

# KISA AI 보안 유망기업 육성 지원 사업 성과공유회 발표

2025.12

CONFIDENTIALITY NOTICE: The contents of this message are intended solely for the addressee(s) and may contain confidential and/or privileged information and may be legally protected from disclosure. If you are not the intended recipient of this message or its agent, or if this message has been addressed to you in error, please immediately alert the sender and delete this message. If you are not the intended recipient, you are hereby notified that any use, dissemination, copying, or storage of this message is strictly prohibited.

# Contents

- 1 사업 개요 및 목표**
- 2 사업 수행 현황**
- 3 제품 개발 및 목표 달성 현황**
- 4 향후 계획 및 목표**

# **1 사업 개요 및 목표**

- 01 과제 소개**
- 02 추진 배경 및 필요성**
- 03 사업 핵심 목표**

|         |  |
|---------|--|
| 사업명     | AI 기반 악성 딥페이크 탐지 및 대응 시스템 개발   |
| 사업 책임기관 | (주)누리랩   |
| 사업 책임자  | 최원혁 대표 / 참여 연구원 : 12명 (총 13명으로 구성)   |
| 사업 기간   | 2025년 5월 ~ 2025년 11월(7개월)  |
| 사업 내용   | <p>AI 기반 악성 딥페이크 탐지 및 대응 시스템 개발</p> <ul style="list-style-type: none"><li>• 모니터링 대상 얼굴 식별 및 검색</li><li>• 딥페이크 합성물 탐지 및 크롤링</li><li>• 악성 딥페이크 합성물 검증</li><li>• 서비스 UI 및 탐지 API 제공</li></ul> |

## ISSUE 01

### AI 역기능으로 인한 악성 딥페이크 합성물 피해 사상 최대 기록

‘디지털 성착취물’, 유명인 사칭 ‘금융투자 사기 광고물’ 등 악성 딥페이크  
합성물 피해 날로 심각,  
올해 7월까지 딥페이크 성범죄 차단 건수 1만5천808건,  
올해 전체 건수는 2만7천여건으로 사상 최다를 기록할 것으로 예상

\*방송통신심의위원회

## ISSUE 02

### “딥페이크 성범죄 근절”을 위한 「딥페이크 성범죄 대응 강화 방안」

2024년 11월 6일, 관계부처 합동 4대 분야 10개 과제 발표  
△강력하고 실효적인 처벌 △플랫폼 책임성 제고  
△신속한 피해자 보호 △맞춤형 예방 교육 등

## REQ 01

### 딥페이크 탐지 기술 고도화 요구

- 날로 정교해지는 딥페이크 기술 발전
- AI 역기능으로 악성 딥페이크 합성물 생성 및 유포 만연

## REQ 02

### 디지털 성착취물 등 악성 딥페이크 합성물 상시 모니터링, 탐지 및 대응 요구

- 유포 상시 모니터링
- 증거 채증 및 삭제 등

## REQ 03

### 「딥페이크 성범죄 대응 강화 방안」 4대 분야 10개 과제 지원

- 플랫폼 사업자 의무 강화
- 신속한 피해자 보호 등
- 삭제 지원 강화, 원스톱 지원 체계 구축

## 시제품 개발(협약기간 내)

### “AI 기반 악성 딥페이크 탐지 및 대응 시스템” 연구개발



- ✓ 모니터링 대상 얼굴 식별 및 검색
- ✓ 하이브리드 딥페이크 합성물 탐지 및 크롤링
  - 딥페이크 메타데이터 스캔 + 최신 AI 탐지 모델
  - 인공지능 생성물 워터마크 검출 모듈
- ✓ 악성 딥페이크 합성물 판별
- ✓ 서비스 UI 및 API 제공

## 사업화(협약기간 이후)

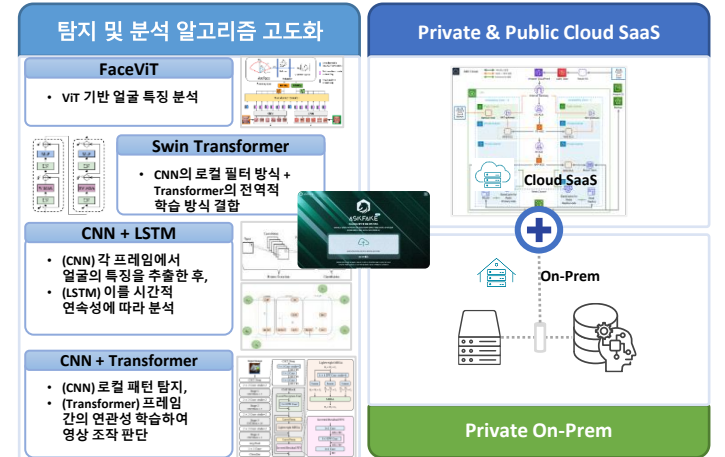
### 「딥페이크 성범죄 대응 강화 방안」 핵심 과제 요구사항 특화 '원스톱 통합 대응 지원' 역량 강화



- ✓ '플랫폼 사업자 의무 준수' 요구사항
  - 청소년 유해물 제공 · 매개 및 불법촬영물 등 유통 방지
  - 성범죄를 게재자 제재 및 모니터링 강화
- ✓ '신속하고 확실한 피해자 보호' 요구사항
  - 딥페이크 성범죄물 신속 삭제
  - 원스톱 지원체계 구축

## 고도화(협약기간 이후)

### 악성 딥페이크 합성물 상시 모니터링 및 탐지, 대응 시스템 고도화



- ✓ 시스템 및 서비스 기능 강화
  - 딥페이크 합성물 탐지 및 분석 알고리즘 고도화
- ✓ 고객 환경 맞춤형 서비스 제공
  - 고객 맞춤형 하이브리드(Cloud/On-Prem) 악성 딥페이크 합성물 검출 시스템

| 사업 내용  | 연구개발 내용   |
|--|---|
| 모니터링 대상 얼굴 식별 및 검색   |   |
| <ul style="list-style-type: none"> <li>모니터링 대상 얼굴 식별 및 학습 모듈 개발</li> </ul> | <ul style="list-style-type: none"> <li>딥러닝 기반 얼굴 검출 모델을 활용하여 이미지/영상에서 얼굴 영역 검출 및 전처리</li> <li>전처리된 얼굴에서 대상 얼굴의 고유 특징 추출</li> <li>탐지된 객체의 특징을 수학적 연산이 가능한 벡터값으로 변환 및 DB 저장</li> <li>머신러닝 기반 모니터링 대상 얼굴 분류 학습</li> </ul>                      |
| <ul style="list-style-type: none"> <li>얼굴 이미지 검색 모듈 개발</li> </ul>          | <ul style="list-style-type: none"> <li>이용자 얼굴 이미지 벡터와 내부 DB에 저장된 벡터 간의 유사도 분석</li> <li>여러 비교값을 종합하여 유사도가 높은 상위 N개 유사 벡터 추출</li> </ul>   |
| 딥페이크 합성물 탐지 및 크롤링 모듈 개발  |   |
| <ul style="list-style-type: none"> <li>딥페이크 메타데이터 스캔 모듈</li> </ul>         | <ul style="list-style-type: none"> <li>이미지/영상 파일의 EXIF, 인코딩 포맷, 생성 로그 등을 추출하여 정상 촬영물과 합성물의 차이 분석</li> <li>비정상적인 압축 패턴·프레임 타임스탬프 불일치 자동 탐지</li> <li>메타데이터 기반 신뢰도 점수 산출 로직 구현</li> </ul>  |
| <ul style="list-style-type: none"> <li>딥페이크 합성물 AI 탐지 모듈</li> </ul>        | <ul style="list-style-type: none"> <li>EfficientNet-B7 등 딥러닝 모델로 이미지/영상의 픽셀 단위 특성을 학습, 합성 여부 판별</li> <li>정확도 향상을 위해 얼굴 특징 및 전역 이미지 특징 동시 활용</li> </ul>  |
| <ul style="list-style-type: none"> <li>인공지능 생성물 워터마크 검출 모듈</li> </ul>      | <ul style="list-style-type: none"> <li>AI 생성물 표시 의무(AI 기본법 제31조)에 맞춰, 삽입된 디지털 워터마크 탐지 기능 구현</li> <li>딥러닝 기술을 활용해 이미지 내 미세한 주파수 변화/패턴 학습을 통한 워터마크 탐지</li> </ul>  |
| <ul style="list-style-type: none"> <li>모니터링 및 탐지 대상 크롤러</li> </ul>         | <ul style="list-style-type: none"> <li>텔레그램, 페이스북, X(트위터), 웹 포럼 등 다양한 채널에서 이미지 자동 수집</li> <li>딥러닝 기반 얼굴 검출 모델을 활용하여 이미지/영상에서 얼굴 영역 검출 및 전처리</li> <li>전처리된 얼굴에서 대상 얼굴의 고유 특징 추출</li> <li>탐지된 객체의 특징을 수학적 연산이 가능한 벡터값으로 변환 및 DB 저장</li> </ul> |
| 악성 딥페이크 합성물 검증   |   |
| <ul style="list-style-type: none"> <li>악성 딥페이크 합성물 검증 모듈</li> </ul>        | <ul style="list-style-type: none"> <li>NSFW(Not Safe For Work) 분류 모델을 사용하여 이미지, 영상에서 성인물(노출, 음란물 등) 여부 판단</li> <li>딥페이크 AI 탐지 점수, 워터마크 유무, 메타데이터 이상 징후를 종합적으로 판단하여 악성 딥페이크 합성물 검증</li> </ul>  |
| 서비스 UI 및 API 제공  |   |
| <ul style="list-style-type: none"> <li>이용자 서비스 UI 개발</li> </ul>            | <ul style="list-style-type: none"> <li>웹 기반 UI로 회원 등록, 얼굴 등록, 얼굴 검색, 딥페이크 탐지 결과 등 정보 제공</li> </ul>  |
| <ul style="list-style-type: none"> <li>악성 딥페이크 탐지 API</li> </ul>           | <ul style="list-style-type: none"> <li>외부 서비스와 연동할 수 있도록 Restful API 형태로 제공</li> </ul>  |

## **2 사업 수행 현황**

### **01 연구개발 추진 일정 및 달성 현황**



| 사업 내용                      | 진척도    | 개발 일정 |    |    |    |    |     |     | 달성<br>현황 |
|----------------------------|--------|-------|----|----|----|----|-----|-----|----------|
|                            |        | 5월    | 6월 | 7월 | 8월 | 9월 | 10월 | 11월 |          |
| 모니터링 대상 얼굴 식별 및 검색         | ✓ 100% |       |    |    |    |    |     |     | 완료       |
| ▪ 모니터링 대상 얼굴 식별 및 학습 모듈 개발 | 100%   |       |    |    |    |    |     |     |          |
| ▪ 얼굴 이미지 검색 모듈 개발          | 100%   |       |    |    |    |    |     |     |          |
| 딥페이크 합성물 탐지 및 크롤링 모듈 개발    | ✓ 100% |       |    |    |    |    |     |     | 완료       |
| ▪ 딥페이크 메타데이터 스캔 모듈         | 100%   |       |    |    |    |    |     |     |          |
| ▪ 딥페이크 합성물 AI 탐지 모듈        | 100%   |       |    |    |    |    |     |     |          |
| ▪ 인공지능 생성물 워터마크 검출 모듈      | 100%   |       |    |    |    |    |     |     |          |
| ▪ 모니터링 및 탐지 대상 크롤러         | 100%   |       |    |    |    |    |     |     |          |
| 악성 딥페이크 합성물 검증             | ✓ 100% |       |    |    |    |    |     |     | 완료       |
| ▪ 악성 딥페이크 합성물 검증 모듈        | 100%   |       |    |    |    |    |     |     |          |
| 서비스 UI 및 API 제공            | ✓ 100% |       |    |    |    |    |     |     | 완료       |
| ▪ 이용자 서비스 UI 개발            | 100%   |       |    |    |    |    |     |     |          |
| ▪ 악성 딥페이크 탐지 API           | 100%   |       |    |    |    |    |     |     |          |

3

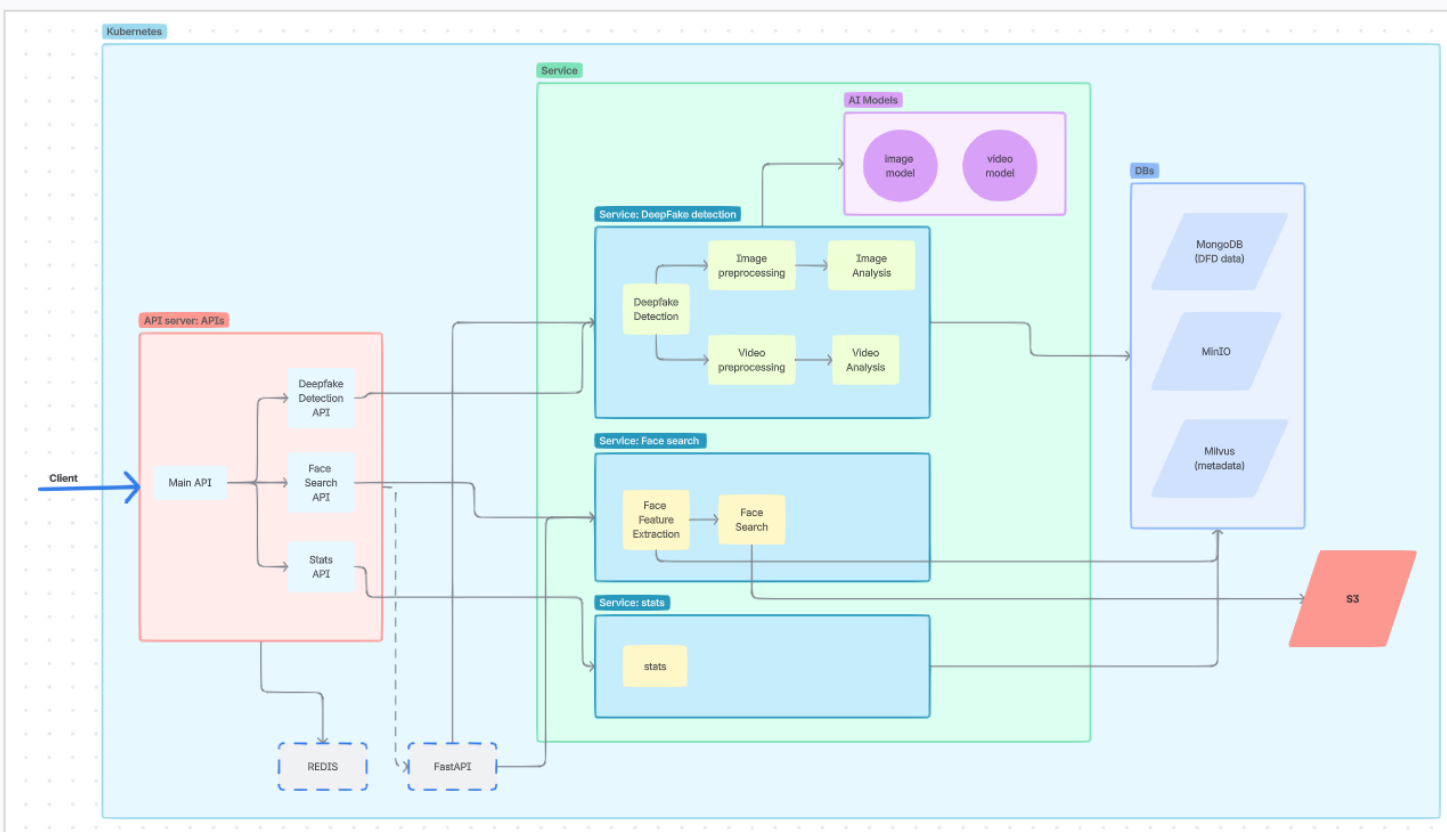
## 제품 개발 및 목표 달성 현황

- 01 기술 개발 현황
- 02 정량/정성적 목표 달성 현황
- 03 제품 시연



## AI 기반 악성 딥페이크 탐지 및 대응 시스템

- ✓ AI 기반 악성 딥페이크 탐지 및 대응 서비스 제공을 위한 K3s microservice 구성
- ✓ (시스템) OS: Ubuntu 22.04.5 LTS(kernel: 6.8.0-65-generic), Memory: 128Gi, SSD: 480GB \* 2, GPU: 24GB (모델명: RTX A5000), 학습 시스템(Tesla A100 40GB)



[AI 기반 악성 딥페이크 탐지 및 대응 시스템 구성도]

### [주요 플랫폼 컴포넌트 (Cluster)]

#### HA K3s Cluster

- 경량화된 Kubernetes(K3s) 기반 고가용성(High Availability) 클러스터 구축
- 마스터 노드 장애에도 서비스 지속, 마이크로서비스를 안정적으로 배포/운영 가능

#### Ceph

- 분산 스토리지(블록, 파일, 오브젝트) 제공, 데이터 무결성과 확장성 확보
- K3s 내에서 사용되는 영속 볼륨(Persistent Volume)을 안정적으로 지원

#### Traefik

- 인그레스 컨트롤러 역할 수행, 서비스 라우팅/로드밸런싱 제공
- 클러스터 외부에서 내부 서비스에 접근할 수 있도록 도메인/SSL 인증서 자동 관리

#### MongoDB

- 비정형 데이터 및 메타데이터 저장용 NoSQL 데이터베이스(통계 등 제한된 용도)
- 로그, 설정값, 사용자 데이터 등을 빠르게 저장/조회 가능

#### MinIO

- 오브젝트 스토리지
- 이미지, 동영상, 대용량 데이터 파일을 안전하게 저장 및 관리(크롤링 이미지 저장)

#### Milvus

- 벡터 데이터베이스로 AI/머신러닝 기반 검색(Embedding 기반 검색) 지원
- 얼굴 인식, 문서 임베딩, 추천 시스템 등 벡터 유사도 검색 성능 최적화

#### RabbitMQ

- 메시지 큐 시스템, 서비스 간 비동기 메시징 처리
- 분산 환경에서 안정적인 이벤트 전송과 마이크로서비스 간 데이터 흐름 제어

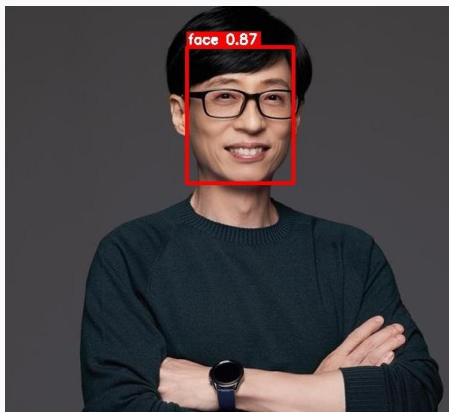
#### Harbor

- 프라이빗 컨테이너 이미지 레지스트리 운영
- 권한 및 보안 강화된 Docker 이미지 저장소 자체 구성, 서드파티 의존성 최소화



## 모니터링 대상 얼굴 식별 및 학습 모듈 개발

- ✓ 딥러닝 기반 얼굴 검출 모델(Facenet 512) 활용하여 이미지/영상에서 얼굴 감지(Detection) → 모니터링 대상 얼굴의 고유 특징 추출 → 벡터값으로 변환 및 DB 저장 → 머신러닝 기반 모니터링 대상 얼굴 분류 학습



## Facenet 512 를 이용한 얼굴 특징 정보 추출 및 검색

### ASKFAKE AI Scan Service

모바일·데스크탑 모두 지원 · 파일·

#### Face: Extract Only

Face Image

파일 선택 20250825\_154412.png

Extract

```
{
  "embedding": [
    1.7981174,
    -0.1056501,
    -0.7082761,
    0.69491684,
    0.20842712,
    -0.09775504,
    -1.2793331,
    1.7525249,
    0.014722697,
    -0.8937305,
    0.12754518,
    ...
  ]
}
```



## 모니터링 대상 얼굴 식별 및 학습 모듈 개발 - 이미지 학습 모델 선정

- ✓ **XceptionNet** 모델 기반으로 **Crop** 된 데이터셋을 학습한 모델 선정
- ✓ 훈련되지 않은 이미지로 식별 테스트 결과: **F1-Score, Accuracy 91.8% 이상 달성**

### XceptionNet vs. EfficientNet\_b0

1. Dataset : Aihub, HuggingFace, Zenodo 204,298장

| Model           | Accuracy | Precision | Recall | F1 Score |
|-----------------|----------|-----------|--------|----------|
| XceptionNet     | 0.9851   | 0.9862    | 0.9827 | 0.9844   |
| EfficientNet_b0 | 0.3936   | 0.4418    | 0.4762 | 0.3407   |

2. Dataset : EfficientNet\_b0 데이터셋 45,796장

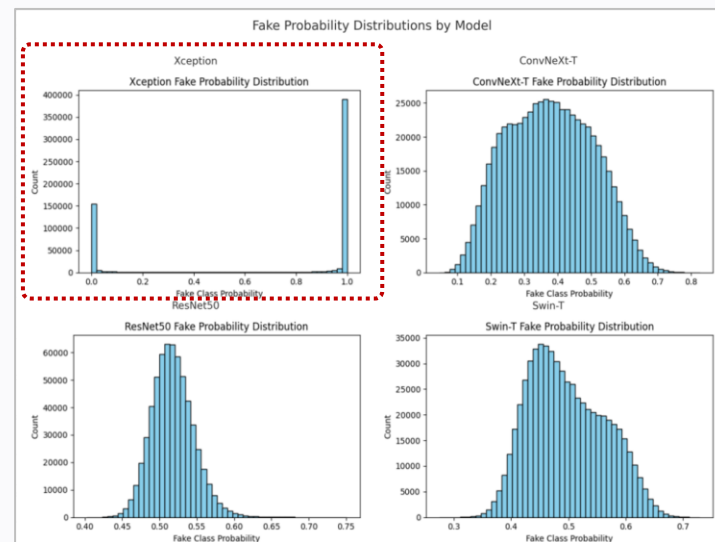
| Model           | Accuracy | Precision | Recall | F1 Score |
|-----------------|----------|-----------|--------|----------|
| XceptionNet     | 0.9388   | 0.9393    | 0.9387 | 0.9387   |
| EfficientNet_b0 | 0.6982   | 0.7204    | 0.6981 | 0.6903   |

### XceptionNet threshold Test

| Threshold | Accuracy | Precision | Recall | F1 Score |
|-----------|----------|-----------|--------|----------|
| 0.6       | 0.9815   | 0.9809    | 0.9808 | 0.9809   |
| 0.7       | 0.9813   | 0.9802    | 0.9812 | 0.9807   |
| 0.9       | 0.9777   | 0.9754    | 0.9792 | 0.9771   |

### XceptionNet vs. ResNet50, ConvNeXt-T, Swin-T

| Model       | Accuracy | Precision | Recall | F1 Score |
|-------------|----------|-----------|--------|----------|
| XceptionNet | 0.9416   | 0.8708    | 0.9322 | 0.9004   |
| ResNet50    | 0.2816   | 0.2807    | 0.9845 | 0.4369   |
| ConvNeXt-T  | 0.3077   | 0.2871    | 0.9745 | 0.4435   |
| Swin-T      | 0.2646   | 0.2545    | 0.8284 | 0.3894   |

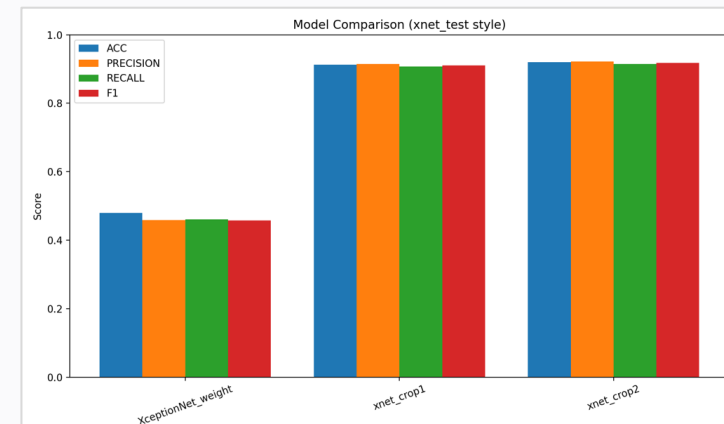


[Fake 확률 분포 히스토그램]

### XceptionNet 가중치 성능 평가

- XceptionNet\_weight.pth(Crop Dataset, Kg\_dataset 존재 X)
- Xnet\_crop2.pth(Fake 비율 = 0.75로 낮춤, mean&std X)
- Xnet\_crop1.pth(Fake 비율 = 0.5로 낮춤, mean&std = [0.5]\*3)

| Weight                 | Accuracy | Precision | Recall | F1 Score |
|------------------------|----------|-----------|--------|----------|
| XceptionNet_weight.pth | 0.4801   | 0.4586    | 0.4612 | 0.4572   |
| Xnet_crop1.pth         | 0.9127   | 0.9145    | 0.9079 | 0.9105   |
| Xnet_crop2.pth         | 0.9201   | 0.9225    | 0.9152 | 0.9181   |



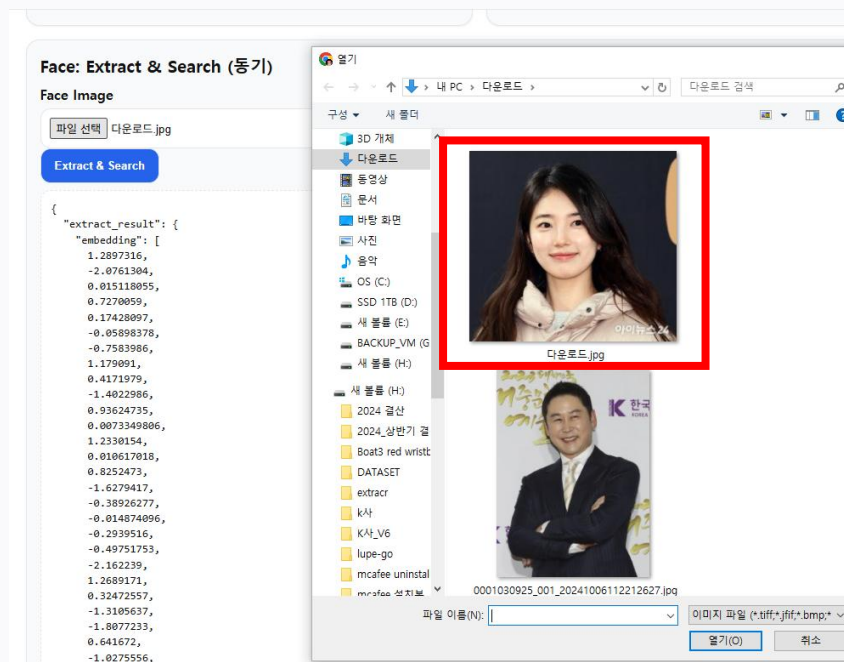
[Fake 확률 분포 히스토그램]



## 얼굴 이미지 검색 모듈 개발

- ✓ 딥러닝 기반 얼굴 검출 모델(Facenet 512) 활용하여 이미지/영상에서 얼굴 감지(Detection) → 모니터링 대상 얼굴의 고유 특징 추출(Embedding) → 벡터값으로 변환 및 DB 저장(Vector Indexing) → 모니터링 대상 얼굴 검색(Search)

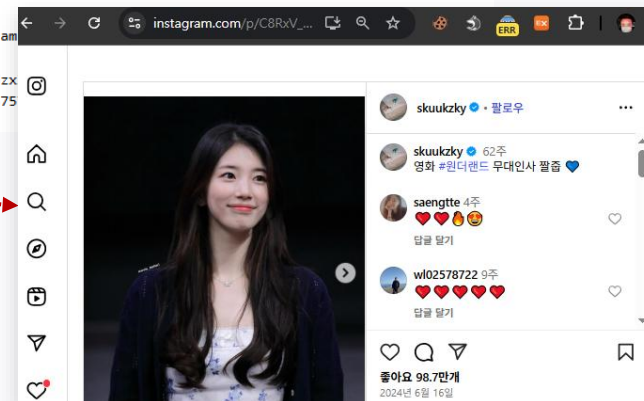
### ❖ 모니터링 대상 얼굴 특성 추출 및 임베딩



### ❖ 모니터링 대상 얼굴 특성 검색 - site\_url

```
"hits": [
  {
    "classification": "",
    "created_at": 1741252129211,
    "download_url": "https://instagram.fb1r2-2.fna.fbcdn.net/v/t51.29350-15/448377897_1217378976301823_1319693642.jpg_e35_p1080x1080_tt6&nc_ht=instagram.fb1r2-2.fna.fbcdn.net&nc_cat=111&nc_ohc=za-FdTpeVh8Q7kNvgEGuLPw&nc_gid=afaa9d4be2884b9c3afe34753bc701f3&edm=APs17CUBAAAA&ccb=7-5&oh=00_AYA1rdL5v1qJKHRwX60yo39o-hLyUlsxStA&oe=675730E6&nc_sid=10d13b",
    "face_confidence": 0,
    "is_nsfw": false,
    "score": 0,
    "sha256": "b7acf1db83971ca48d5beb15485330e4d054ffa9078e2f3aa85cdc3eaabfcb13",
    "similarity": 0.9251569,
    "site_domain": "",
    "site_id": "",
    "site_url": "https://www.instagram.com/p/C8RxV_mvCco/"
  },
  {
    "classification": "",
    "created_at": 1741252105779,
    "download_url": "https://scontent-phx1-1.cdninstagram.com/jpg_e35_p1080x1080&nc_ht=scontent-phx1-1.cdninstagram.com&nc_cat=109&nc_ohc=UwTHn2kjtCQ7kNvgGiHx5&oh=00_AYBbcuO_-okiAiwehQCT0JCyu_L005Arrkja6z_tz4m9tQ&oe=675",
    "face_confidence": 0,

```





## 딥페이크 메타데이터 스캔 모듈 (C2PA)

- ✓ C2PA: 디지털 콘텐츠의 출처 및 변경 이력 추적, 위변조 방지 및 진위성 보장, 딥페이크·허위정보 대응을 위한 표준화 프레임워크
- ✓ C2PA 존재하는 주요 AI 기반 이미지 합성물 생성기: OpenAI(DALL-E, Sora), Adobe Firefly, Bria, Runway, Bing Image Creator 등

| OpenAI (DALL-E, Sora)   |                                 |
|-------------------------|---------------------------------|
| GPT-4o                  | 생성한 AI 모델에 대한 이름                |
| OpenAI API              | OpenAI의 클라우드 API                |
| Sora                    | 생성한 비디오 AI 모델에 대한 이름            |
| ChatGPT                 | OpenAI의 대화형 AI 웹 인터페이스          |
| org.cai.c2pa_rs         | C2PA 메타데이터를 삽입, 서명 등을 수행한 톨의 출처 |
| trainedAlgorithmicMedia | AI 모델이 직접 생성했다는 것 의미            |

| Adobe Firefly               |                               |
|-----------------------------|-------------------------------|
| Adobe Firefly               | 생성한 AI 모델에 대한 이름              |
| com.adobe.firefly.version   | Adobe Firefly의 버전에 대한 Key     |
| com.adobe.firefly.operation | Adobe Firefly가 수행한 작업에 대한 Key |
| Adobe Inc                   | 서명을 발행한 기관                    |
| trainedAlgorithmicMedia     | AI 모델이 직접 생성했다는 것 의미          |

| Bria                    |                                     |
|-------------------------|-------------------------------------|
| Bria AI                 | 이미지는 Bria AI 플랫폼을 사용                |
| Bria_Ai c2pa-rs         | Bria AI가 C2PA 표준을 기반으로 메타데이터 서명을 생성 |
| XML:com.adobe.xmp       | Adobe XMP 형식의 메타데이터 사용              |
| Generated by Bria       | 이미지가 Bria 플랫폼에 의해 생성                |
| trainedAlgorithmicMedia | AI 모델이 직접 생성했다는 것 의미                |

| Runway                         |                                 |
|--------------------------------|---------------------------------|
| CN=eng-leadership@runwayml.com | Runway AI 소속 리더십 계정으로 서명되었음을 의미 |
| O=RUNWAY AI                    | Runway AI 법인 이름으로 생성            |
| DE-7124526                     | Runway AI의 등록 식별자 또는 사업자 등록 번호  |
| c2pa-rs                        | C2PA 메타데이터를 삽입, 서명 등을 수행한 톨의 출처 |
| trainedAlgorithmicMedia        | AI 모델이 직접 생성했다는 것 의미            |

| Bing Image Creator                        |   |
|---|---|
| Microsoft_Responsible_AI                  | Microsoft Responsible AI 가이드라인에 따라 생성   |
| Microsoft Responsible AI Image Provenance | Microsoft AI 출처 정보                      |
| Microsoft Corporation                     | Microsoft 저작자 정보                        |
| Image Creator from Designer               | Bing Image Creator 만든 이미지에 추가되는 생성 톨 정보 |
| trainedAlgorithmicMedia                   | AI 모델이 직접 생성했다는 것 의미                    |

| Titan                    |   |
|--------------------------|---|
| Titan Image Generator G1 | Amazon의 생성 모델 Titan의 첫번째 버전 의미          |
| Amazon_Bedrock c2pa-rs   | Amazon의 AI 서비스 플랫폼 AWS Bedrock으로 이미지 생성 |
| Amazon Bedrock           | AWS에서 제공하는 생성형 AI 서비스 통합 플랫폼            |
| Amazon Web Services      | Amazon이 제공하는 클라우드 컴퓨팅 서비스               |
| trainedAlgorithmicMedia  | AI 모델이 직접 생성했다는 것 의미                    |





## 딥페이크 메타데이터 스캔 모듈(이외 META DATA)

- ✓ C2PA가 존재하지 않는 이미지 합성물 생성기
- ✓ Midjourney, Automatic 1111, HunYuan, Kling 등 특정 패턴화된 이미지 합성물 메타데이터 식별 및 탐지

| Midjourney  |                                     |
|---|-------------------------------------|
| --ar  | Midjourney 명령어                      |
| --v   | Midjourney 전용 옵션                    |
| Job ID  | Midjourney 고유 형식                    |
| Creation Time, Author, Description, XML:com.adobe.xmp | Midjourney에서 생성된 이미지에 흔히 나타나는 메타데이터 |
| trainedAlgorithmicMedia                               | C2PA 메타데이터를 삽입, 서명 등을 수행한 톨의 출처     |

| HunYuan         |                               |
|-----------------|-------------------------------|
| HunYuan         | ilst 부분에 모델명 존재               |
| Tencent         | ilst 부분에 기업명 존재               |
| ContentProducer | AI가 만든 콘텐츠 의미                 |
| Propagator      | 해당 콘텐츠 배포에 대한 의미로 AI가 배포함을 의미 |

| Automatic 1111 |                                    |
|----------------|------------------------------------|
| parameters     | 대부분의 AI 생성기에서 사용                   |
| Sampler        | Sampler 알고리즘 종류에 대한 설명             |
| CFG scale      | 대부분의 WebUI에서 사용                    |
| Model hash     | 모델의 SHA256 해시로 Automatic 1111에서 사용 |
| Lora hashes    | LORA 모듈의 해시값으로 Automatic 1111에서 사용 |

| Kling                       |   |
|-----------------------------|---|
| metadata0                   | Kling에서 해당 명칭으로 Keys가 생성                    |
| metadata1                   | Kling에서 해당 명칭으로 Keys가 생성                    |
| kling-ai                    | AI 모델에 대한 명칭이 존재                            |
| security.kling.meta_encrypt | Kling에서 protobuf를 이용하고, Metadata Encryption |





## 딥페이크 메타데이터 스캔 모듈

- ✓ 딥페이크 생성 Application 별 특정 메타데이터를 패턴화하여 딥페이크 합성물을 생성한 AI Application 및 딥페이크 합성물 여부 판단
- ✓ 패턴 매치 결과: application, category, encoding, subcategory, AI 여부 및 확률, 앱/모델 정보 제공

```
-----매치 패턴-----  
{'application': ['OpenAI'],  
  'category': ['AI'],  
  'encoding': ['SEI', 'FYTP'],  
  'subcategory': ['Sora']}  
AI 여부: True  
AI 확률: 1.0  
앱/모델: OpenAI/Sora
```

### ❖ 딥페이크 메타데이터 스캔 결과: 매치 패턴

'application': 사용한 앱 / 서비스 (ChatGPT, ComfyUI 등) 표기

'category': 미디어 진위성 (AI, Real) 표기

'encoding': 메타데이터 인코딩 (c2pa, exif 등) 구조 표기

'subcategory': 앱 내의 하위 범주 (GPT-4o 등) 표기

AI 여부: True/False

AI 확률: 0.98742 (정량적인 확률 값 제공)

앱/모델: application/subcategory



## 딥페이크 합성물 AI 탐지 모듈

- ✓ **XceptionNet** 모델을 이용한 딥페이크 합성물 학습/탐지
- ✓ 정상 이미지 → 낮은 Score, **딥페이크 합성물** → **높은 Score** (Ex. 'score': '98.7648')

### **XceptionNet** 모델을 이용한 딥페이크 합성물 학습/탐지



### 정상 이미지 탐지 시 낮은 Score

#### Image Scan (Callback)

Image File

파일 선택 crop\_1756112030194.png

Scan Image

TID: 37ec2f9f630b4cba95b565ef59fe11a8

카메라 바로 열림(기기/브라우저에 따라 다를 수 있음).

```
{
  "elapsed_time": "1.4962369269924238",
  "score": "11.0211",
  "sha256":
    "0e460992c3755039ec10fd9b1c788e68a73c3e26969d261496d0
    1645a19d6de7",
  "status_code": "200",
  "tid": "37ec2f9f630b4cba95b565ef59fe11a8"
}
```



## 인공지능 생성물 워터마크 검출 모듈

- ✓ 딥러닝 방식의 워터마크 탐지 모델 선정: **DWNet(Dual-Domain Watermark Network)**  
픽셀/주파수 이중 도메인 특징을 결합하여 워터마크의 미세한 주기/위상 변화를 배경 질감과 분리, 낮은 임베딩 강도에서도 신호를 뽑아냄  
→ 비가시성 워터마크 식별에 가장 효과적
- ✓ 총 9개의 모델을 통해 비가시성 워터마크 탐지율 테스트: LSB, DCT, DWT, SVD 방식의 비가시성 워터마크 이미지 100만장 생성 후 학습에 이용

| 모델                          | Accuracy      | Precision     | Recall        | F1-score      |
|-----------------------------|---------------|---------------|---------------|---------------|
| EfficientNet_b0             | 0.5120        | 0.5098        | 0.5990        | 0.5508        |
| ResNet18                    | 0.5012        | 0.5006        | 0.7459        | 0.5991        |
| ResNet50                    | 0.5007        | 0.5016        | 0.2161        | 0.3020        |
| ConvNeXtV2-Base             | 0.6666        | 0.6666        | 1.0000        | 0.8000        |
| ConvNeXtV2-Tiny             | 0.5605        | 0.6630        | 0.6928        | 0.6776        |
| <b>DWNet</b>                | <b>0.6401</b> | <b>0.6638</b> | <b>0.9321</b> | <b>0.7754</b> |
| Siamese_Network (LSB)       | 1.0000        | 1.0000        | 1.0000        | 1.0000        |
| Siamese_Network (DctDwtSVD) | 1.0000        | 1.0000        | 1.0000        | 1.0000        |
| Siamese_Network (All)       | 1.0000        | 1.0000        | 1.0000        | 1.0000        |



**모델 선정**

※ 비가시성 워터마크(Invisible Watermark) 생성 방법

LSB: Least Significant Bit(최하위 비트 변조), 이미지의 각 픽셀 값의 최하위 비트, 즉 가장 중요도가 낮은 마지막 비트를 워터마크 데이터로 바꾸는 변조 방식

DCT: Discrete Cosine Transform, 이미지를 여러 주파수 성분으로 분해하는 기술로, JPEG 압축의 핵심 원리이기도 함

DWT: Discrete Wavelet Transform, 이미지를 저주파와 고주파 성분으로 반복적으로 분해하여 계층적인 구조로 표현하는 기술

SVD: Singular Value Decomposition(특이값 분해), 행렬을 세 개의 다른 행렬의 곱으로 분해하는 기술, 이미지를 행렬로 간주했을 때, 대각 행렬 내 특이값의 안정성(이미지에 압축, 노이즈 추가, 잘라내기 등의 변형이 있어도 특이값은 다른 성분들에 비해 상대적으로 변화가 적음)을 이용하여 해당 특이값에 워터마크 정보를 삽입하는 방식

※ Siamese Network

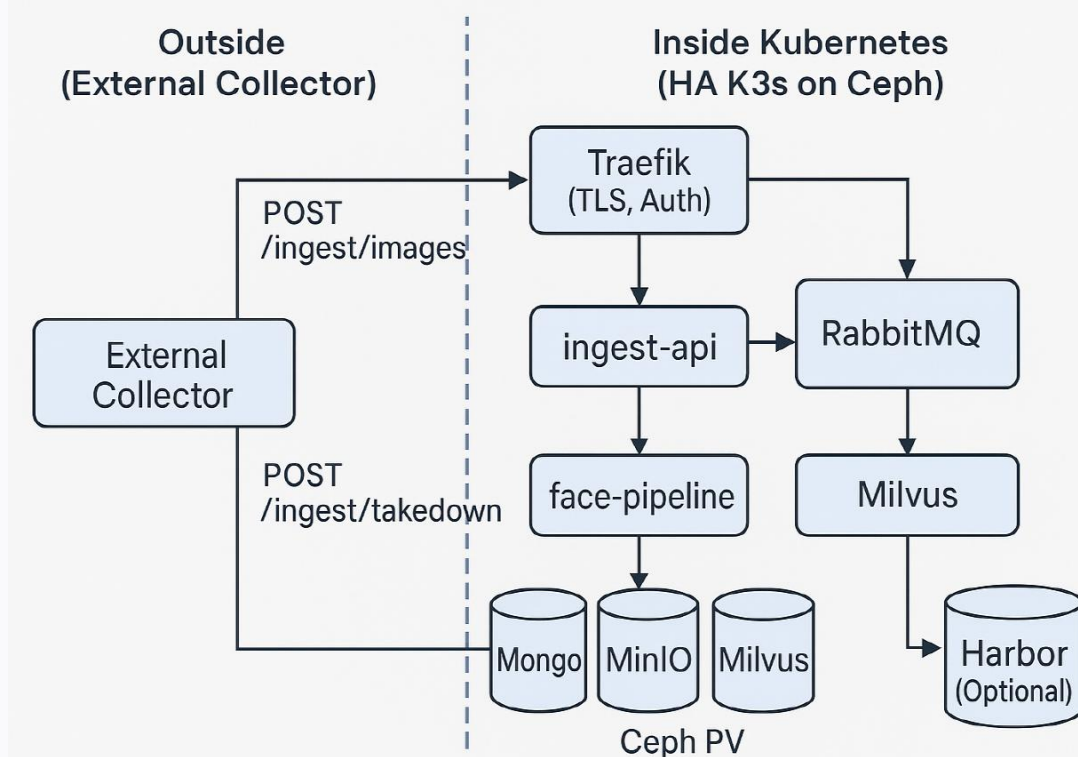
두 이미지 임베딩 값의 거리/유사도를 측정하는데 실제로 이미지를 외우거나 거리/유사도의 차이가 존재하지 않아 모든 값이 1.0이 나오지만 실제 테스트는 0.5 수준밖에 되지 않음



## 모니터링 및 탐지 대상 크롤러 개발

### ✓ 크롤러 엔진 구성

Outside External Collector & Inside K3s (HA K3s on Ceph)



### ❖ 크롤러 아키텍처

Traefik : 외부 엔드포인트 TLS/라우팅/인증.

ingest-api : 이미지 메타 수신 큐 접수.

RabbitMQ : 비동기 작업 큐(백프레셔/재시도).

face-pipeline : 얼굴탐지 -> 정렬 -> 임베딩 -> 중복제거 -> 저장.

MongoDB : 메타데이터, 잡 상태, 감사 로그, 삭제 티켓.

Milvus : 벡터 인덱스/검색.

MinIO : 썸네일 저장.

K3s(HA)/Ceph/Harbor: 배포, 영속, 볼륨, 이미지 레지스트리



## 모니터링 및 탐지 대상 크롤러 개발

- ✓ 텔레그램, 페이스북, X(트위터), 웹 포럼 등 다양한 채널에서 이미지 자동 수집
- ✓ 딥러닝 기반 얼굴 검출 모델을 활용하여 이미지/영상에서 얼굴 영역 검출 및 전처리 → 대상 얼굴의 고유 특징 추출 → 벡터값으로 변환 및 DB 저장

### 크롤링 대상 사이트 선정: (예시) AskURL에서 수집한 adult, illegal 사이트

ASKURL Console

대시보드

현황

분석

사용자별 현황

스캔 요청

URL

SMS

GetAPK

스캔 결과

URL

SMS

GetAPK

classification="adult" || classification="illegal"

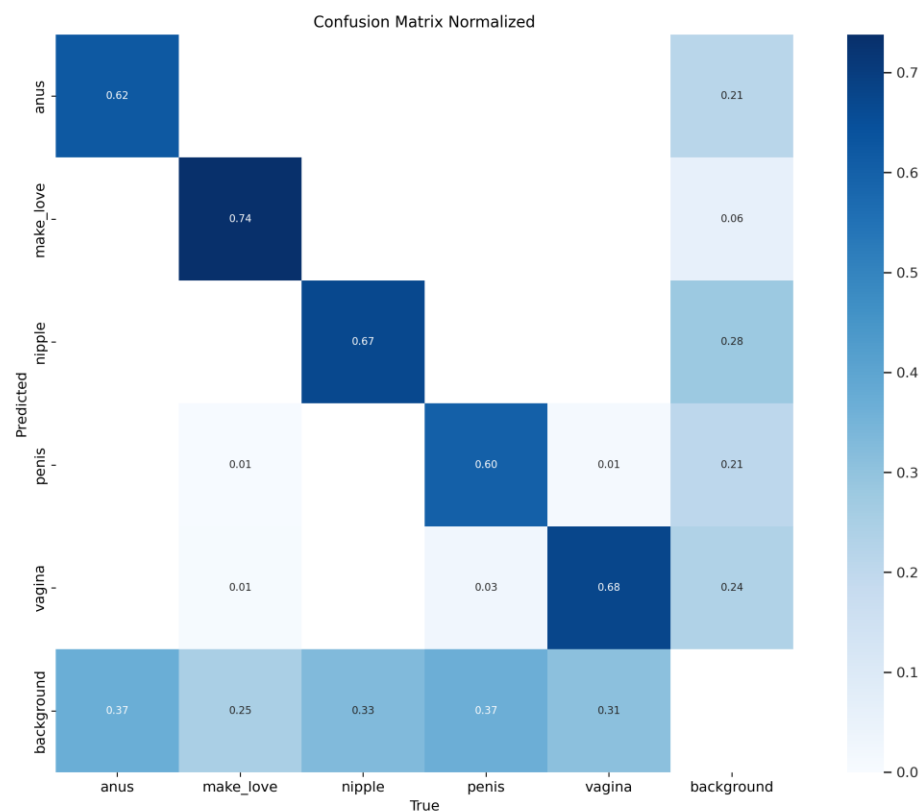
MN +

| <input type="checkbox"/> | No. | 날짜                  | URL   | 제목  |
|--------------------------|-----|---------------------|---|---|
| <input type="checkbox"/> | 1   | 2024-07-12 19:46:03 | https://filemong.com/                               | 파일몽 - 신규 웹하드 p2p사이트 순위 1위 추천 다운로드사이트                      |
| <input type="checkbox"/> | 2   | 2024-07-12 20:02:59 | https://absolutewire.com/vital-information-regar... | Vital Information Regarding Massage Therapy - Absolute... |
| <input type="checkbox"/> | 3   | 2024-07-12 20:32:59 | https://cheapbookmarking.com/story17375059/...      | 출장안마  |
| <input type="checkbox"/> | 4   | 2024-07-12 20:33:14 | https://m1bar.com/user/heightgroup2                 |   |
| <input type="checkbox"/> | 5   | 2024-07-13 02:15:26 | https://bookmarkrange.com/storv18684979/%FC...      | 출장마사지   |



## 악성 딥페이크 합성물 검증 모듈 개발

- ✓ 모델은 NSFW로 라벨링된 신체 부위(예: nipple, penis 등)를 bounding box로 학습해, 이미지에서 해당 객체를 빠르게 탐지하도록 학습
- ✓ 추론 시 입력 이미지를 스캔하여 NSFW 클래스에 해당하는 객체가 감지되면 해당 영역과 확률을 보고 NSFW 여부 결정



NSFW 라벨링 및 혼동행렬



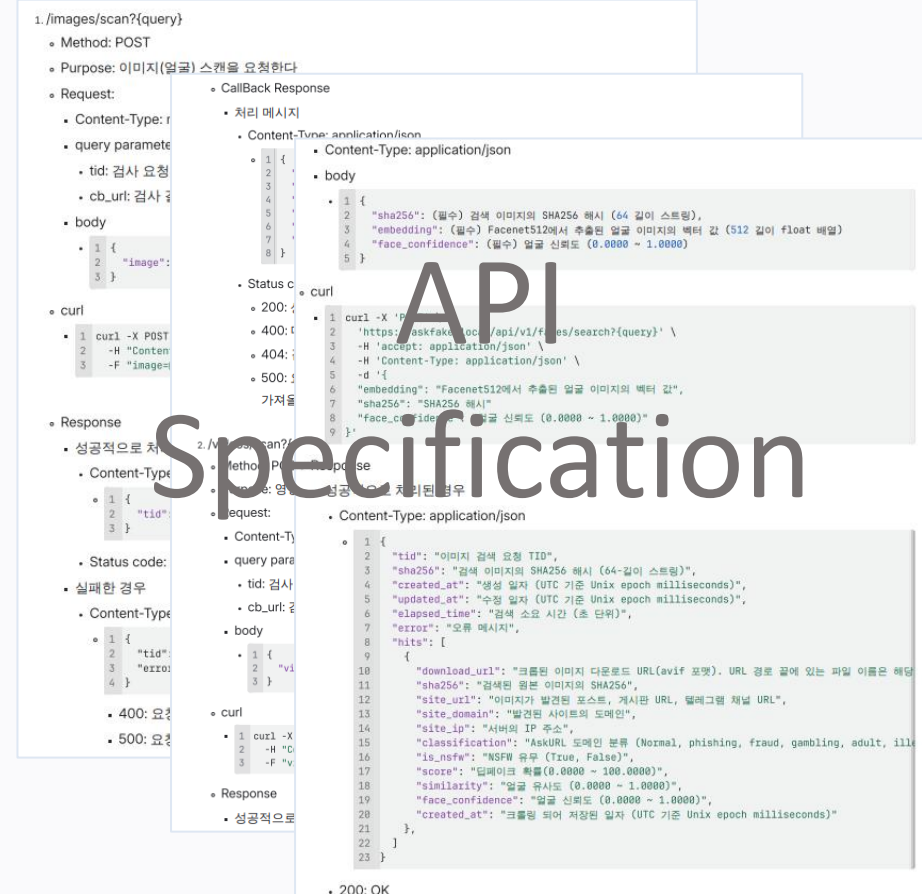
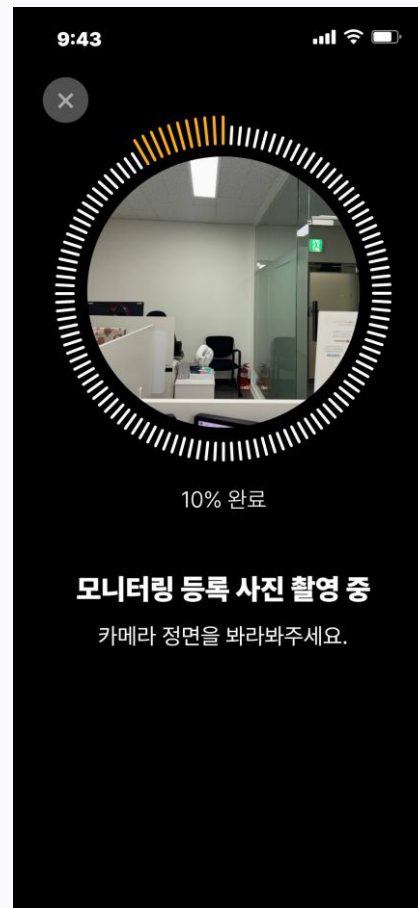
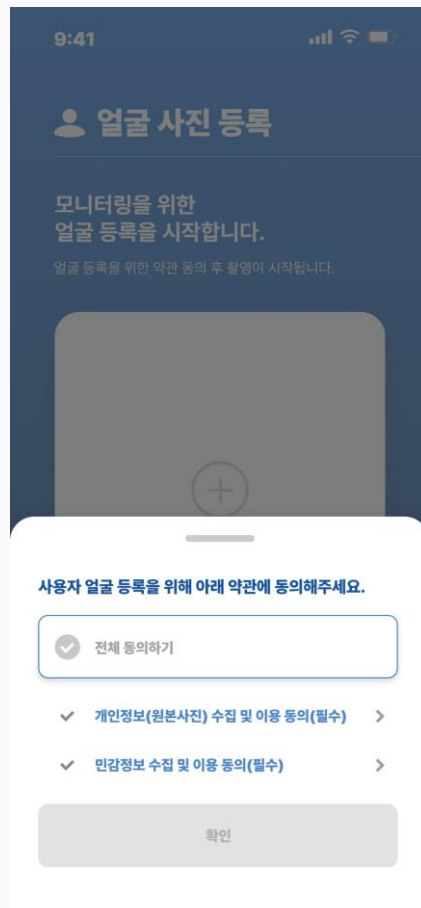
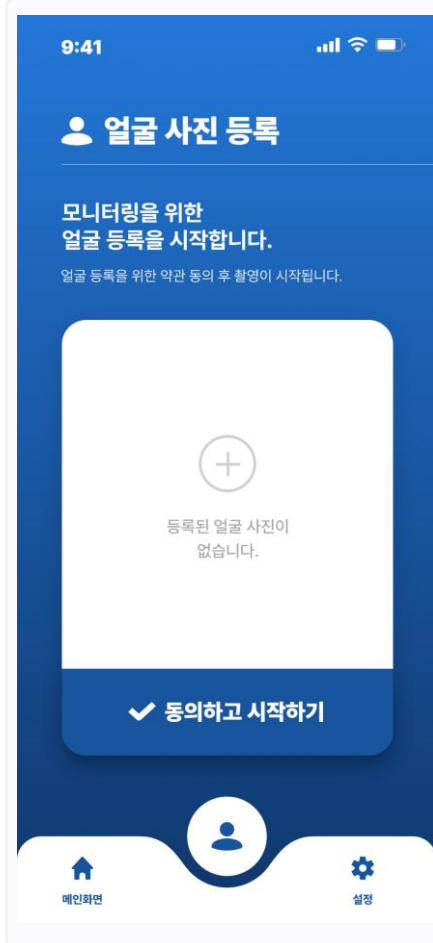
NSFW 탐지 결과 예시(추론 결과 시각화)





## 이용자 서비스 UI 및 API

- ✓ 등록된 이용자 얼굴 기준으로 웹/커뮤니티/성인사이트 등에서 유사 얼굴을 검색하여 딥페이크 여부 판별
- ✓ 얼굴 검색 및 딥페이크 탐지 기능을 외부 서비스에서 연동·활용할 수 있도록 표준 인터페이스 제공





## 정량(성능)적 목표 달성 현황

- ✓ 제3자 공인기관 한국시험인증원(KOTCA)에서 시제품 검증(한국시험인증원 시험실)
- ✓ AI 기반 악성 딥페이크 탐지 및 대응 시스템 시제품은 5개 모든 시험 항목에서 목표(90%)를 초과 달성함

| 구분      | 성과지표           | 목표치 | 목표치 산출근거   | 한국시험인증원 (KOTCA) 성능 시험 결과 | 비고   |
|---------|----------------|-----|--|--------------------------|--|
| 정량 (성능) | Accuracy(정확도)  | 90% | • Accuracy = (탐지한 정상 및 딥페이크 콘텐츠 중 올바르게 분류된 비율)     | 93.13%                   | <p>AI 기반 악성 딥페이크 탐지 및 대응 시스템은 한국시험인증원 검증을 통해 <b>고성능·고신뢰도의 모델임을 공식적으로 증명</b></p> <div> <div> <p>G4B(www.g4b.go.kr)진위확인코드 : 3w85K0GBw+I=</p> </div> <div> <p>G4B(www.g4b.go.kr)진위확인코드 : 3w85K0GBw+I=</p> </div> </div> |
|         | Precision(정밀도) | 90% | • Precision = (실제 딥페이크로 탐지된 콘텐츠 중 진짜 딥페이크 콘텐츠의 비율) | 94.58%                   |  |
|         | Recall(재현율)    | 90% | • Recall = (전체 딥페이크 콘텐츠 중 탐지된 비율)                  | 92.74%                   |  |
|         | F1-score       | 90% | • 정밀도와 재현율의 조화 평균을 통한 성능 평가                        | 93.65%                   |  |
|         |                |     |  |                          |  |





## 정성적 목표 달성 현황

✓ **특허출원 완료:** 사용자 얼굴 기반 딥페이크 모니터링 및 탐지 방법 및 장치(METHOD AND APPARATUS FOR USER FACE-BASED DEEPFAKE MONITORING AND DETECTION)

| 구분 | 성과지표        | 목표치 | 목표치 산출 근거  | 달성 현황  |
|----|-------------|-----|--|--|
| 정성 | 특허 등 지재권 확보 | 1건  | <ul style="list-style-type: none"> <li>특허 등 출원 건수</li> </ul> | <p>사용자 얼굴 기반 딥페이크 모니터링 및 탐지 기술<br/>METHOD AND APPARATUS FOR USER FACE-BASED DEEPFAKE MONITORING AND DETECTION</p> <div> <div>2025-11-14</div> <div> <div>【서지사항】</div> <div> <div>【서류명】</div> <div>특허출원서</div> </div> <div> <div>【출원구분】</div> <div>특허출원</div> </div> <div> <div>【출원인】</div> <div>주식회사 누리랩</div> </div> <div> <div>【명칭】</div> <div>특허고객번호</div> <div>1-2017-024535-8</div> </div> <div> <div>【대리인】</div> <div>명칭</div> <div>특허법인시공</div> </div> <div> <div>【대리인번호】</div> <div>9-2023-100041-2</div> </div> <div> <div>【지정된변리사】</div> <div>조예찬, 이준호, 신진현</div> </div> <div> <div>【포괄위임등록번호】</div> <div>2023-069327-5</div> </div> <div> <div>【발명의 국문명칭】</div> <div>사용자 얼굴 기반 딥페이크 모니터링 및 탐지 방법 및 장치</div> </div> <div> <div>【발명의 영문명칭】</div> <div>METHOD AND APPARATUS FOR USER FACE-BASED DEEPFAKE MONITORING AND DETECTION</div> </div> <div> <div>【발명자】</div> <div>성명</div> <div>최원혁</div> </div> <div> <div>【성명】</div> <div>성명의 영문표기</div> <div>Wonhyuk Choi</div> </div> <div> <div>【국적】</div> <div>KR</div> </div> </div> </div> <div> <ul style="list-style-type: none"> <li>출원인: (주)누리랩</li> <li>출원일: 2025년 11월 14일</li> <li>청구항 수: 16항</li> </ul> </div> |



### 정성적 목표 달성 현황

- ✓ **MOU 체결 완료:** 누리랩 – NHN Service Technology Corp(大連慧搜網絡有限公司) 간 업무 제휴 협약서 체결(2025년 11월)
- ✓ **협약 목적:** AI 기반 딥페이크 탐지 서비스를 포함한 보안·위협 분석 솔루션의 중국 시장 내 상용화 및 확산을 공동으로 추진

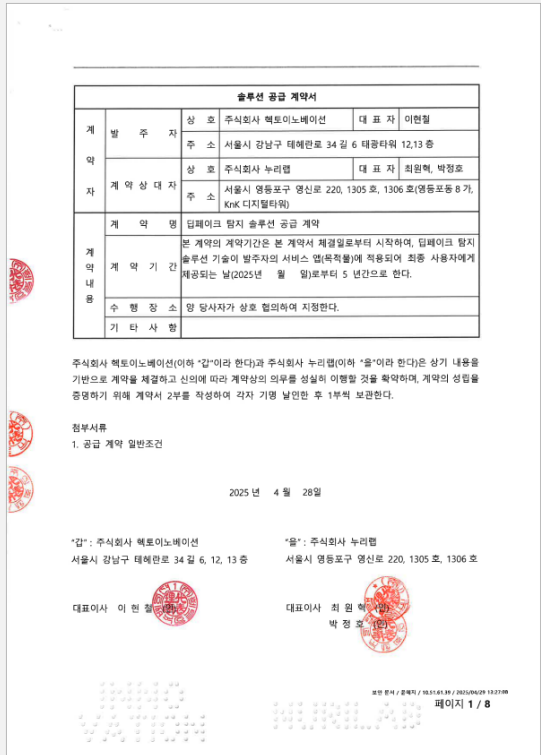
| 구분 | 성과지표   | 목표치 | 목표치 산출 근거   | 달성 현황  |
|----|--------|-----|-------------|--|
| 정성 | MOU 협약 | 1건  | • MOU 체결 건수 | <div> <div> <p>업무 제휴 협약서</p> <p>MEMORANDUM OF UNDERSTANDING</p> <p>주식회사 누리랩</p> </div> <div> <p>업무 제휴 협약서</p> <p>주식회사 누리랩</p> </div> </div> <div> <p>业务合作协议</p> <p>业务合作协议</p> <p>株式会社 NHN</p> </div> |



## 정성적 목표 달성 현황

- ✓ **솔루션 공급 계약 체결 완료:** 누리랩-헥토이노베이션 간 딥페이크 탐지 솔루션 공급 계약 체결
- ✓ **계약 목적:** 누리랩이 개발한 딥페이크 탐지 솔루션을 헥토이노베이션 서비스 앱에 적용 및 공급하기 위한 제반 사항 규정

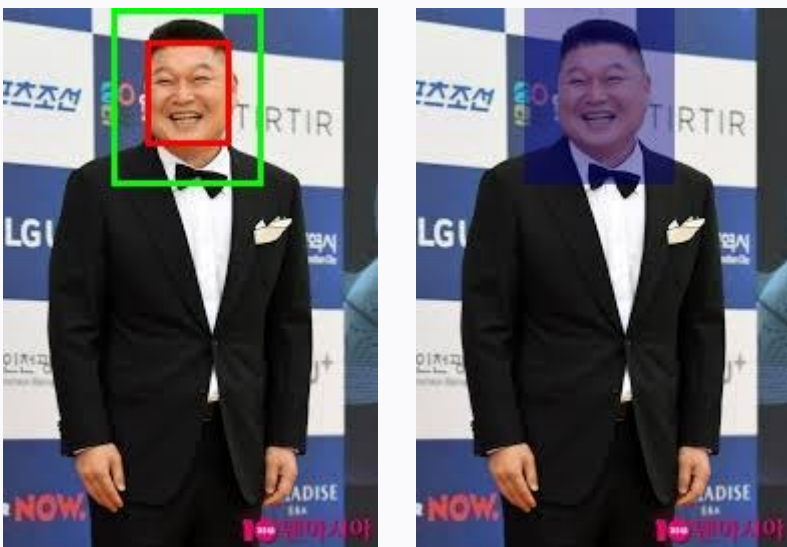
| 구분 | 성과지표     | 목표치 | 목표치 산출 근거  | 달성 현황  |
|----|----------|-----|------------|--|
| 정성 | 수요-공급 협업 | 1건  | • 공급 계약 건수 | <ul style="list-style-type: none"> <li>■ <b>계약명:</b> 딥페이크 탐지 솔루션 공급 계약</li> <li>■ <b>계약 당사자:</b> 누리랩, 헥토이노베이션</li> <li>■ <b>솔루션 공급 기능</b> <ul style="list-style-type: none"> <li>- 딥페이크 합성물 탐지</li> <li>- 딥페이크 유출 모니터링</li> <li>- 얼굴 검색 및 식별</li> <li>- 검색 및 탐지 결과 안내</li> </ul> </li> <li>■ <b>대상 시장:</b> 국내 이동통신사 및 금융권</li> <li>■ <b>계약금액:</b> 5억원</li> </ul> |





## 모니터링 대상 얼굴 식별 및 검색 - 모니터링 대상 얼굴 식별 및 학습 모듈

- ✓ 입력된 사진의 얼굴 영역을 식별하는 화면
- ✓ 식별된 얼굴 영역을 추출하여 학습하며, 각 학습 모델 버전별 평가지표 기록



```
def save_heatmap_on_image(orig_img_rgb: np.ndarray, cam_hw01: np.ndarray,
                          box_xyxy: Tuple[int, int, int, int], save_path: str, alpha: float = 0.45):
    """Overlay resized CAM onto the given bbox region and save."""
    import cv2
    from PIL import Image as PILImage # avoid shadowing

    x1, y1, x2, y2 = [int(v) for v in box_xyxy]
    H, W = orig_img_rgb.shape[:2]
    x1 = max(0, min(x1, W-1)); x2 = max(0, min(x2, W-1))
    y1 = max(0, min(y1, H-1)); y2 = max(0, min(y2, H-1))
    if x2 <= x1 or y2 <= y1:
        return

    cam = (cam_hw01 * 255.0).astype(np.uint8)
    cam = cv2.resize(cam, (x2 - x1, y2 - y1), interpolation=cv2.INTER_LINEAR)

    cam_color = cv2.cvtColor(cam, cv2.COLORMAP_JET)[:, :, :1] # BGR->RGB
    overlay = orig_img_rgb.copy()
    roi = overlay[y1:y2, x1:x2]
    blended = (alpha * cam_color + (1 - alpha) * roi).astype(np.uint8)
    overlay[y1:y2, x1:x2] = blended

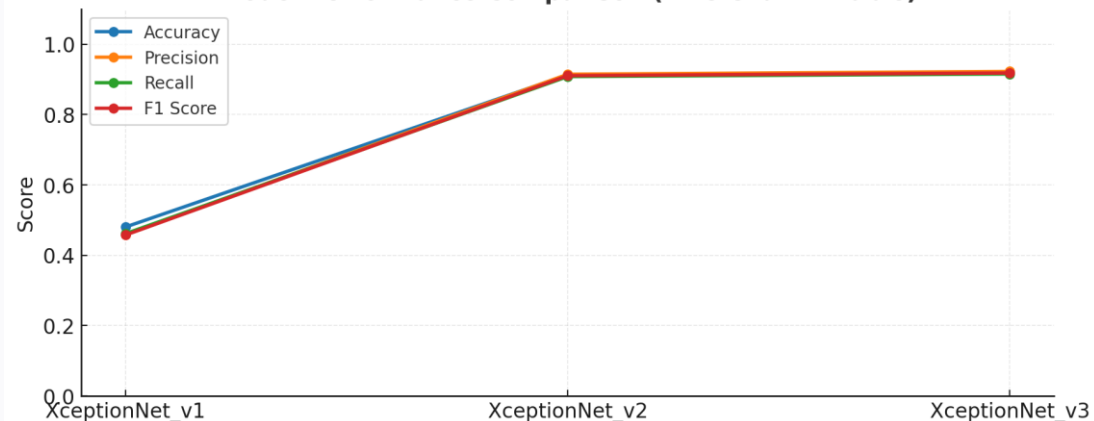
    PILImage.fromarray(overlay).save(save_path, quality=95)

# pick best by confidence
best_i = idxs[np.argmax(confis[idxs])]
x1, y1, x2, y2 = [int(v) for v in xyxy[best_i]]
x1 = max(0, min(x1, w-1)); x2 = max(0, min(x2, w-1))
y1 = max(0, min(y1, h-1)); y2 = max(0, min(y2, h-1))
if x2 <= x1 or y2 <= y1:
    return None, None, info

# padded bbox
bw, bh = x2 - x1, y2 - y1
pad = self.padding_ratio
x1p = max(0, int(x1 - pad * bw)); y1p = max(0, int(y1 - pad * bh))
x2p = min(w, int(x2 + pad * bw)); y2p = min(h, int(y2 + pad * bh))
if x2p <= x1p or y2p <= y1p:
    return None, None, info

face_raw = img_rgb[y1:y2, x1:x2]
face_padded = img_rgb[y1p:y2p, x1p:x2p]
```

Model Performance Comparison (Line Chart + Table)



| Model          | Accuracy | Precision | Recall | F1 Score |
|----------------|----------|-----------|--------|----------|
| XceptionNet_v1 | 0.4801   | 0.4586    | 0.4612 | 0.4572   |
| XceptionNet_v2 | 0.9127   | 0.9145    | 0.9079 | 0.9105   |
| XceptionNet_v3 | 0.9201   | 0.9225    | 0.9152 | 0.9181   |



## 모니터링 대상 얼굴 식별 및 검색 - 얼굴 이미지 검색 모듈

- ✓ 얼굴이 포함된 검색할 사진 선택
- ✓ 검색 결과의 download\_url 링크를 통해 유사한 얼굴이 존재하는 사진을 받을 수 있음. (각 검색 된 사진별 Deepfake 여부, 원본 얼굴과 얼굴 유사도 등이 검색 결과에 포함되어 있음)

ASKFAKE Deepfake Scan Service (Development SERVER)  
모바일 데스크탑 지원 · 화상 파일은 로컬에서 선택 후 업로드됩니다

AI-powered by NURILAB

**Video Scan (Callback)**

Video File

파일 선택 선택된 파일 없음

Scan Video

TID: -

모바일에서는 비디오만 선택 항목에 표시됩니다.

**Image Scan (Callback)**

Image File

파일 선택 선택된 파일 없음

Scan Image

TID: -

키메라 바로 올림(기기/브라우저에 따라 다를 수 있음)

**Face: Extract & Search (동기)**

Face Image

파일 선택 선택된 파일 없음

Extract & Search

↓

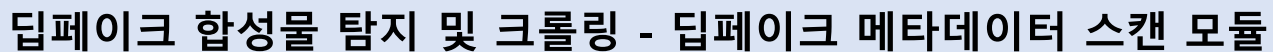
```

{
  "classification": "telegram",
  "created_at": 1757680383936,
  "download_url": "https://search.askfake.com/api/v1/faces/media/eadd1eb736fe",
  "face_confidence": 0.87,
  "is_nsfw": false,
  "phash": "",
  "score": 49.9082,
  "sha256": "eadd1eb736fe8e21a2de417648a56cfb795c2055ec06d837a8e14f4c16624c8a",
  "similarity": 0.85787445,
  "site_domain": "telegram.org",
  "site_ip": "",
  "site_url": "https://t.me/+7oH5OjHkhe3m9mV1"
},
{
  "classification": "telegram",
  "created_at": 1757602261700,
  "download_url": "https://search.askfake.com/api/v1/faces/media/eadd1eb736fe",
  "face_confidence": 0.87,
  "is_nsfw": false,
  "phash": "",
  "score": 72.2616,
  "sha256": "eadd1eb736fe8e21a2de417648a56cfb795c2055ec06d837a8e14f4c16624c8a",
  "similarity": 0.85787445,
  "site_domain": "telegram.org",
  "site_ip": "",
  "site_url": "https://t.me/+0K0MgdLQbgQpJNzg1"
},
{
  "classification": "telegram",
  "created_at": 1757602261700,
  "download_url": "https://search.askfake.com/api/v1/faces/media/eadd1eb736fe",
  "face_confidence": 0.87,
  "is_nsfw": false,
  "phash": "",
  "score": 72.2616,
  "sha256": "eadd1eb736fe8e21a2de417648a56cfb795c2055ec06d837a8e14f4c16624c8a",
  "similarity": 0.8544887,
  "site_domain": "telegram.org",

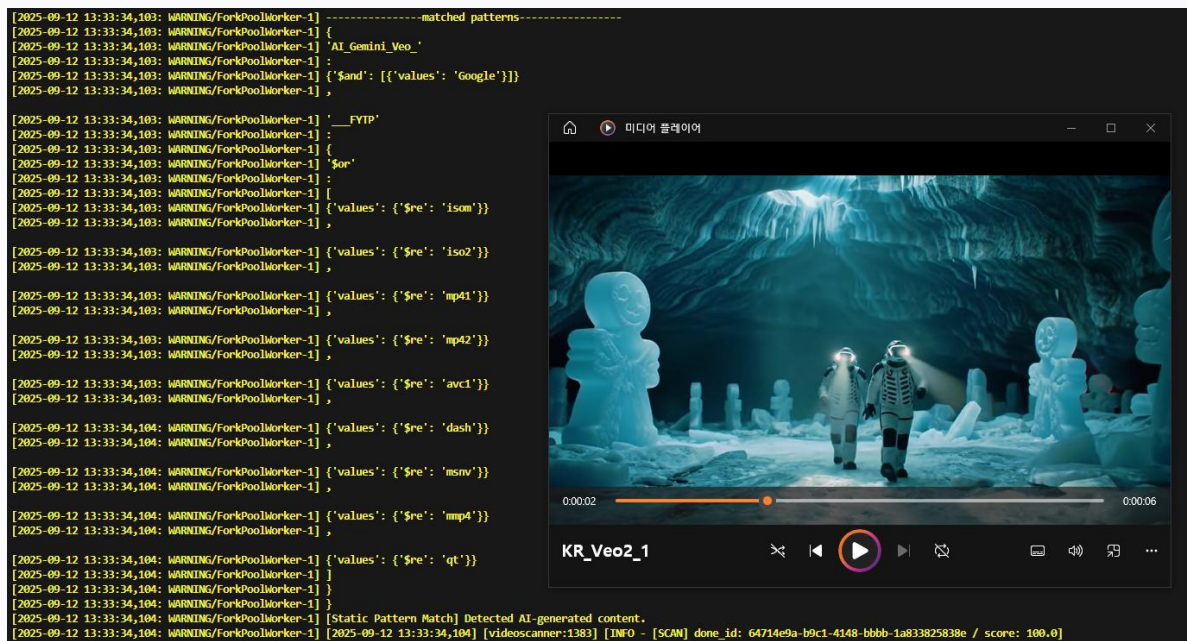
```

↓





- ✓ 이미지에 대한 메타데이터 탐지 결과
- ✓ 동영상에 대한 메타데이터 탐지 결과





## 딥페이크 합성물 탐지 및 크롤링 - 딥페이크 합성물 AI 탐지 모듈

- ✓ Video 및 이미지에 대한 탐지 모델이며, 파일을 선택하면 업로드 후 탐지
- ✓ Metadata 및 C2PA 등 헤더 정보를 우선 탐색 하므로 딥페이크 생성물에 대한 높은 탐지율을 보임

**ASKFAKE Deepfake Scan Service** (Development SERVER)  
모바일 데스크탑 지원 · 의심 파일은 로컬에서 선택 후 업로드됩니다

AI-powered by NURILAB

**Video Scan (Callback)**

Video File

파일 선택 선택된 파일 없음

**Scan Video**

TID: -  
모바일에서는 비디오만 선택 목록에 표시됩니다.

**Face: Extract & Search (동기)**

Face Image

파일 선택 선택된 파일 없음

**Extract & Search**

**ASKFAKE Deepfake Scan Service** (Development SERVER)  
모바일 데스크탑 지원 · 의심 파일은 로컬에서 선택 후 업로드됩니다

AI-powered by NURILAB

**Image Scan (Callback)**

Image File

파일 선택 선택된 파일 없음

**Scan Image**

TID: -  
카메라 바로 열람(기기)/브라우저에 따라 다를 수 있음).

**Video Scan (Callback)**

Video File

파일 선택 kling\_20250909\_Image\_to\_Video\_\_\_\_\_2400\_0 (1).mp4

**Scan Video**

TID: c6fbfa25b68c4b76be48bfd0aebae173  
모바일에서는 비디오만 선택 목록에 표시됩니다.

```
{
  "elapsed_time": "0.9309361490013544",
  "score": "100.0000",
  "sha256": "8045cb55a36b0796492918f6b62d2808758f79722484ebaa5921c28670894f69",
  "status_code": "200",
  "tid": "c6fbfa25b68c4b76be48bfd0aebae173"
}
```

**Image Scan (Callback)**

Image File

파일 선택 0005244448\_001\_20250902144216090.jpg

**Scan Image**

TID: c1a80eb37c154f849a47dd0342783aa8  
카메라 바로 열람(기기)/브라우저에 따라 다를 수 있음).

```
{
  "elapsed_time": "1.3032221750036115",
  "score": "0.2428",
  "sha256": "f6caeec00eeb90f0c2c8b5d949de056c532aa88a068fb850f51a58231a05d265",
  "status_code": "200",
  "tid": "c1a80eb37c154f849a47dd0342783aa8"
}
```



## 딥페이크 합성물 탐지 및 크롤링 - 딥페이크 합성물 AI 탐지 모듈

- ✓ Video 및 이미지에 대한 탐지 모델이며, 파일을 선택하면 업로드 후 탐지
- ✓ Metadata 및 C2PA 등 헤더 정보를 우선 탐색 하므로 딥페이크 생성물에 대한 높은 탐지율을 보임

### Image File

파일 선택 fake\_0013.jpg

Scan Image

TID: d9fffc8701604325bf30bc1980a251e4

카메라 바로 열람(기기/브라우저에 따라 다를 수 있음).

```
{
  "elapsed_time": "0.31879525631666183",
  "heatmap": "rendered image (hidden)",
  "score": "99.9982",
  "sha256":
    "be5498a20586e5f539d3615866851007a9dd606bbcc9cc61278cecbc1ce09d90"
,
  "status_code": "200",
  "tid": "d9fffc8701604325bf30bc1980a251e4"
}
```

- heatmap image -



딥페이크  
이미지

### Image File

파일 선택 test02.png

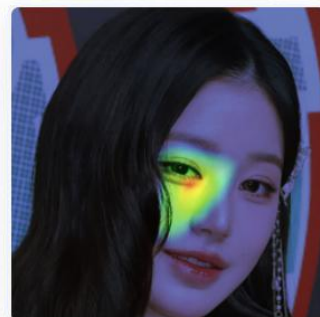
Scan Image

TID: faa7414597a9476fb3b61f4fce618877

카메라 바로 열람(기기/브라우저에 따라 다를 수 있음).

```
{
  "elapsed_time": "0.5947003327310085",
  "heatmap": "rendered image (hidden)",
  "score": "1.5045",
  "sha256":
    "4b105b240ef0db359e5c84c763f2fa0f6a46f784f1f1a6cd1d7d1195f57d167"
,
  "status_code": "200",
  "tid": "faa7414597a9476fb3b61f4fce618877"
}
```

- heatmap image -



정상  
이미지

### Video File

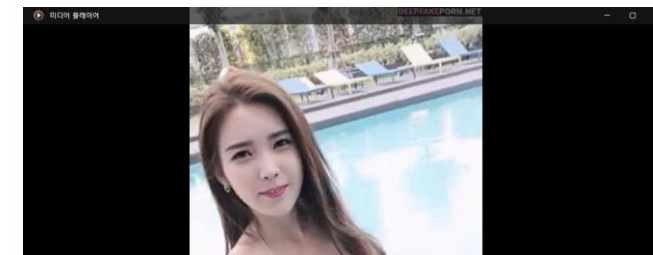
파일 선택 ( ) IU likes to pick up guys near pools - solo face swap.mp4

Scan Video

TID: 1b21614f6ebe4561b210c4d55a3faed6

모바일에서는 비디오만 선택 목록에 표시됩니다.

```
{
  "elapsed_time": "4.454337882809341",
  "score": "87.6775",
  "sha256":
    "9872fb722b0210ac6e939178fb51611430e2ee01fb04a6a7282b1e42f1e87e1a"
,
  "status_code": "200",
  "tid": "1b21614f6ebe4561b210c4d55a3faed6"
}
```



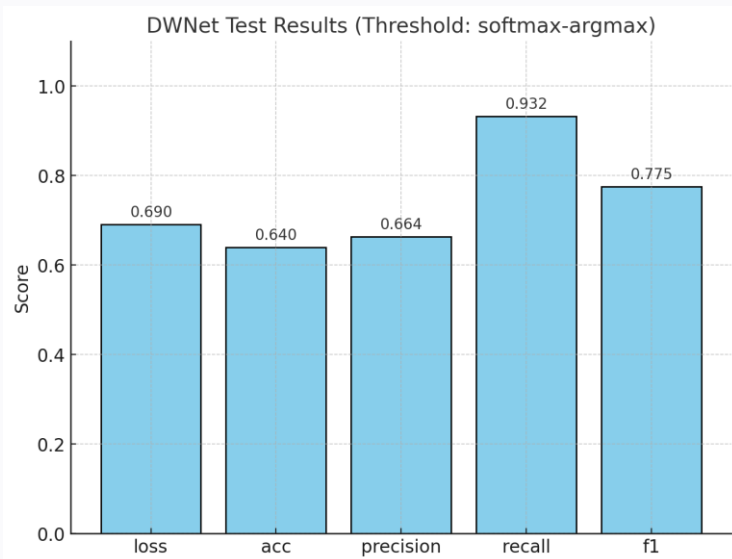
딥페이크  
동영상



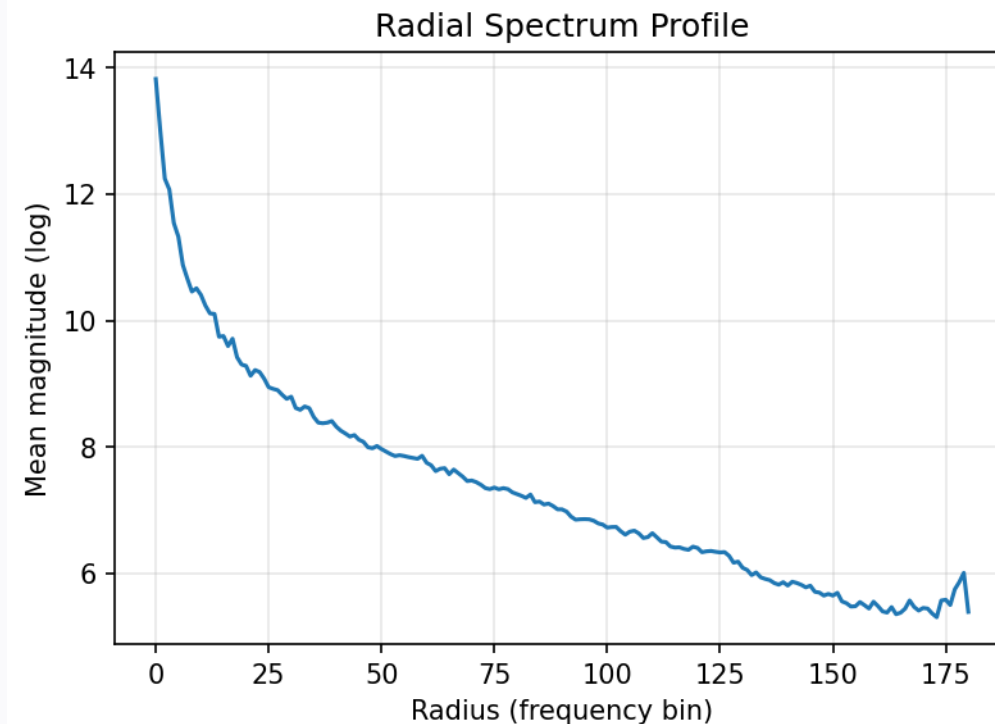


## 딥페이크 합성물 탐지 및 크롤링 - 인공지능 생성물 워터마크 검출 모듈

- ✓ 인공지능 생성물 워터마크 검출 모듈은 DWNNet 모델을 사용하여 학습 진행 중
- ✓ 현재는 모든 비가시성 워터마크 삽입 방식을 통합 학습하고 있으며, 향후 각 워터마크 삽입 방식을 유형별로 구분해 개별 학습을 진행할 예정



```
def forward(self, x):  
    spatial_feat = self.spatial_path(x)  
    freq_input = self.wavelet(x)  
    freq_input = self.wavelet_attn(freq_input)  
    freq_feat = self.freq_conv(freq_input)  
  
    if spatial_feat[2:] != freq_feat.shape[2:]:  
        freq_feat = F.interpolate(freq_feat, size = spatial_feat.shape[2:], mode = 'bilinear', align_corners = False)  
    # Feature Fusion  
    feat = torch.cat([spatial_feat, freq_feat], dim = 1)  
    feat = self.cbam(feat)  
    feat = self.fusion(feat)  
  
    out = self.pool(feat)  
    return self.classifier(out)
```





## 딥페이크 합성물 탐지 및 크롤링 - 모니터링 및 탐지 대상 크롤러

- ✓ 텔레그램 초대 URL 입력 후 크롤링이 완료되면 status\_code 200 으로 콜백 호출
- ✓ 이후 얼굴을 검색하면 해당 텔레그램을 통해 크롤링 된 사진 이미지가 검색



### Telegram Crawl (Callback)

#### Telegram Invite URL

<https://t.me/+PEd-dhGblzowNjA1>

Crawl

TID: -

초대 링크(공개/비공개)에 따라 접근 가능 범위가 달라질 수 있습니다.

### Telegram Crawl (Callback)

#### Telegram Invite URL

<https://t.me/+PEd-dhGblzowNjA1>

Crawl

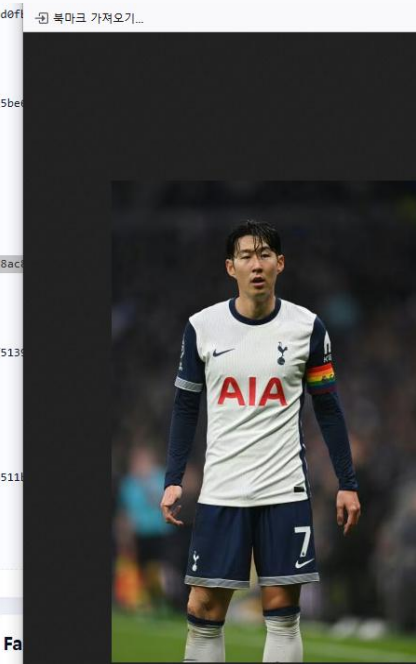
TID: b4c75cfa4d1048c08579048b68aa5e3c

초대 링크(공개/비공개)에 따라 접근 가능 범위가 달라질 수 있습니다.

```
{
  "channel_id": "2684292953",
  "status_code": "200",
  "tid": "b4c75cfa4d1048c08579048b68aa5e3c"
}
```

```
{
  "download_url": "https://search.askfake.com/api/v1/faces/media/76ced0fb1cd1300f8542c767c04520637cae9aac359c2b422bada3f35be",
  "face_confidence": 0.89,
  "is_nsfw": false,
  "phash": "",
  "score": 8.7012,
  "sha256": "76ced0fb1cd1300f8542c767c04520637cae9aac359c2b422bada3f35be",
  "similarity": 0.9911426,
  "site_domain": "telegram.org",
  "site_ip": "",
  "site_url": "https://t.me/+PEd-dhGblzowNjA1"
},
{
  "classification": "telegram",
  "created_at": 1757683762165,
  "download_url": "https://search.askfake.com/api/v1/faces/media/35e98ac82a6ab16867a3c3514e837191bbf77e0920ffd5effc241c3f513",
  "face_confidence": 0.86,
  "is_nsfw": false,
  "phash": "",
  "score": 10.7212,
  "sha256": "35e98ac82a6ab16867a3c3514e837191bbf77e0920ffd5effc241c3f513",
  "similarity": 0.8566345,
  "site_domain": "telegram.org",
  "site_ip": "",
  "site_url": "https://t.me/+PEd-dhGblzowNjA1"
},
{
  "sha256": "29bc8d29a5fcf4b6187d449ad936bb34bc78475ef9ca7b9af4ae4361e89b511",
  "tid": "d8ec051aca0a442b81ff4d2efea0a789",
  "updated_at": 1757651444855
}
```

ce: Extract Only



4

## 향후 계획 및 목표

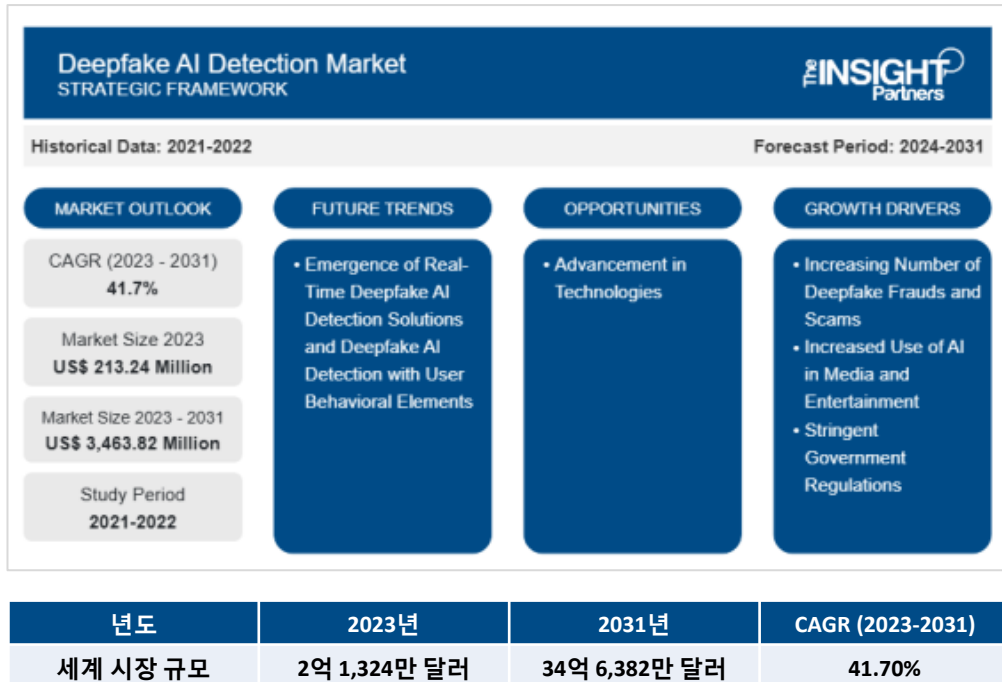
- 01 국내외 현황 조사
- 02 투자 및 생산 계획
- 03 국내외 사업화 전략



## 'AI 기반 악성 딥페이크 탐지 및 대응 시스템' 국내외 현황 조사

- ✓ 글로벌 딥페이크 탐지 시장은 2023년 2억 1,324만 달러에서 2031년 34억 6,382만 달러로 글로벌 CAGR (2023-2031) 41.7% 규모로 성장 전망
- ✓ 국내 시장: 「딥페이크 성범죄 대응 강화 방안」 등 법 규제 준수 → 정부 및 관계 기관, 소셜 네트워크 플랫폼 기업 등에서의 수요 크게 증가 예상

### 글로벌 시장 현황 및 전망



### 국내 시장 현황 및 전망

| 기관/업체명          | 사업명                        | 사업 내용   |
|-----------------|----------------------------|---|
| 경찰청             | 「딥페이크 탐지 소프트웨어」 고도화        | <ul style="list-style-type: none"> <li>• 딥페이크 탐지 소프트웨어 고도화</li> <li>- 2025년 예산 5억원 편성</li> <li>• 허위 영상물 범죄에 대응하기 위해 이미지합성기술(딥페이크) 방식의 허위 영상물뿐만 아니라 최신 인공지능(AI) 기법을 활용한 허위영상물까지 탐지할 수 있도록 '소프트웨어 고도화'</li> </ul>   |
|                 | 허위 조작 콘텐츠 복합 탐지 기술 개발(R&D) | <ul style="list-style-type: none"> <li>• 이미지합성기술(딥페이크) 및 딥보이스 등 진화하는 허위 콘텐츠 범죄에 대응하기 위해 딥러닝에 기반한 '허위 조작 콘텐츠 복합 탐지 기술 개발(R&amp;D)' 추진</li> <li>- 2025년에 27억원 투입</li> <li>- 2025~2027년 총 91억원 배정</li> </ul>   |
| 과학기술정보통신부       | 딥페이크 범죄 예방 연구개발(R&D) 과제    | <ul style="list-style-type: none"> <li>• 적대적 생성 신경망(GAN) 방식의 딥페이크 탐지 고도화 및 생성 억제 기술 개발: 예산 10억 원</li> <li>• 자가진화형 딥페이크 탐지 기술 개발: 예산 10억 원</li> </ul>  |
| 방송통신심의위원회       | AI기반 자동 식별·채증 시스템 구축       | <ul style="list-style-type: none"> <li>• 「딥페이크 성범죄 대응 강화 방안」 '2. 플랫폼 책임성 제고' - '플랫폼 사업자 의무 강화' 과제에 'AI기반 자동 식별·채증 시스템 구축' 예정</li> </ul>   |
| 여성가족부           | 디지털성범죄피해자 지원센터 예산          | <ul style="list-style-type: none"> <li>• 2025년도 디지털성범죄피해자지원센터 예산 32억6900만원으로 확정</li> </ul>  |
| 서울시             | 서울 디지털성범죄 안심지원센터           | <ul style="list-style-type: none"> <li>• 서울시 여성가족실 주관 딥페이크 성착취물 피해자 지원</li> <li>• 딥페이크 성착취물 유포 모니터링 · 삭제, 수사, 의료, 심리상담, 법률 지원 등</li> <li>• 수요예보: [여성가족실]서울디지털성범죄안심지원센터 업무시스템 및 AI기반 삭제지원시스템 통합 (계속사업, 유지보수)</li> <li>- 예산: 3억 5천만원</li> </ul>   |
| 소셜 네트워크 서비스 플랫폼 | 딥페이크 성착취물 상시 모니터링 및 탐지 서비스 | <ul style="list-style-type: none"> <li>• 「딥페이크 성범죄 대응 강화 방안」 '2. 플랫폼 책임성 제고' - '플랫폼 사업자 의무 강화' 과제 준수하기 위한 '딥페이크 성착취물 등 악성 딥페이크 합성물에 대한 상시 모니터링 및 탐지, 대응 서비스' 도입 사업 예상</li> <li>- '청소년유해물 제공 · 매개', '불법촬영물 등' 유통방지 의무 등을 강제</li> <li>- 불법촬영물 등 유통방지 노력, 신고·삭제 처리 결과, 책임자 지정 내용 등 포함한 투명서보고서 내실화</li> <li>- 성범죄를 게재자 제재</li> <li>- 모니터링 강화 등</li> </ul> |



## 'AI 기반 악성 딥페이크 탐지 및 대응 시스템' 투자 및 생산 계획

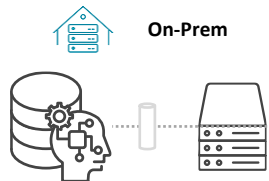
- ✓ 딥페이크 탐지 및 대응 서비스는 기술 개발 완료 후 즉시 시장에 출시 가능한 아키텍처로 구성
- ✓ AI 기반 탐지 성능 고도화와 글로벌 서비스 확장을 위한 클라우드 인프라 최적화를 병행하여, 다양한 시장 요구에 맞춘 SaaS 및 온프레미스(On-Premise) 형태로 제품을 생산할 계획

### 제품 생산 계획

#### 온프레미스(On-Premise) 제품

- 공공기관 및 특수 목적 기업을 위해 개발되는 온프레미스 제품은 사용자의 내부 인프라에 직접 설치되어 운영되는 형태로, 높은 보안성과 데이터 관리 요구 충족

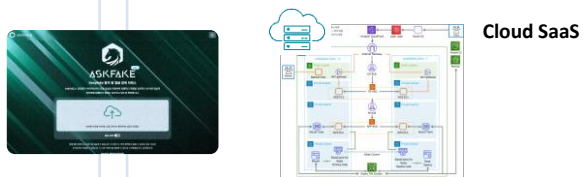
1. SaaS 제품 출시 후 온프레미스 제품 개발 및 생산
2. 공공기관, 금융기관, 보안이 중요한 대기업을 대상으로 맞춤형 솔루션 제공
3. 보안 인증 및 데이터 보호 요구사항을 충족하는 커스터마이징 지원



#### SaaS(서비스형 소프트웨어) 제품

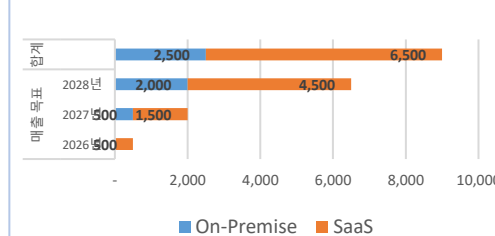
- 클라우드 기반의 딥페이크 탐지 및 분석 플랫폼을 SaaS 형태로 제공, 국내외 다양한 채널에서 유입되는 딥페이크 콘텐츠를 자동으로 탐지 및 분석하는 서비스

1. 기술 개발 완료 후 즉시 SaaS 서비스 론칭
2. 클라우드 기반으로 글로벌 사용자를 대상으로 서비스 확장
3. API 및 SDK 제공으로 다양한 기업 시스템과 연동 가능
4. 자동화된 탐지 및 대응 프로세스 구축으로 운영 효율성 극대화

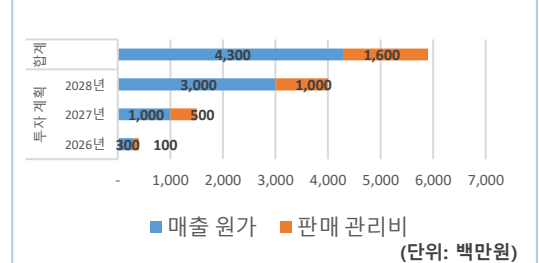


### 투자 계획 및 매출 목표

#### 매출 목표



#### 투자 계획



| 구분         | 매출 목표                |                      |                      | 합계    |
|------------|----------------------|----------------------|----------------------|-------|
|            | 2026년<br>(개발종료 후 1년) | 2027년<br>(개발종료 후 2년) | 2028년<br>(개발종료 후 3년) |       |
| On-Premise | 0                    | 500                  | 2,000                | 2,500 |
| SaaS       | 500                  | 1,500                | 4,500                | 6,500 |
| 합계         | 500                  | 2,000                | 6,500                | 9,000 |

| 구분     | 투자 계획                |                      |                      | 합계    |
|--------|----------------------|----------------------|----------------------|-------|
|        | 2026년<br>(개발종료 후 1년) | 2027년<br>(개발종료 후 2년) | 2028년<br>(개발종료 후 3년) |       |
| 매출 원가  | 300                  | 1,000                | 3,000                | 4,300 |
| 판매 관리비 | 100                  | 500                  | 1,000                | 1,600 |
| 합계     | 400                  | 1,500                | 4,000                | 5,900 |



## 'AI 기반 악성 딥페이크 탐지 및 대응 시스템' 국내외 사업화 전략

- ✓ 전 세계적으로 증가하는 딥페이크 범죄와 강화되는 규제에 따라 높은 글로벌 수요를 기대
- ✓ 클라우드 기반 SaaS 모델과 다국어 지원을 통해 다양한 국가에서 확장 가능하며, AI 기반 고도화된 탐지 기술로 글로벌 경쟁력 확보
- ✓ 현지화 전략과 국제 보안 인증을 통해 공공기관 및 기업 시장에서 신뢰 구축

### 2026년(국내)

| 구분                | 추진 계획  |
|-------------------|--|
| 프로모션으로 레퍼런스 확보    | <ul style="list-style-type: none"> <li>국내 공공기관 및 대기업과의 협업을 통한 성공 사례 확보</li> <li>초기 고객 인프라를 활용하여 프로모션 및 할인 정책으로 사용자 확대</li> </ul>   |
| 유통채널 확보를 통한 판로 개척 | <ul style="list-style-type: none"> <li>유통사 및 해외 리셀러와의 파트너십 체결로 글로벌 판매 전략 수립</li> <li>글로벌 SaaS 플랫폼(AWS Marketplace 등) 입점</li> </ul>   |
| 온라인 홍보            | <ul style="list-style-type: none"> <li>SNS, 웹사이트, 블로그 등을 통한 글로벌 온라인 홍보 강화</li> <li>딥페이크 분석 보고서 및 각종 통계 자료를 게재하여 품질의 우수성 홍보</li> <li>글로벌 인플루언서 및 보안 전문가와 협업하여 신뢰성 있는 온라인 홍보 진행</li> </ul> |
| 국내 시장 점유율 확대      | <ul style="list-style-type: none"> <li>국내 공공기관 및 대기업 중심으로 시장 점유율 확대</li> <li>기술력 우위를 바탕으로 국내외 시장 점유율 확대</li> <li>정부 지원 사업 및 네트워크를 활용한 현지 진출</li> </ul>                                   |

### 2027년(해외)

| 구분         | 추진 계획   |
|------------|---|
| 해외 파트너사 발굴 | <ul style="list-style-type: none"> <li>해외 거점을 기반으로 파트너사 발굴</li> <li>정부 지원 사업을 통해 해외 네트워크를 활용한 파트너 확보</li> </ul>                           |
| 동남아 시장 진출  | <ul style="list-style-type: none"> <li>태국, 인도네시아, 필리핀, 베트남 등 동남아 시장 개척</li> <li>현지 정부기관과의 협업 및 지역 보안 기업과의 파트너십 체결</li> </ul>              |
| 일본 시장 진출   | <ul style="list-style-type: none"> <li>KDDI, 알서포트, NTT 도코모 등과의 협력을 통한 일본 시장 진출</li> <li>일본의 사이버 보안 법규 대응 및 특수 규제에 맞춘 제품 커스터마이징</li> </ul> |



### 2028년(해외)

| 구분          | 추진 계획   |
|-------------|---|
| 북미권 파트너사 발굴 | <ul style="list-style-type: none"> <li>미국, 캐나다 정부 지원 사업 및 현지 네트워크 활용하여 파트너사 확보</li> </ul>   |
| 북미권 시장 진출   | <ul style="list-style-type: none"> <li>현지 보안 박람회 및 컨퍼런스(RSA Conference 등) 참가로 인지도 확대</li> <li>프로모션 및 공동 마케팅을 통한 북미 시장 점유율 확대</li> <li>연방 정부 기관과의 협업으로 공공부문 시장 진입</li> <li>헥토이노베이션, 폴라리스오피스 등과의 협력을 통한 북미 시장 개척</li> </ul> |
| 유럽 시장 진출    | <ul style="list-style-type: none"> <li>북미 시장 판로 개척 이후 유럽 시장으로 확장</li> <li>영국, 독일, 프랑스 등 주요 국가의 법적 요구사항에 맞춘 제품 현지화</li> <li>유럽 내 공공기관 및 금융기관 대상 맞춤형 솔루션 제공</li> </ul>  |





# 감사합니다.

CONFIDENTIALITY NOTICE: The contents of this message are intended solely for the addressee(s) and may contain confidential and/or privileged information and may be legally protected from disclosure. If you are not the intended recipient of this message or its agent, or if this message has been addressed to you in error, please immediately alert the sender and delete this message. If you are not the intended recipient, you are hereby notified that any use, dissemination, copying, or storage of this message is strictly prohibited.