Innovation Academy

「2025 IA x AI 해커톤」

개발 완료 보고서

팀 명 :	Cosmic Ocean
프로젝트명 :	AI Smell

「2025 IA x AI 해커톤」

※ 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

프로젝트명	AI Smell
프로젝트 목표	사용자가 검색 결과 페이지에서 웹사이트를 직접 방문하지 않아도 각 콘텐츠가 AI 생성 콘텐츠일 가능성을 한 눈에 확인할 수 있는 크롬 익스텐션 프로그램을 개발한다. 구체적으로, 검색 결과의 각 링크 옆에 AI 생성 가능성을 확률로 표시하여 사용자가 정보의 신뢰도를 빠르고 직관적으로 판단할 수 있도록 돕는다. 이를 통해 불필요한 탐색 시간을 절감하여 신뢰할 수 있는 정보를 효율적으로 찾아볼 수 있는 사용자 경험을 제공한다.
개발 환경	모델 학습 : AWS g6e.xlarger 인스턴스 (CPU : 4 vCPU, 메모리: 32 Gib, GPU : Nvidia L40S GPU) CUDA : 13.0 Fine-tuning : PEFT (LoRa), 16bit quantization
구현 기능	[검색 결과 필터링] • 구글 검색 페이지에서 각 사이트의 AI 생성 글을 판별 [설정 기능] • 측면 패널을 통해 AI 확률별로 모아주는 기능
코드 주요 설명	[크롬 확장 기능] • Server : 클라이언트로부터 URL을 받아서 해당 웹페이지 본문 텍스트 추출하여 AI 모델에 전달한다. 그리고 AI모델이 LLM이 작성한 글일 확률과 판단 근거를 json형태로 클라이언트에게 반환한다. • Chrome-extension : content.js가 구글 검색 결과 페이지의 링크들을 감지 후 서비스 워커인 background.js를 통해 각 링크의 URL을 서버로 전송한다. 이후, 서버로부터 받은 AI 확률 점수를 바탕으로 검색 결과 옆에 시각적 표시(태그)를 추가한다. panel.js로 사이드 패널에 검색 결과 링크들을 확률에 따라 필터링해서 보여준다.

	[Fine-Tuning] • LoRA 모델을 사용하여 카카오 'kanana-nano-2.1b-instruct' 모델의 Low-Rank 행렬만 학습을 수행함. 랭크 r=16, 계수 lora_alpha=64 적용한 결과 23,003,136 파라미터가 학습되었다. llama.cpp 을 사용하여 모델을 16bit 양자화하여 더 작고 빠르게 만들어 메모리 사용량을 줄이고 추론 속도를 향상시킬 수 있었다.
개발 내용	[주요 기능] • AI 작성 글을 판별하는 Chrome Extension 프로그램. 구글 검색 결과 페이지에서 링크 옆에 글이 AI로 생성되었을 확률을 표시한다. • 사람이 작성한 신뢰도 높은 글을 필터링해준다. [LLM모델 선정 및 파인튜닝] • AI, Human 판별에 있어 각 LLM 모델 별 응답 속도와 정확도를 비교하여 최적의 모델을 선정한다. • Prompting 기법과 Fine-Tuning을 통해 응답 속도와 정확도를 개선한다.
시연 영상	https://youtu.be/mzN_qvzKYS4
실행 파일 (선택)	구글폼 제출
기타 (선택)	김승렬 - 프로젝트 총괄 및 구조 설계 김동은, 양승조 - AI/인간 작성 문서 데이터셋 수집 후 파인튜닝 이지연 - UI/UX 및 크롬 익스텐션 기능 구현 정찬수 - Backend (서버 구현 및 테스트)

붙임1

모델	평균 응답 시간(초)	정확도(%)
local_ai_score	6.53	87.80
gpt_ai_score	9.52	90.00
gpt4_ai_score	**1.42**	92.00
gemini_ai_score	9.49	**93.87**

[모델별 응답 속도 및 정확도 비교]