

「2025 IA x AI 해커톤」

개발 완료 보고서

팀 명 : 진실은 나의 빛

프로젝트명 : LLM 모델을 이용한 브라우저 확장
프로그램 기반 팩트 체크 서비스

※ 유의사항

1. 본 보고서의 내용은 최대 2 page이내로 작성 (**본 표지 제외**)
2. 보고서의 설명을 보충하기 위해 필요한 사진 또는 그래프 첨부 가능
3. 제출 서류는 일체 반환을 하지 않음
4. 제출 파일명 작성 요령
 - 파일명: [2025 IA x AI 해커톤]_팀명
5. 서체: 맑은고딕, 크기: 12p, 줄간격: 160%
6. 제출처: 깃허브에 업로드

「2025 IA x AI 해커톤」

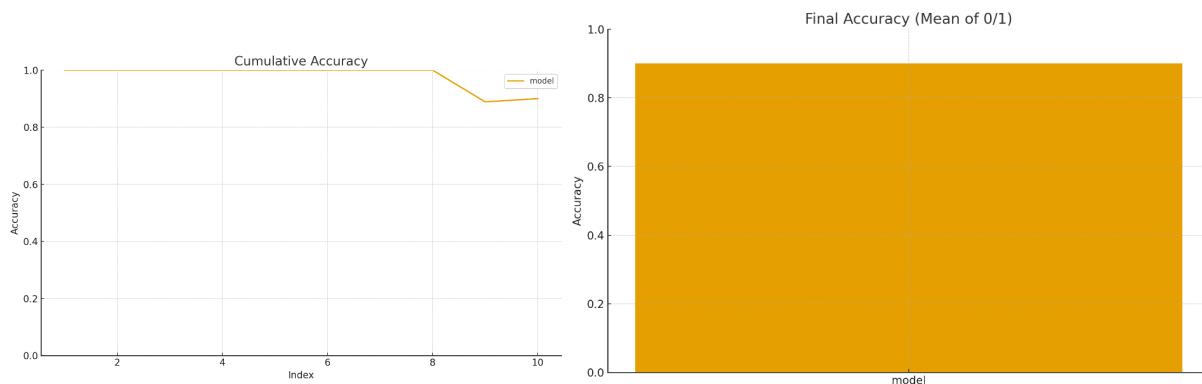
* 개발물의 내용을 2쪽 이내 분량으로 작성(분량 초과 시, 불임으로 작성)

프로젝트명	뉴스, 커뮤니티, SNS에서 허위정보를 즉시 검증할 수 있는 AI 팩트체크 서비스 도구 제공
프로젝트 목표	사용자가 웹을 탐색하는 흐름 안에서 즉시 텍스트, 이미지, 영상의 신뢰도를 점검해 거짓정보의 노출을 막고 근거가 담긴 결과를 한번에 공유할 수 있게 함
개발 환경	메인 프로덕트(크롬 익스텐션) : JS 프론트엔드(팩트체크 결과 공유하기) : JS, React.js, Vite 백엔드 : Python, FastAPI 인프라 : AWS EC2, RDS
구현 기능	PC 인터넷 브라우저 (크롬) 확장프로그램 기반 팩트체크 - (1) 선택한 텍스트의 진실일 확률, 확률의 근거, 상세 설명, 레퍼런스 url 제공, 결과 공유하기 제공 - (2) 선택한 이미지의 AI 생성형 이미지 여부 판별 - (3) 선택한 유튜브/인스타그램 영상의 AI 생성형 이미지 여부 및 거짓 주장 여부 판별 - (4) 백그라운드에서 현재 방문한 페이지에서 거짓 정보 포함 경고
코드	- backend/: FastAPI 기반 검증 API 서버로, 텍스트·이미지·영상 입력을 비동기 처리해 표준 JSON 결과를 생성·저장하며 PostgreSQL 연동, 입력 검증, CORS/로깅, 환경변수 기반 설정을 포함합니다. 영상은 yt-dlp 다운로드→프레임 샘플링→FFT/모션 분석→Whisper 전사까지 자동 파이프라인을 제공하고, 캐시·재시도·명확한 에러 응답으로 운영 안정성을 확보했습니다.
주요 설명	- extension/: Manifest V3 Chrome 확장. background.js가 컨텍스트 메뉴와 팩트체크 비즈니스 로직을 처리하고, content.js는 오버레이 UI, popup.js는 API URL 설정 및 기능 토글을 제공합니다. - frontend/: Vite + React 기반 랜딩페이지 및 팩트체크 결과공유 열람 가능. /share?id=<record_id> 접속 시 ShareResultPage.jsx가 백엔드 결과를 정규화해 ShareResult.jsx 컴포넌트로 렌더링합니다.
개발 내용	기능별 설명 - 출처 중심 텍스트 사실검증 파이프라인을 자체 설계해 핵심 주장

	<p>추출→근거 수집→정확도(%) 산출을 표준 JSON으로 정규화하고, 결과를 record_id와 함께 PostgreSQL에 영속화.</p> <ul style="list-style-type: none"> - 다중 키 라운드Robin·재시도·콘텐츠 필터 대응을 내장해 외부 응답을 안정화하고, 한국어 판단 사유·근거 URL만 엄격 스키마로 채택하여 저장. - 비동기 I/O 기반으로 응답 지연을 줄이고, 검증 결과·원문을 원자적으로 기록해 공유 링크로 재조회·재사용 가능한 데이터 자산화 완성. - Chrome Extension을 개발하여 브라우저에서 방문한 페이지의 모든 텍스트/이미지/영상 소스의 진실 판별 및 실시간 감지/경고, 감별 결과 공유 기능 등 쉬운 유저 경험을 제공 - 검증 가능한 출처 우선 전략으로 하위·불명확 응답을 배제하고, 정확도 산출 배경(한국어)을 강제해 설명가능성과 감사 추적성을 확보.
시연 영상 실행 파일	구글폼참조 구글폼참조
기타 (선택)	<p>팀 구성</p> <ul style="list-style-type: none"> - AI/백엔드: 한현진, 김지은, 임진호 - AI 모델 조사 및 Fact-check 프롬프팅 알고리즘 적용, LLM 기반 팩트체크 로직 개발, Web Search API 연동, 백엔드 아키텍처 설계 및 API 개발 - 프론트엔드/디자인: 전준성 - 크롬 확장 프로그램 개발, 서비스 소개 페이지 및 팩트체크 결과 공유 서비스 개발 - 데이터 리서치: 전체 팀원 - 텍스트/이미지/영상 사례 수집, 데이터셋 구축, 퍼포먼스 테스트 및 검증 - PM/문서화: 전준성, 임진호 - 프로젝트 기획, 회의록 및 기술 문서 작성, 데모 영상 제작 <p>협업 체계</p> <ul style="list-style-type: none"> - 커뮤니케이션: Slack을 통한 실시간 소통 - 개발: GitHub 기반 코드 관리 및 Pull Request 기반 코드 리뷰 (Copilot AI 를 사용한 코드리뷰도 포함해 리뷰 퀄리티 향상), Swagger 및 Notion을 통한 API 문서화 - 프로젝트 관리: Notion으로 회의록·기술 스펙 관리, GitHub Issues로 버그 및 개선사항 트래킹 - 데이터/테스트: Notion으로 데이터셋 공유 및 개발 담당자 자체 테스트

불임1

1) 텍스트 성공률



2) AI 생성 이미지 판별. gemini vs Huggingface deepfakedetect 모델 성능 비교(데이터 49건 비교(참28, 거짓21))

- gemini 성공률 약 90%, Huggingface deepfacedetect 성공률 약 60%

