

2025년 새싹 해커톤(SeSAC Hackathon) AI 서비스 기획서

팀명	채용사기가멈추지않는탓일까..?T.T
팀 구성원 성명	류현정, 신수지, 박민서

1 AI 서비스 명칭

고해성사

고: 수익
해: 외취업
성: 매매
사: 기

2 활용 인공지능 학습용 데이터

	활용 데이터명	분야	출처
1	해외 취업 사기 공고 글	분류	웹 크롤링
2	텍스트 윤리 검증 데이터	검증	AI hub

3 핵심내용

- 서비스 : **고해성사** - 해외 취업 사기 위험도 분석 솔루션
- 제안배경 :
최근 사회 문제로 떠오르는 해외 취업 사기
그렇듯한 직책과 숙식 제공, 월 수백만 원 상당의 고수익을 미끼로 내건 해외 취업 사기 문제가 주목 받고 있습니다. 점차 교묘해지는 수법에, 이를 미리 알고 예방할 수 있는 방법이 있을지 고민해본 결과 위 서비스를 제안하게 되었습니다.
- 핵심기능 :
고수익 해외 취업 사기 위험도 분석
취업 사기로 의심되는 채용 공고 내용이나 모집책과의 연락 내용을 캡처한 뒤 업로드하면 키워드와 맥락 등을 파악하여 사용자에게 사기 위험도를 알립니다.

○ 기대효과 :

구직자 피해 예방

잠재적인 취업 사기 위험에 대한 구직자들의 경각심을 높이고
안전한 결정을 내릴 수 있도록 하여 피해를 예방합니다.

4 제안배경 및 목적

○ 제안배경 :

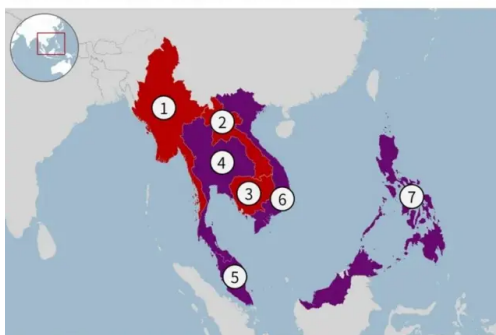
최근 캄보디아를 비롯한 동남아시아 지역에서 발생하는 **해외 취업 사기**가 우리 사회의 심각한 문제로 대두되고 있습니다. 대한민국 정부뿐만 아니라 다수의 언론에서도 해당 문제에 관심을 기울이며 대책 마련을 촉구하고 있는 실정입니다.

지난 7월, 현지 박람회 참관을 목적으로 캄보디아로 출국했던 한 대학생이 3주 뒤 고문으로 인해 싸늘한 주검으로 발견되는 사건이 있었습니다. 또한 온라인 구인 글을 통해 캄보디아로 출국한 한국인이 범죄 가담을 거부하자 100여 일간 감금된 채 쇠파이프와 전기충격기로 구타를 당했다는 충격적인 사실이 밝혀진 바 있습니다. 이러한 문제의 심각성은 신고 건수에서도 명확하게 드러나고 있습니다. 지난 5년간 캄보디아 내에서 한국인을 대상으로 한 취업 사기 및 감금 신고는 2021년 4건, 2022년 1건이었습니다. 그러나 2023년 13건, 2024년 220건에 이어 2025년 8월까지의 무려 330건으로 폭발적인 증가세를 보이고 있습니다.

이와 같이 한국인을 대상으로 한 취업 사기 강력 범죄는 캄보디아에 국한되지 않고, 미얀마, 라오스 등 동남아시아 여러 지역에서 지속적으로 발생하고 있기에 해외 구직자들의 각별한 주의가 절실히 요구됩니다.

범죄조직들의 사기 수법은 점차 교묘해지고 있습니다. 이들은 'IT 전문가', '해외 마케터', '온라인 상담원' 등 그럴듯한 직책을 내세워 구직자들에게 신뢰를 얻은 후, 숙식 제공과 고수익을 미끼로 유혹하고 있습니다. 이처럼 지능화된 수법으로 인해 구직자가 해당 구인 공고의 위험도를 자체적으로 파악하는 데 어려움이 있습니다. 따라서 **해외 구인 글의 위험도를 판단하여 잠재적 피해를 예방할 수 있는 서비스**의 필요성이 절감되고 있습니다.

동남아시아 내 온라인 사기 범죄 중심지



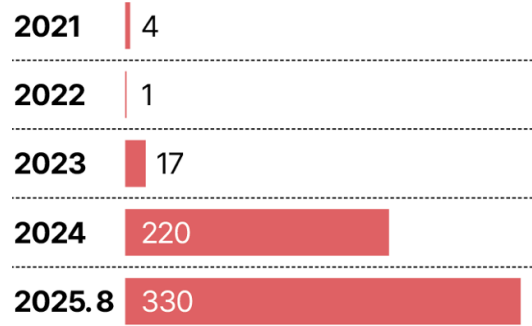
- 주요 인신매매 중심 국가
① 미얀마 ② 라오스 ③ 캄보디아
■ 그외 중심 국가
④ 태국 ⑤ 말레이시아 ⑥ 베트남
⑦ 필리핀

출처: 국제형사경찰기구(인터폴)

BBC

최근 5년간 캄보디아 내 한국인 취업사기 후 감금 신고 현황

●자료 박찬대 더불어민주당 의원실·외교부



○ 목적 : 해외 구인 글의 위험도를 판단하여 잠재적 피해를 예방합니다.

5 세부내용

○ 활용 데이터 및 AI 모델 :

- AI 모델 개요 및 기능

주요 모델: 채용 사기 분류 모델

기능: 사용자가 제공하는 이미지를 분석하여 채용 사기 여부를 판별합니다.

단순 분류를 넘어, 추후 통계 및 분석 자료로도 활용할 수 있습니다.

기반 기술: KoELECTRA (base 또는 small 버전) 모델을 핵심으로 사용합니다.

- 데이터 구축 전략

수집: 웹 크롤링을 통해 대량의 텍스트 데이터를 확보한 뒤, 특징을 추출합니다.

증강: 1차 증강 - 추출된 특징을 기반으로 LLM(거대 언어 모델)을 활용하여
텍스트 데이터를 증강합니다.

2차 증강 - 한국어 텍스트를 영어로 번역한 후 다시 한국어로 재번역하는 기법을
통해 추가 데이터를 확보하여 모델의 학습 다양성을 높입니다.

- 적용 기술 및 구현 방식

기반 기술 스택: Regex, KoELECTRA, HuggingFace Trainer, PyTorch

향후 선정: LLM 및 OCR(광학 문자 인식) 기술은 추후 최적의 솔루션으로 선정하여
적용할 계획입니다.

서비스 운영 흐름:

1. 사용자가 채용 공고 이미지를 업로드합니다.
2. 서버에서 OCR 기술을 통해 해당 이미지를 텍스트로 변환합니다.
3. 변환된 텍스트는 KoELECTRA 모델에 입력되어 채용 사기 여부를 예측합니다.
4. 예측 결과는 '사기' 또는 '정상' 여부와 함께 정량화된 정확도(위험도 %)로
사용자에게 제공됩니다.

- 구현 가능성

KoELECTRA 모델: 한국어 소량 데이터 환경에서도 효율적인 학습이 가능하여
모델 구현에 적합합니다.

OCR 기술: 한국어를 지원하는 다양한 오픈소스 솔루션이나 상용 API를 활용할 수
있어, 기술적 구현에 어려움이 없을 것으로 판단합니다.

○ 세부내용 :

- 서비스 아이디어 개요

궁극적 목표: 해외 취업을 희망하는 구직자들이 온라인 구인 공고의 잠재적 위험
요소를 사전에 정확히 인지하고, 이를 통해 취업 사기 피해를 예방하는
것을 궁극적인 목표로 합니다.

문제점: 교묘해지는 해외 취업 사기 수법과 구직자들의 정보 부족으로 인한 피해 증가

해결책: 인공지능 기반의 '해외 구인 공고 위험도 분석 서비스'를 제공하여,
구직자들이 게시글의 안전도를 정량적으로 판단할 수 있도록 지원합니다.

핵심 가치: 막연한 경고가 아니라 정확한 위험도(%)와 그 근거를 제시하여 구직자의
주체적인 의사결정을 돕는 예방 중심의 서비스입니다.

- 적용 기술

AI 모델 및 학습

- KoELECTRA: 한국어 텍스트 분석에 특화된 모델로, 채용 사기 분류 및 위험도 예측의 핵심 엔진
- 데이터 증강: 웹 크롤링 및 LLM, Back-translation 기법을 활용하여 학습 데이터 양과 다양성을 확보
- 프레임워크: 모델 개발 및 학습 관리를 위해 PyTorch와 HuggingFace Trainer 사용

데이터 처리

- OCR (광학 문자 인식): 사용자 이미지 속 텍스트를 인식하여 분석 가능한 형태로 변환
- Regex (정규 표현식): 텍스트 내 특정 패턴을 식별하고 추출하는 데 활용

애플리케이션 개발

- 프론트엔드: Next.js, TailwindCSS, Axios
- 백엔드: Express.js

- 서비스 방법

AI 기반 위험도 분석:

1. 공고 이미지 업로드: 사용자가 해외 구인 공고가 담긴 이미지를 서비스에 업로드
2. OCR 텍스트 변환: 업로드된 이미지는 서버에서 OCR 기술을 통해 텍스트 데이터로 정확하게 변환
3. AI 모델 분석: 변환된 텍스트는 KoELECTRA 기반의 채용 사기 분류 모델로 전달되어 심층적인 언어 분석 진행. 이 과정에서 텍스트를 어조, 특정 키워드 사용 여부, 비현실적인 조건 제시 등 다양한 특징 분석
4. 위험도(%) 및 근거 제시: 분석 결과를 바탕으로 해당 공고의 채용 사기 위험도를 퍼센트(%)로 표시하여 제공. 이와 함께 위험도를 산출하게 된 주요 근거를 상세히 함께 안내

자가진단 체크리스트 (부가기능):

역할: AI 분석 외에, 사용자가 직접 구인 공고의 위험 요소를 확인하고 즉각적인 피드백을 받을 수 있도록 돕는 보조 기능

1. 체크리스트 확인: 사용자는 서비스에서 미리 정의된 사기 의심 사례 항목들을 직접 확인하고 체크
2. 즉각적인 위험 피드백: 체크된 위험 요소의 개수에 따라 실시간으로 해당 공고의 대략적인 위험도 산정하여 사용자에게 즉각적으로 안내

- 서비스 아이디어의 창의성 및 구현 가능성

창의성

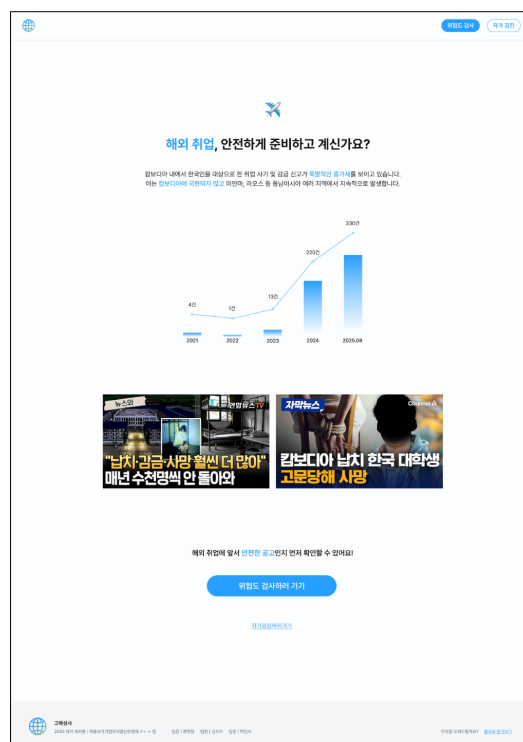
- 능동적인 피해 예방: 단순 정보 전달을 넘어, AI 분석을 통해 위험 요소 사전에 예측하고 경고하여 구직자가 능동적으로 피해를 예방할 수 있도록 돕습니다.
- 정량적 위험도 제시: '사기/정상' 여부만을 제공하는 것이 아니라 위험도를 퍼센트(%)로 수치화하여 직관적이고 객관적인 판단 기준을 제공한다는 점이 독창적입니다. 이는 구직자의 주체적인 판단 능력을 극대화합니다.
- AI 기술 접목: 최신 자연어 처리 기술(KoELECTRA, LLM)과 OCR을 결합하여 복잡한 한국어 채용 공고를 효과적으로 분석하는 데 창의성이 있습니다.

가능성

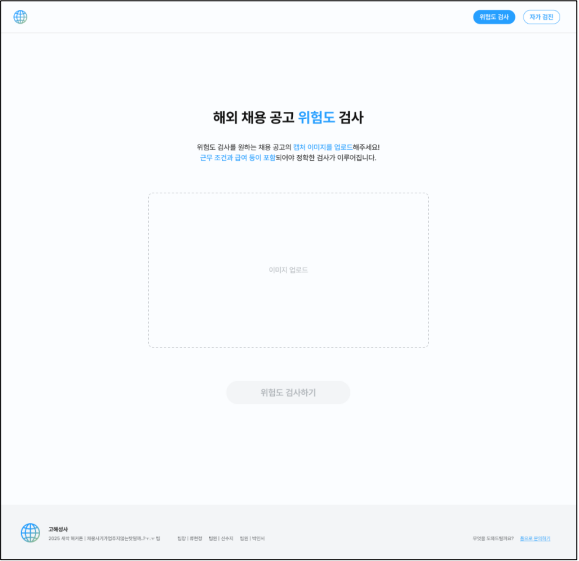
- 검증된 AI 모델 활용: KoELECTRA는 한국어 처리 분야에서 이미 성능이 검증된 모델로, 소량의 데이터로도 효과적인 학습이 가능하여 실제 서비스 개발에 유리합니다.
- 오픈소스 및 API 활용: OCR 기술은 시중에 다양한 고성능 오픈소스 라이브러리 및 상용 API가 존재하여, 기술 선정 및 도입이 용이합니다.
- 체계적인 데이터 전략: 웹 크롤링과 LLM 기반 데이터 증강, back-translation 등 다각적인 데이터 확보 및 가공 전략을 통해 모델 학습에 필요한 양질의 데이터를 효과적으로 구축할 수 있습니다.

○ 서비스의 예상 UI/UX 이미지 시각화

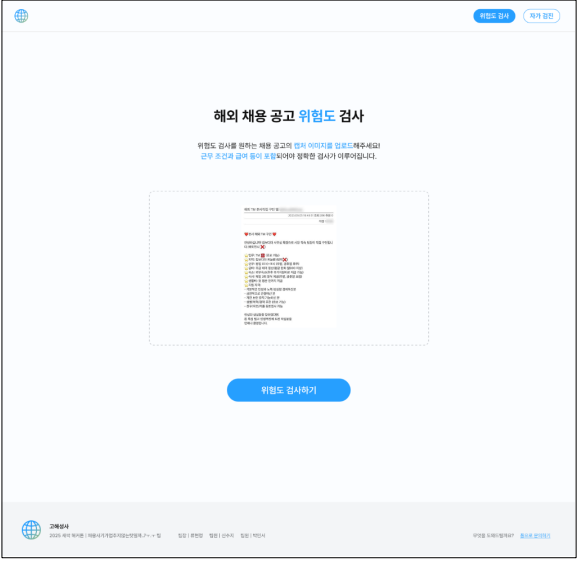
- 메인 화면



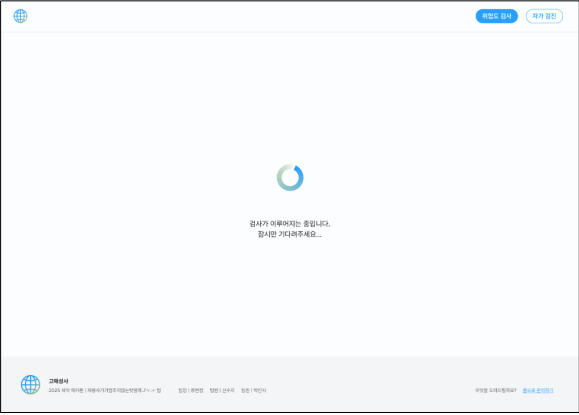
- 위험도 검사



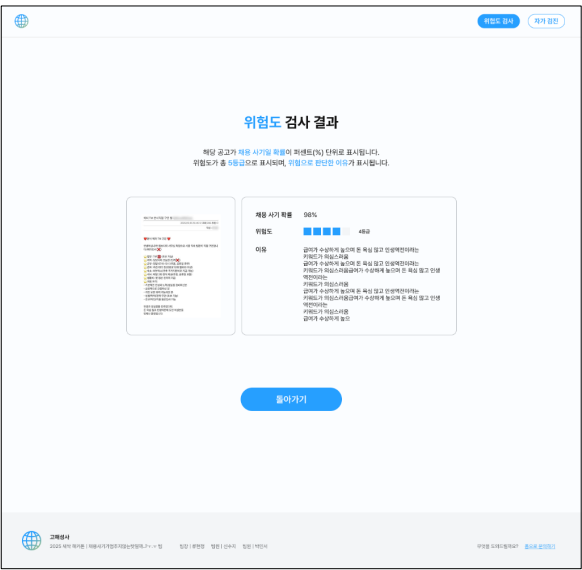
이미지 업로드 전



이미지 업로드 후



결과 대기



위험도 검사 결과

[illegible]

6개 이상 - 매우 위험한 공고

: 구직자들이 해외 구인 공고의 위험도를 정량적인 퍼센트(%) 형태로 명확하게 인지합니다. 이에 따라 해외 취업 결정 과정에서 더욱 신중하고 주체적인 의사결정을 내릴 수 있는 역량을 강화할 수 있습니다.