17조
조은비, 박범창, 좌진우

한국어 기반 인공지능 생성 텍스트 탐지

5주차 - 유스케이스 다이어그램

Contents 목차

17조 '한국어 기반 인공지능 생성 텍스트 탐지' 프로젝트
5주차 유스케이스 다이어그램 과제 발표자료

Introduction 연구 배경

“왜 한국어 기반 가짜뉴스 탐지가 중요한가?”

- 가짜뉴스 확산 피해 급증
- 대부분 연구는 영어 중심
- 영어권 모델 → 한국어 적용 시 정확도 급감 (90% → 50~60%)

Introduction 연구 목적

한국어 기반 AI 생성 가짜뉴스 탐지 방법

- 가짜뉴스로 인한 피해 최소화

- 기사 신뢰도 향상

→ 가짜뉴스로 인한 피해 감소 및 건강한 정보 사회에 기여

Introduction 연구 질문 및 가설

연구 질문

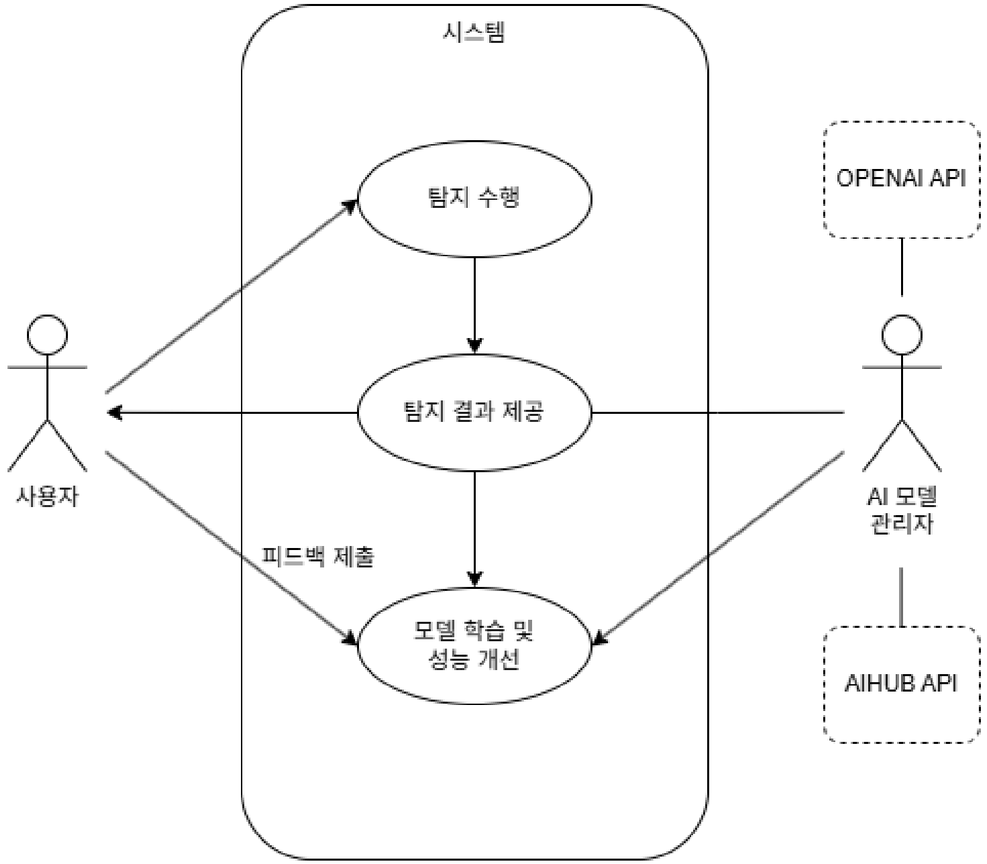
RQ1.
연구한 모델이 기존 연구모델에 비해 더 높은 정확도로 가짜뉴스를 판별하는가?

가설

H1.
AI 생성 기사를 만들기 위한 기초 기사 데이터는 AI로 생성된 기사가 아니다.

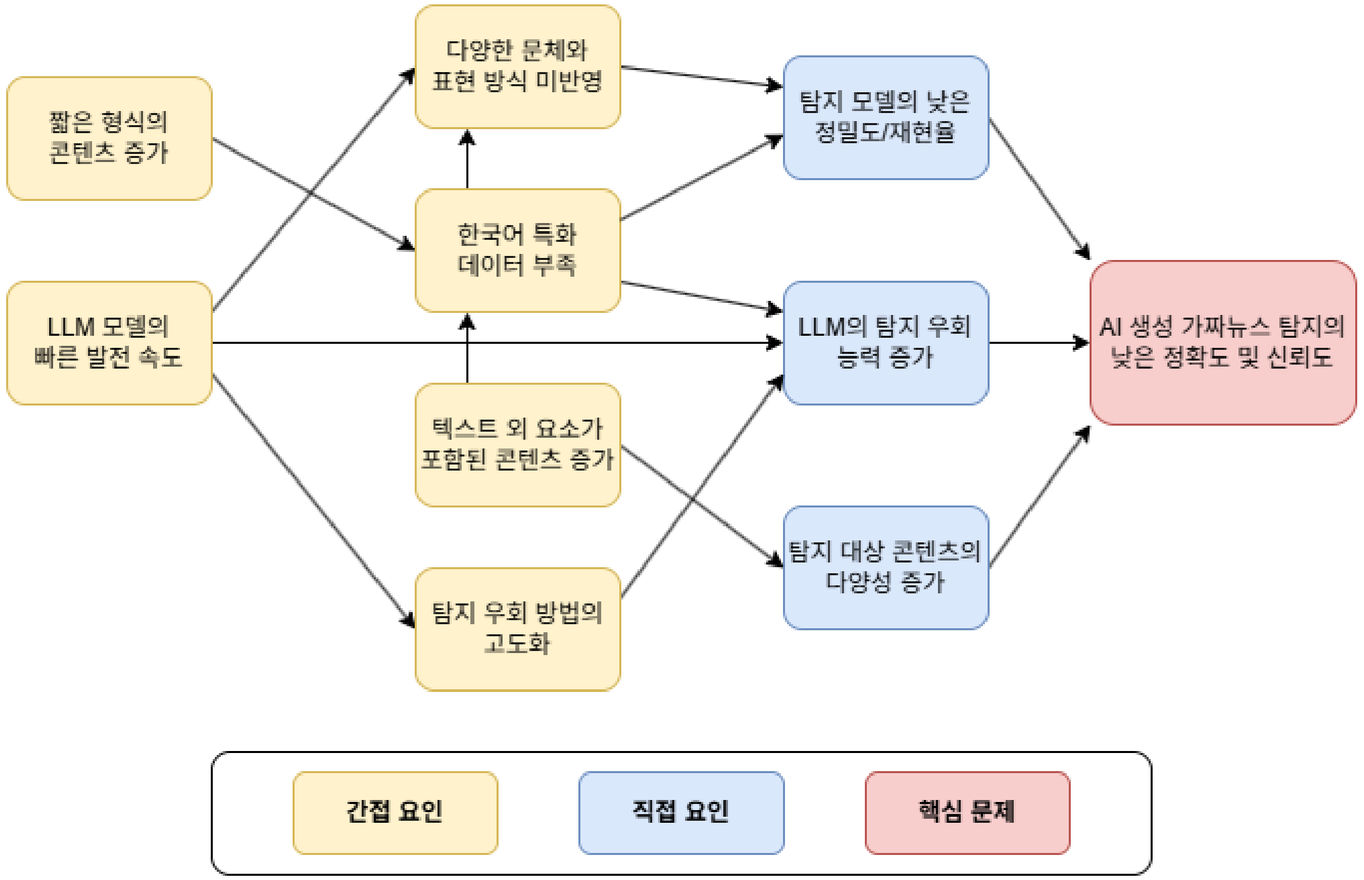
Usecase Diagram 소프트웨어의 사용 사례 Diagram

- Actor: 사용자, 관리자
- 기능: 뉴스 접근 → 탐지 수행 → 결과 확인 → 피드백



Usecase Diagram 문제 해결에 대한 사용 사례 Diagram

• 탐지 정확도 향상 → 정보 신뢰성 확보



Usecase Specification 소프트웨어 활용 사례

주요 Actor	AI 모델 관리자: 탐지 모델을 학습 및 배포하며 성능을 지속적으로 개선 일반 사용자: 뉴스 소비자. 탐지된 결과를 통해 뉴스의 신뢰도를 확인함
주요 기능 구성 요소	- 탐지 모델: KoBART / KoELECTRA 등 - AL 뉴스 생성기 (LLM 모델 기반) - 탐지 결과 시각화
입/출력 데이터	입력 데이터(결과): 실제 뉴스 기사 제목 및 내용 : AI를 통해 생성된 AI 뉴스 텍스트 출력 데이터(결과): AI 생성 가짜 뉴스 여부 (O, X, 모름) : 탐지 확률, 정밀도, 재현율 등의 결과 지표
데이터 Flow	① [뉴스 콘텐츠 접근] : 사용자가 뉴스에 접근 ② [탐지 수행] : 해당 뉴스를 탐지 ③ [탐지 결과 제공] : 화면에 다음과 같은 형태로 결과 제공 - ✔: 가짜 뉴스 아님 - ×: 가짜 뉴스 - ? : 모름 ④ [사용자 피드백] : 사용자가 해당 뉴스에 대한 의견 제출 가능
외부 시스템 연계	OPENAI API : GPT 모델을 사용하여 AI 생성 뉴스 텍스트를 생성 AIHUB API: 뉴스 기사 데이터셋 제공 그 외 다른 모델 사용 가능

- 구성 요소: KoBART / KoELECTRA, GPT 뉴스 생성기, 시각화 모듈
- 입력: 실제 뉴스, AI 생성 뉴스
- 출력: O, X, ? (탐지 결과 및 확률 등)
- 데이터 흐름: 뉴스 접근 → 탐지 → 결과 제공 → 피드백

Usecase Specification 문제 해결에 대한 사용 사례

핵심 문제	AI 생성 가짜 뉴스 탐지의 낮은 정확도 및 신뢰도
직접 요인	- 탐지 모델의 낮은 정밀도/재현율 - LLM의 탐지 우회 능력 증가 - 탐지 대상 콘텐츠의 다양성 증가
간접 요인	- LLM 모델의 빠른 발전 속도 - 짧은 형식의 콘텐츠 증가 - 한국어 특화 데이터 부족 - 탐지 우회 방식의 고도화 - 다양한 문체와 표현 방식 미반영 - 텍스트 외 요소가 포함된 콘텐츠 증가
활용 맥락	- 사회, 정치적 이슈에 대한 여론 조작 방지 - 언론사 및 플랫폼의 콘텐츠 검증 시스템 고도화 - 선거, 사회운동 등 민감한 시기에 가짜 뉴스 확산 억제 - SNS 및 커뮤니티 내 정보 확산의 신뢰도 향상 - 다양한 미디어 포맷에서의 위험 콘텐츠 사전 탐지

- 직접 요인: 모델 정밀도 낮음, 콘텐츠 다양성, 탐지 우회 증가
- 간접 요인: 한국어 데이터 부족, 짧은 콘텐츠 증가, 표현 다양성
- 여론 조작 방지 (선거/사회 이슈 등)
- 언론사/플랫폼의 콘텐츠 검증
- SNS/커뮤니티 내 정보 신뢰도 향상

Thank You