P R O J E C T D O C U M E N T

문제점 목록

Project Name	한국어 기반 인공지능 생성 텍스트 탐지

17 조

202202501 조은비
202002494 박범창
202002565 좌진우

지도교수: 이종률 교수님 (서명)

Document Revision History

Rev#	Date	Affected Section	Author
1	2025/03/29	문제점 목록 문서 작성	조은비

Table of Contents

1. 이해당사자(stakeholder)의 문제 이해

문제점	문제점 파악 방법	문제 상세 기술		
		이해당사자	고충/니즈	이유
한국어 기반 AI 생성 텍스트 탐지의 어려움	인터뷰 및 설문조사 (교수 1명, 관련 업계 종사자 1명, 일반 사용자 5명)	NLP 연구자, 교수, 일반 사용자 (SNS/뉴스 이용자)	AI가 생성한 글을 사람이 쓴 글과 구별하기 어려워 정확한 판별 도구 필요	영어 기반 도구는 한국어에서 탐지 정확도가 낮고, 실제 사용자들은 혼란을 겪고 있음
기존 AI 탐지 모델의 한계 (짧은 문장 탐지 실패)	AI 관련 논문 조사 및 사용자 대상 설문 응답 분석	언론 소비자 및 SNS 사용자	댓글이나 뉴스 제목 등 짧은 문장에서도 AI 탐지가 가능했으면 좋겠음	기존 탐지 모델은 긴 문장 위주로 학습되어 짧은 문장에서 탐지 성능이 매우 낮음
AI 생성 가짜 뉴스의 확산 위험	AI 뉴스 사례 분석 및 사용자 인터뷰	일반 뉴스 소비자, 커뮤니티 사용자	신뢰할 수 있는 정보와 가짜 정보를 구분하고 싶어함	가짜 뉴스가 AI로 쉽게 생성되어 실제 뉴스처럼 유포되고 있어 혼란 발생
생성형 AI 기술에 대한 사회적 불신	SNS 커뮤니티 사례 및 뉴스 댓글 분석	일반 사용자, 교육기관 관계자	AI 생성 정보의 출처나 신뢰도를 표시해주는 시스템 원함	사용자들은 정보의 진위를 판별하기 어려워 AI 콘텐츠에 대한 신뢰가 낮아지고 있음

한국어 기반 인공지능 생성 텍스트 탐지 기술의 필요성과 현실적인 문제점을 파악하기 위해 총 7명을 대상으로 인터뷰 및 설문조사를 진행하였다. 인터뷰는 AI 및 자연어처리(NLP) 분야의 전문가를 대상으로 진행하였으며, 설문조사는 최근 ChatGPT 및 SNS 기반 콘텐츠를 접한 일반 사용자 5명을 대상으로 실시하였다. 이를 통해 실제 사용자 및 전문가의 관점에서 공통적으로 나타나는 문제점과 요구사항을 도출하였다.

2.AS-IS 파악

현재 한국어 기반의 인공지능 생성 텍스트 탐지 시스템은 초기 단계에 머물러 있으며, 실제 적용에는 여러 가지 한계가 존재한다.

1. **기존 탐지 기술의 언어적 한계**
   현재 시중에 존재하는 AI 생성 텍스트 탐지 모델(DetectGPT, GPTZero, OpenAI Text Classifier 등)은 대부분 영어를 대상으로 개발되었으며, 영어 문장 구조와 어휘 특성을 기반으로 작동한다. 그러나 한국어는 교착어로서 어순이 유동적이고 조사와 어미 변화가 많기 때문에 동일한 구조를 적용했을 때 성능이 급격히 떨어진다.
2. **짧은 텍스트에서의 탐지 성능 저하**
   KoELECTRA나 KoBART 등 일부 한국어 특화 모델을 활용한 연구에서는 에세이나 기사 본문처럼 문장 길이가 긴 경우 비교적 높은 탐지 성능을 보였으나 뉴스 제목, 댓글, 짧은 SNS 포스트와 같은 짧은 문장에서는 탐지 정확도가 현저히 낮다.
3. **한국어 학습 데이터의 부족**
   대부분의 탐지 모델은 영어 기반의 XSum, SQuAD 등 대규모 정형 데이터셋을 사용하여 학습되었으며 한국어에 특화된 학습 데이터는 뉴스 기사에 한정되어 있다. 실제 사용 환경에서는 커뮤니티 글, 트위터, 블로그 등 다양한 형태의 텍스트가 존재하지만, 이들을 포함한 데이터셋은 거의 없는 상태이다.
4. **사용자 경험 및 실생활 적용 한계**
   일반 사용자들은 생성형 AI가 만든 콘텐츠를 자주 접하고 있지만 이를 즉각적으로 식별할 수는 없다 AI가 만든 콘텐츠를 사용자에게 알림으로 표시하거나 위험도를 평가해주는 서비스가 없어 정보 소비 과정에서 혼란을 겪고 있다. 특히, 정치적·사회적 이슈가 걸린 콘텐츠에서는 신뢰성 판단의 기준이 모호하여 잘못된 정보가 빠르게 확산되는 문제로 이어진다.

현재 시스템은 영어 중심, 긴 문장 중심, 정형화된 환경 중심으로 설계되어 있어 한국어 환경과 실제 사용 맥락에서는 정확성·실용성·적용성 모두에서 한계가 뚜렷하다. 이를 해결하기 위한 요구사항은 아래와 같다.

- 한국어 특화 탐지 모델 구축
- 짧은 문장에 강한 모델 구조 설계
- 실사용 환경을 반영한 데이터셋 수집 및 학습
- 탐지 결과를 사용자에게 직관적으로 제공하는 UI/UX 설계