AI Image Detector

AI 생성 이미지 판별기

경영학부
20251450 양지선

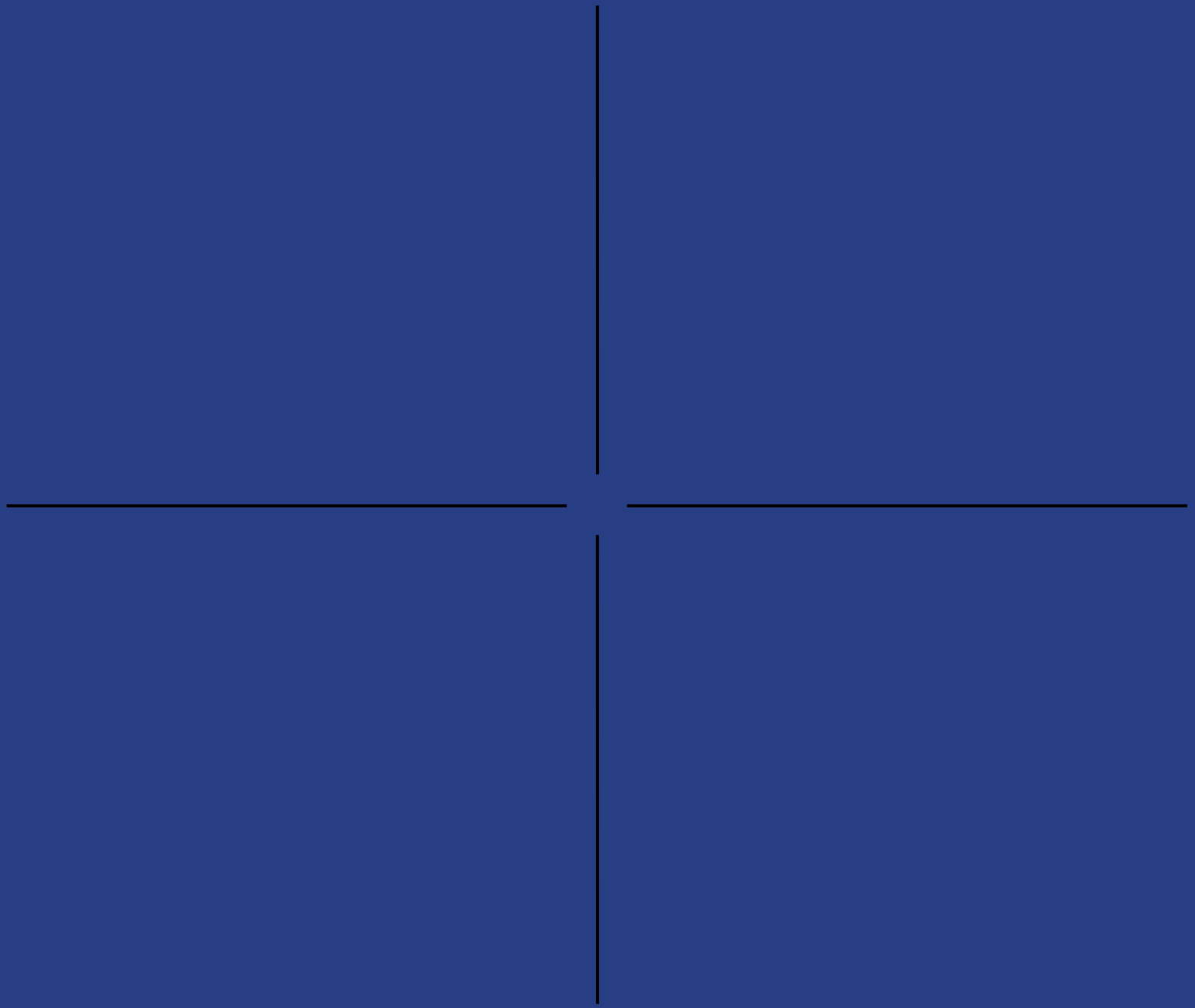
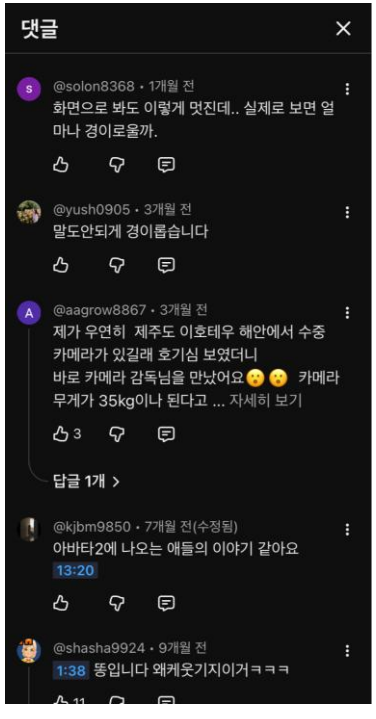


TABLE OF CONTENTS



추진 배경. + AI

유튜브에서 고래영상을 시청하다 한 댓글을 발견하게 됐다.

"AI의 발전으로 진짜가 진짜로 인정받지 못하는 세상이다"

현실의 장면조차 AI로 오해받는 세상이라면

앞으로 우리는 무엇을 믿을 수 있을까?

댓글 한 줄이 문제의식을 자극했고,

누구나 쉽게 이미지의 진위를 판별할 수 있는 AI 생성 이미지 판별앱을 만들었다.

1

허위 정보 확산. Deepfake

AI 생성 이미지가 실사와 구분되지 않아,
SNS·커뮤니티에서 조작된 정보가
빠르게 증폭되고 있다.

2

진위 판별의 어려움.

사용자·플랫폼 모두 이미지 출처와
진위를 눈으로만 판단하기 어려워,
검증 과정이 사실상 불가능에 가까워졌다.

3

저작권 문제.

창작자의 작품이 AI 생성물에 의해
무단 모방·혼동되며,
실제 창작물의 권리 침해 가능성이 커졌다.

4

문화적 혼란.

딥페이크·가짜 이미지의 반복적 노출로
사회적 신뢰가 떨어지고, 공론장과 문화 전반에서
왜곡과 혼란이 발생하고 있다.

내가 만든.
App.
Service.

---

app purpose

tech stack

function

---

notandum

expected effect

01

App purpose

APP을 통해 이루고자 하는 최소 목적

사용자가 업로드한 이미지를 분석해서

AI가 생성/합성한 이미지 vs 사람이 직접 만든 이미지를

구분하는 것을 목적으로 가집니다.

02

Tech stack

기반이 되는 기술과 프로그램

프롬프트엔드 : Streamlit

모델 : Hugging Face의 AI 이미지 판별 공개 모델

- Ateeqq/ai-vs-human-image-detector(바탕)
- Smogy(백그라운드 분석용)

03

Function

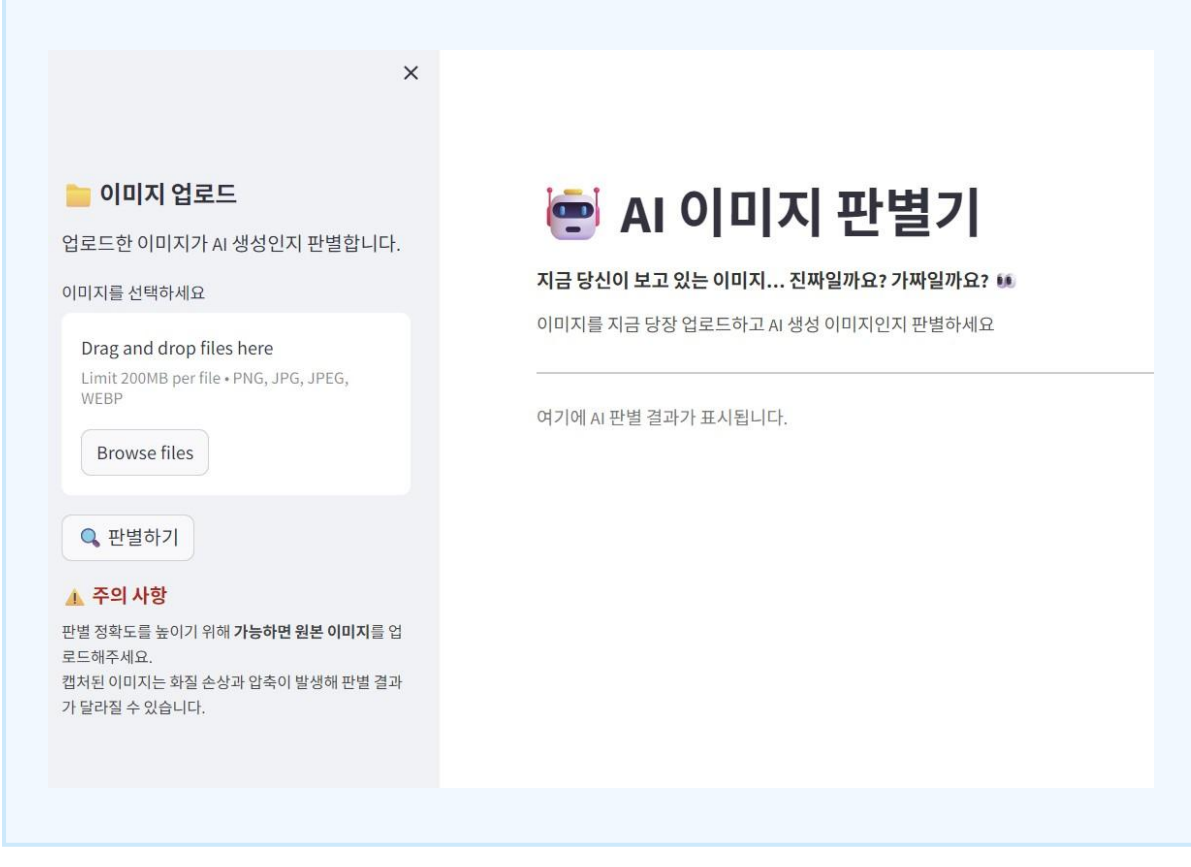
이미지 업로드

- 사용자가 PC나 모바일에서 이미지 선택
- 여러 장 한 번에 업로드 가능

AI 판별

- 모델을 활용해서 이미지 특징 분석
- 판별한 결과를 확률로 표시

주요 기능



04

Notandum

모델/서버 부하

많은 이미지를 동시에 분석하면
모델/서버 부하 증가한다.

오판 가능성 존재

사람이 만든 창작물이라도 창작자가
편집/보정을 하면 AI감지 모델이 오판하는
경우가 매우 흔하게 발생한다.

photo-1763315152539-06fc234b526c.avif

Ateeqq AI 확률: 0.9995

Smogy AI 확률: 0.9708

평균 AI 확률: 0.9851

최종 판정: AI 생성 이미지로 의심됩니다.

- AI 생성물을 사람이 편집/보정하는 경우
- 사람 작품을 사람이 편집/보정하는 경우

구분X

- 아무 편집이 없어도 최신 AI모델은 감지기가 구별 못 할 경우도 존재

*현 기술로는 감지 정확도가 제한적이고, 참고용 정도라고 보면 된다

05

Expected effect

기대 효과

허위 정보 확산 방지

창작 보호 효과

접근성 높은 도구

허위 정보 확산 방지

- AI 이미지 판별 앱을 통해 딥페이크·조작 이미지·AI 생성물의 무분별한 공유를 사전에 차단할 수 있다.
- 사용자는 이미지를 업로드하는 즉시 진위 여부를 판단할 수 있어, SNS에서 잘못된 정보가 빠르게 퍼지는 것을 줄이고 신뢰성 있는 정보 유통에 기여한다.

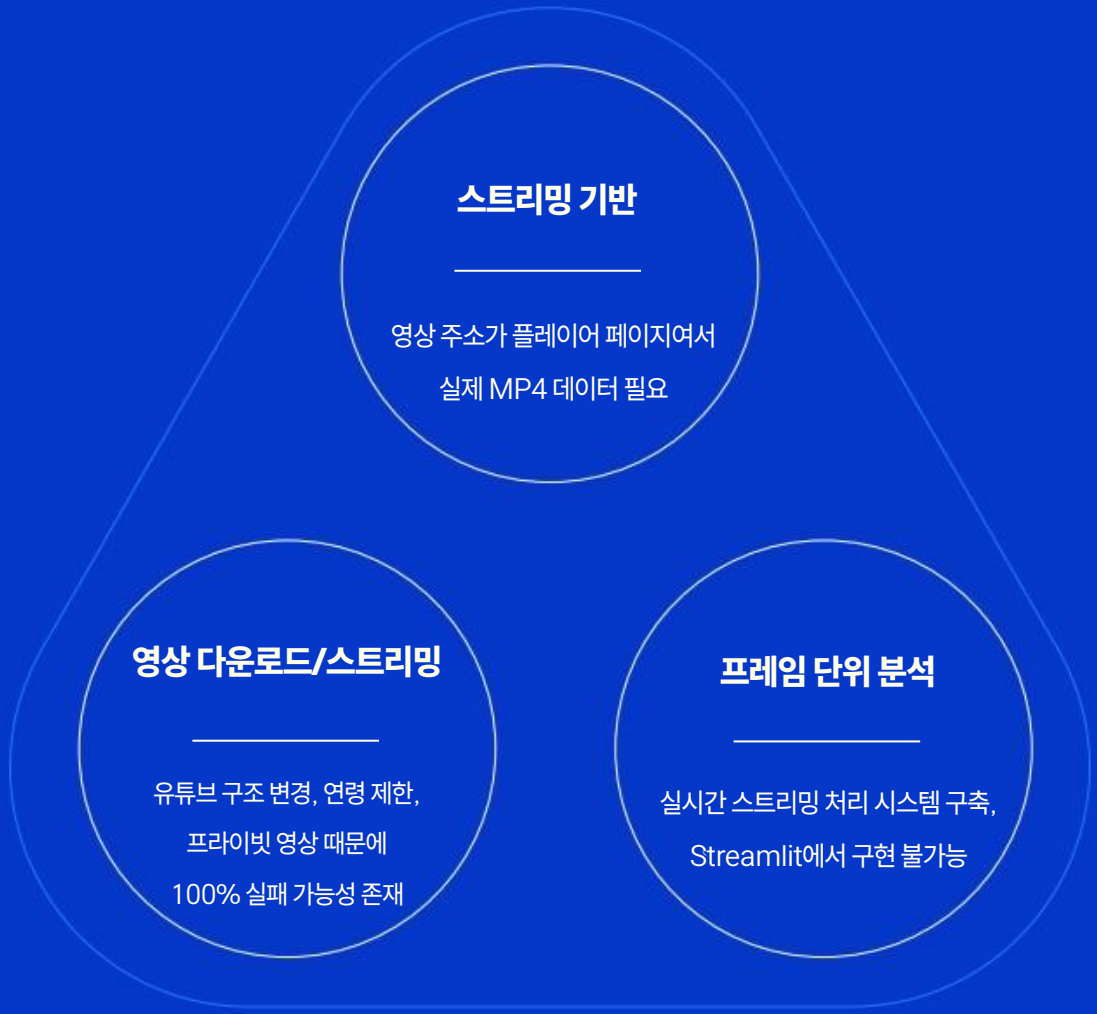
창작 보호 효과

- 실제 작가·디자이너·사진가가 만든 콘텐츠가 AI 생성물과 혼동되는 문제를 완화한다.

접근성 높은 도구

- 별도의 설치나 전문 도구 없이 누구나 웹에서 바로 사용할 수 있는 앱 형태이기 때문에, 다양한 계층이 손쉽게 이미지 진위 여부를 확인할 수 있다.

유튜브 URL만으로 영상을 실시간 판별하고 싶었지만,
유튜브 영상이 스트리밍 기반이라는 점, 프레임 분석이 Streamlit
수준에서 거의 불가능하다는 것과 더불어 영상 다운로드/스트리밍
문제로 100% 실패 가능성이 존재했다.



현재는 이미지만 판별할 수 있지만,
텍스트와 영상, 보다 더 나아가 다양한 형태의 창작물들을
판별할 수 있도록 확장하고 싶다.

이상 발표를 마무리하겠습니다