도, 이름을 바꾸면 참조 이미지에서 알아볼 수 있는 디자인의 문제가 해결된다는 근거도 아니다.

3. 날씨와 신체적 노력 구체화하기

나는 유용한 단어가 경고 목록에 있다는 이유만으로 바꾸는 대신, 물기의 눈에 보이는 원인을 유지한다. 원래 묘사가 모호할 때는 노트의 더 구체적인 표현을 사용한다.

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

실제 이미지를 설명하는 표현을 사용하세요. 비에 관한 묘사는 비가 오는 상황에, 신체적 노력에 관한 묘사는 그 노력이 눈에 보이는 연기에 맞아야 합니다. 수정은 장면을 명확하게 해야 하며, 다른 원인을 만들어 내서는 안 됩니다.

4. 실제로 일어나는 무해한 행동 묘사하기

과장된 동사는 평범한 행동도 부상을 입는 상황처럼 들리게 할 수 있다. 의도한 행동이 잠들기나 의도적인 자세 변화라면, 내 노트에서는 그 행동 자체와 눈에 보이는 징후를 우선한다.

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

이 문구들은 의도한 행동에 충실할 때만 적절하다. 나는 위해 행위를 수면으로 바꿔 부르거나 장면이 묘사하는 내용을 숨기기 위해 사용하지 않는다.

5. 눈에 보이는 상태를 긍정문으로 명시하기

부정형 지시는 필요한 이미지를 설명하지 못하면서 원하지 않는 개념을 끌어들일 수 있다. 나는 인물과 행동은 그대로 유지한 채, 그 불확실성을 실제 자세, 의상, 신체 상태에 대한 묘사로 바꾼다.

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

이는 눈에 보이는 안정성을 지시하는 표현이다. 모든 콘텐츠 심사 시스템이 부정문을 같은 방식으로 해석한다는 증거가 아니며, 다른 내용을 묘사하는 참조 자료의 문제를 해결하지도 않는다.

06 첨부 자료 검토하기

- [ ] 참조 자료가 텍스트와 동일한 무해한 행동이나 대상을 묘사한다.
- [ ] 가상의 인물이라는 표기에만 의존하지 않고, 알아볼 수 있는 얼굴을 확인했다.
- [ ] 눈에 보이는 로고와 알아볼 수 있는 제품 디자인을 확인했다.
- [ ] 신원이나 동의에 관한 문제는 표현으로 감추지 않고 해당 문제 자체로 해결했다.
- [ ] 명시한 인물의 나이와 눈에 보이는 외모가 일치한다.
- [ ] 특정 영화제의 연령 제한을 모든 모델에 관한 주장과 구분했다.

나는 참조 자료를 중립적인 장식이 아니라 요청의 일부로 다룬다. 문장을 정리해도 함께 제출한 나머지 자료는 그대로 남으므로, 이미지에 실제로 무엇이 보이는지까지 검토해야 한다.

07 결과를 기록하며 마무리하기

- [ ] 완성된 결과물을 시청하고, 요청한 무해한 행동이 그대로 있는지 확인했다.
- [ ] 누락되거나 바뀐 요소를 요청 거부와 별개로 기록했다.
- [ ] 환불 정책을 추측하지 않고 실제 크레딧 처리 결과를 확인했다.
- [ ] 허용된 수정본을 재현성의 약속이 아니라 관찰 결과로 기록했다.

- [ ] 밝혀지지 않은 원인은 미해결로 표시해 두었다.

내가 원하는 것은 더 명확한 묘사와 믿을 수 있는 기록이지, 마지막으로 바꾼 단어를 둘러싼 미신이 아니다. 근거가 더 이상 없다면 원인을 단정하지 않고, 다음 결정도 의도한 숏에 충실하게 내린다.

JEFF ZAMBRANO. · AI 영화감독

당신의 영화를 함께 만듭시다

저는 각본부터 최종 편집까지 AI로 영화, 시리즈, 브랜드 작품을 연출합니다. 구상하신 내용을 알려 주시면 직접 답장드리겠습니다.



영화·시리즈·예고편 제작 의뢰
영화의 형식, 예정 길이와 납품 일정부터 알려주시고, 이미 준비한 이야기나 자료도 공유해주세요.

브랜드 영화
브랜드와 영화의 목적을 소개해주세요. 대상 관객과 공개 일이 정해졌다면 함께 알려주세요.

AI 플랫폼 또는 후원
도구나 모델을 소개한 뒤, 구상 중인 영화나 후원 내용을 알려주세요.

영화제·언론·배급
행사, 매체 또는 배급사를 소개하고, 일정과 제안에 관련된 영화나 자료를 명시해주세요.

hello@jeffzambrano.es
WhatsApp +34 688 351 935
jeffzambrano.com