

문제정의서 (연구개발계획서)

프로젝트명	정치, 경제, 시사분야 가짜 뉴스 판별 모델 개발
-------	-----------------------------

조	02 조
지도교수	정상근 교수님 (서명)
조원	201803851 양성욱 201802074 김재윤 202104524 박규수

1. 연구개발의 필요성

2024년 3월 28일 정부는 대국민 담화문에서 가짜뉴스와 허위선동을 중대 범죄로 엄정한 수사를 진행할 것으로 선언했다. 또한 2023년 8월 24일 오영국 대국민 담화에서도 국민을 가장 위협하는 것은 가짜뉴스와 허위선동이라는 말을 강조했다.

이와 같이 각종 뉴스 미디어 매체가 발전함에 따라서 추측과 과장이 섞인 뉴스가 나오고 이는 분명히 개인의 판단을 저해하고 국가의 위협이 되는 문제이다. 그러나 개인의 입장으로서 이러한 뉴스를 일일이 교차검증할 노력과 수고를 들이기 어려운 입장이다. 그렇기에 뉴스를 판단하기 위한 교차검증을 할 비용과 시간을 절감하는 방법이 필요하다.

현재 기술적으로 가짜 뉴스 문제를 해결하기 위해 연구개발된 가짜 뉴스 판별 ai모델들의 경우, 특정 사이트의 관리자를 대상으로 하여 가짜 뉴스를 ai가 탐지하고 관리자가 이를 배제하도록 돕는 방향으로 연구가 진행 중이다. 그 예시로 국내에서는 한국보건사회연구원에서 “기계학습 기반 보건복지분야 가짜뉴스(fake news) 탐지 방법 연구” 연구보고서가 있고, 국외에서는 최근 FOX NEWS에서 VERIFY라는 ai기반 툴이 있다.

그러나 정치, 경제, 시사 분야의 경우, 사람의 의견에 따라 정보의 참,거짓이 달라지기도 하여 일괄적으로 참,거짓을 나누기 어렵고, 또한 ai를 통한 배제가 정보와 의견에 대한 검열이 될 수도 있는 윤리적 문제가 있다. 그래서 관리자를 대상으로 하는 방식이 아닌 새로운 접근 방식을 가진 가짜 뉴스 판별 방안이 필요하다.

가짜 뉴스에 대응하는 새로운 방법으로 미디어 리터러시 교육을 통해 개인이 직접 뉴스를 판별하게 하는 방안이 주목되는바 개인이 직접 가짜 뉴스를 판별하는 것을 도와 기존 방식의 단점을 타파한 새로운 방식의 연구가 필요하다.

2. 연구개발의 목표 및 내용

-연구개발 목표

본 연구는 뉴스를 교차 검증할 시간이 부족한 사회 참여계층들로 하여금 정치, 경제, 시사 분야의 뉴스를 직접 교차 검증하는 시간을 줄일 수 있도록하고, 사안에 대한 의견을 가지도록 돕는 것을 궁극적인 목표로 한다.

-연구개발 내용 및 범위

첫 번째 기능으로 뉴스의 목적을 파악하기 위한 유형별 분석을 수행하는 예측 모델을 개발하고자 한다.

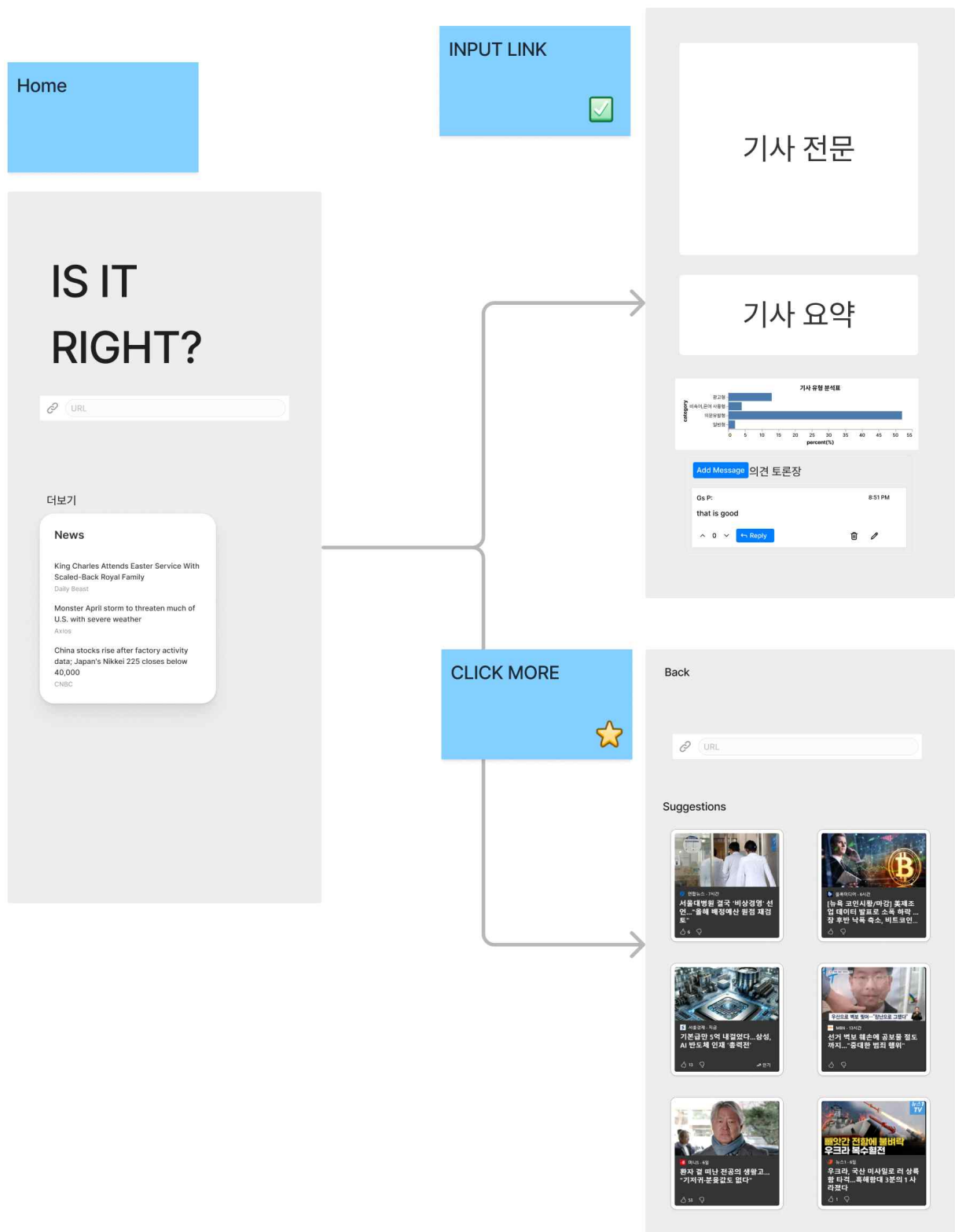
두 번째 기능으로 뉴스 기사 전문을 파악하는 시간을 단축하기 위해 뉴스의 내용을 분석하고 핵심내용을 요약해주는 텍스트 마이닝 모델을 개발하고자 한다.

부속 기능으로 유저들이 자신의 의견이나 검증 내용을 공유할 수 있는 토론 댓글창을 구현하고자 한다.

그리고 타인이 분석한 뉴스에 대해서도 추천해주는 기능을 구현하고자 한다.

-프로토타입 계획

아래와 같이 UI를 만들어 이해당사자에게 시나리오를 전달하고자 한다



3. 연구개발의 추진전략 및 방법

- 뉴스 유형별 예측 모델의 경우 AI-HUB의 ‘뉴스성 뉴스 판별’데이터 셋을 통해 LLM 모델을 사용하여 뉴스의 키워드와 문단을 요약한다.
- 자연어 처리 기술을 이용하여 유형별 결과를 얻어 classification을 수행한다.
- 웹 페이지의 경우 html, css, javascript 로 제작하여 분석할 뉴스의 링크를 검색할 창, 더보기를 통한 추천 뉴스, 요약한 내용 분석표, 그리고 토론 창을 제작할 계획이다.
- 웹 페이지의 back end의 경우 FAST API를 사용하여 구축할 것이다.

4. 연구개발 팀의 구성 및 프로젝트 추진 일정

팀 구성	구분	성명	참여내용 및 역할			
	팀장	양정욱	자연어 처리 및 모델 제작			
	팀원	김재운	데이터 셋 수집 및 전처리			
	팀원	박규수	웹 제작 및 DB구축			
추진 일정	세부내용		수행기간(월)			
			3	4	5	6
	데이터셋 수집					
	데이터셋 전처리					
	자연어 처리					
	DB 구축					
	웹 구축					
	모델 제작					

자연어 처리는 뉴스 본문의 주요 내용을 파악하고 키워드를 파악하기 위해 필수적으로 사용된다.

전처리 과정을 통해 뉴스의 키워드를 가져오고 해당 뉴스의 유형을 파악한다.

웹은 해당 모델의 기능을 사용하기 위해 링크를 첨부하는 기능, 이전의 파악을 위해 저장한 뉴스의 결과를 저장하기 위해 DB와 웹의 구축이 필요하다. 또한 추가적으로 토론 기능을 더 할 경우 네트워크의 구축도 필요하다.

해당 프로젝트를 진행하기 위해 데이터셋의 전처리와 키워드 수집이 가장 많은 시간을 소모할 것으로 확인된다

※ 반드시 5쪽 이외의 분량으로 작성

－ 참고문헌(Reference)

- － 송상현. "가짜뉴스에 대한 규범적·기술적 규제방안 연구." 국내박사학위논문 성균관대학교 일반대학원, 2020. 서울
- － 오미애등. (2022). "기계학습 기반 보건복지분야 가짜뉴스(fake news)탐지 방법 연구". 한국보건사회연구원
- － <https://www.verify.fox/> (VERIFY TOOL-FOX NEWS)
- － 이동호 외6명.(2018.5). "딥러닝 기법을 이용한 가짜뉴스 탐지". 2018년도 춘계학술발표대회 논문집 제25권 제1호(pp.384~387). 한국정보처리학회