

11조 Fact Shield

중간발표

20224151 김봉민
20213673 최지우
20222228 오주연

2025-04-19
화 234 인공지능

목차

Table of Contents

1 프로젝트 개요

프로젝트 목적 및 목표

2 진행 현황

현재까지 완료한 작업 내역

3 향후 계획

9주 ~ 14주차의 주차별 세부 계획

4 변경 사항 요약

일정 or 기능 변경 사유 및 내용 설명

5 시연 계획

최종 결과물 형태
시연 영상 제작 방식 구체화

프로젝트 개요

Project Overview

배경 요인

Background Factors

● 가짜뉴스 확산

온라인상에서 유명인의 사망과 관련한 루머, 조작된 썸네일 영상 등 허위 정보 유포가 급증하며 사회적 혼란을 초래하고 있음

한국경제 · 3주 전 · 네이버뉴스

추성훈, 故 김새론 장례비 전액부담? "가짜뉴스" [공식]

사이버 렉카 유튜버들이 돈벌이를 위해 가짜뉴스를 조직적으로 제작하고 있으나 1인 미디어는 '방송'으로 인정되지 않기 때문에 언론중재법 개정안이나 방송법의 적용을 받지 않는다. 따라서 피해자가 직접 해당 콘텐츠...



서울신문 PICK · 3일 전 · 네이버뉴스

中, 가짜 '서울신문' 등 사이트 만들어 '시진핑 정권' 홍보

국정원은 아직 해당 사이트에서 국내 언론사 기사의 무단 도용이나 가짜 뉴스를 퍼뜨리는 모습은 확인되지 않았다고 밝혔다. 그러나 해당 사이트가 국내 언론사의 신뢰도를 악용해 이런 행위를 할 가능성이 있다고 보...



매일경제 PICK · 5일 전 · 네이버뉴스

'박성광 아내' 이솔이 "암투병 끝 사망? 가짜뉴스 어떡해야..."

또한 해당 메시지에는 이솔이의 사망설 가짜뉴스를 터뜨린 영상 링크도 포함됐다. 이솔이는 이에 "이런 DM(다이렉트 메시지)가 하루에 수십통씩 온다"며 "저도 유튜브에 신고를 해봤는데 더이상 어떻게 해야할지 모...



SS 스포츠서울 · 2025.03.19. · 네이버뉴스

신기루, 사망설 가짜 뉴스에 분노 "진짜 너무하네" [★SNS]

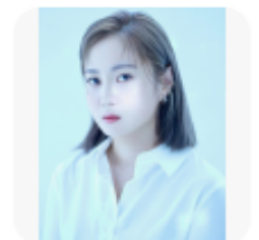
신기루 사례 이전에도 가짜 뉴스로 인한 피해 사례가 많았다. 허위 사실을 담은 가짜 뉴스 유포가 빈번하게 벌어지며 사회 문제시 되고 있다. 특히 각종 플랫폼을 통해 가짜 뉴스가 우후죽순 확산되며 애꿎은 피해자들만...



T 톱스타뉴스 · 6일 전

장도연 절도 루머 사실 아냐, 박나래 측 가짜뉴스에 강경 대응

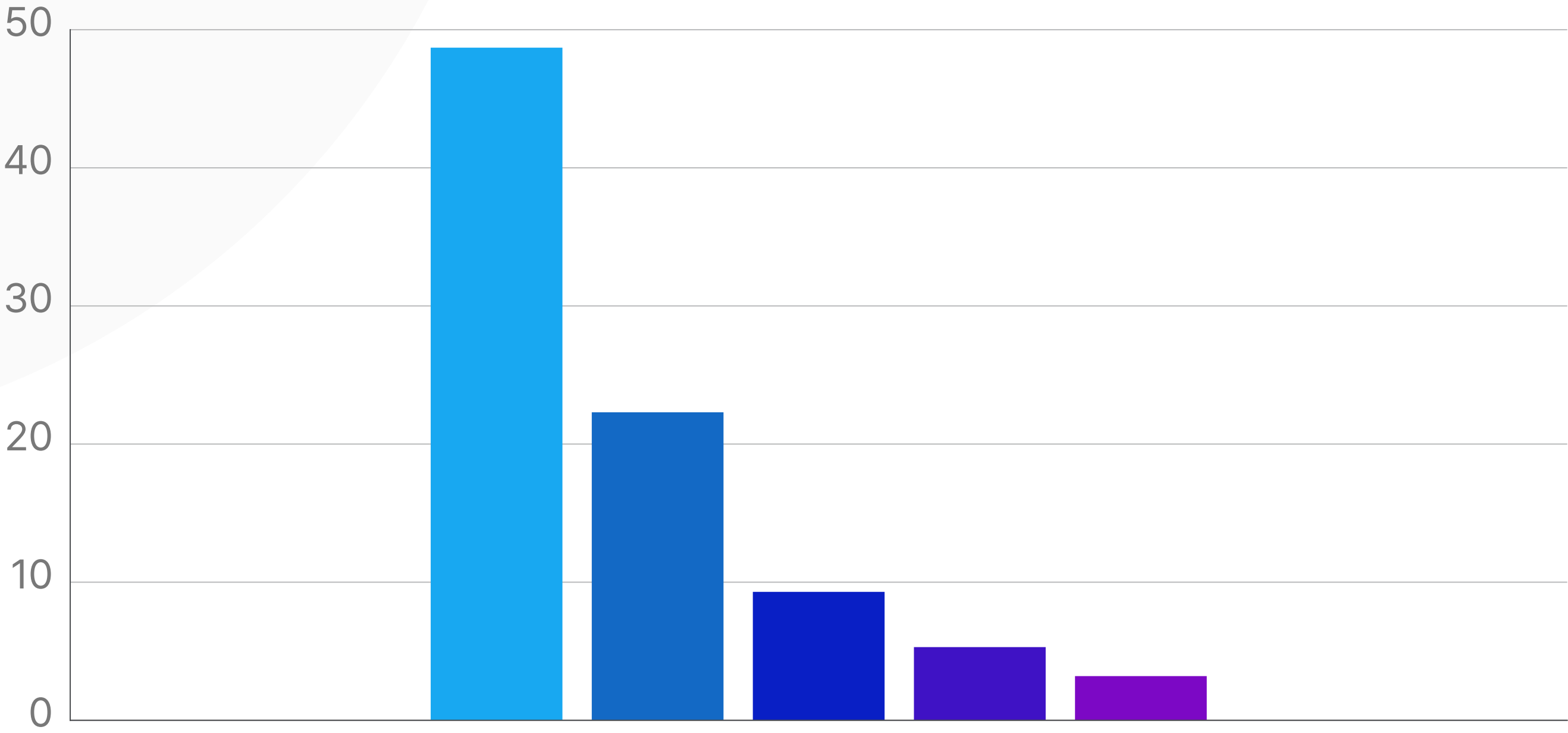
지목하는 가짜뉴스가 퍼졌고, '박나래 장도연'을 검색하면 관련 영상을 쉽게 찾아볼 수 있는 상황이었다. 이에 대해 박나래 측은 "근거 없는 허위사실 유포에 대해서는 어떠한 선처 없이 법적 조치를 취하겠다"고 밝혔다....



● 가짜 뉴스로 인한 언론의 신뢰도 저하

반복되는 허위 콘텐츠는 언론 및 플랫폼에 대한 대중의 신뢰를 약화시키고 있음

- 허위/조작정보
- 편파적 기사
- 자사 이기주의 기사
- SNS를 그대로 복붙한 기사
- 기타



한국언론의 문제점 <한국언론진흥재단.2020>

프로젝트의 목표



가짜 뉴스 탐지 AI Fact Shield를 통해 정보
신뢰성 확보 및 피해 최소화가 이 프로젝트의
목표임

진행 현황

Progress

현재까지 완료한 작업

활동	착수보고	5주차	6주차	7주차	중간보고
언어모델 조사					
데이터 수집 및 정제					
사전 학습 모델 검토					
API 연동 설계					
추진보고서 작성					
시스템 설계도 작성					
UI 디자인					
산출물			모델 제안서	정제된 데이터셋 추진보고서 시스템 설계도	Figma 디자인

AI 언어모델 비교

Model Comparison

모델	문맥 이해	학습 속도	리소스 요구	난이도
KoBERT				
KoELECTRA				
KLUE-RoBERTa				

- 1

KoBERT: 가볍고 학습 빠름, 실습 자료 풍부함
KoELECTRA: 빠르지만 세팅이 복잡함
KLUE-RoBERTa: 문맥 이해 뛰어나지만 무겁고 느림
- 2

Colab 기반 실습 환경, 빠른 반복 테스트 필요
GPU 자원이 제한적 (8~12GB 수준)

선택 모델

KoBERT

- 구현 난이도가 낮고, 학습속도 빠름
- 한국어 뉴스에 적합한 성능과 풍부한 예제 제공
- 시간 내 달성 가능성이 높음

데이터 수집 및 정제

Data Collection

Data Refining



데이터 출처

- AI Hub에서 제공한 낚시성 기사 탐지 데이터셋 사용
- 기사 제목과 본문의 불일치, 주제 일관성 부족 기사 포함



데이터 규모

- 733,427건



데이터 유형

- 신문기사 기반 텍스트
- JSON (제목/본문 불일치)

데이터 성능 지표 (Classification)

본문 주제 불일치 탐지

Accuracy



F1-Score



데이터 수집 대신 검증된 AI HUB 데이터셋을 활용함으로써
정제 과정의 신뢰성과 효율성을 확보

사전 학습 모델 검토

모델 구조

- KoBERT 모델 + Linear

학습 세팅

- Epoch: 3
- 손실함수: CrossEntropy + 클래스 가중치 적용
- GPU 사용 (cuda)

Metric	참	거짓
Precision	0.64	0.65
Recall	0.70	0.59
F1-score	0.67	0.62

```

(kobert_env) C:\Users\win\Desktop\kobert_news_project>python train_kobert.py
python: can't open file 'C:\Users\win\Desktop\kobert_news_project\train_kobert.py': [Errno 22] Invalid argument

(kobert_env) C:\Users\win\Desktop\kobert_news_project>python train_kobert.py
KoBERT 전체 데이터 학습 시작!
사용 디바이스: cuda
라벨 분포:
Label:
0 36643
1 34527
Name: count, dtype: int64
[Train] Epoch 1: 100% | 3559/3559 [23:38<00:00, 2.51it/s]
Epoch 1 완료 - Loss: 2343.5519, Accuracy: 0.5880
[Train] Epoch 2: 100% | 3559/3559 [23:34<00:00, 2.52it/s]
Epoch 2 완료 - Loss: 2182.6679, Accuracy: 0.6499
[Train] Epoch 3: 100% | 3559/3559 [23:37<00:00, 2.51it/s]
Epoch 3 완료 - Loss: 2012.1717, Accuracy: 0.6952

분류 성능 보고서:
      precision    recall  f1-score   support

   참         0.64         0.70         0.67         7329
   거짓        0.65         0.59         0.62         6905

 accuracy         0.65         0.64         0.65         14234
 macro avg        0.65         0.64         0.64         14234
 weighted avg        0.65         0.65         0.64         14234

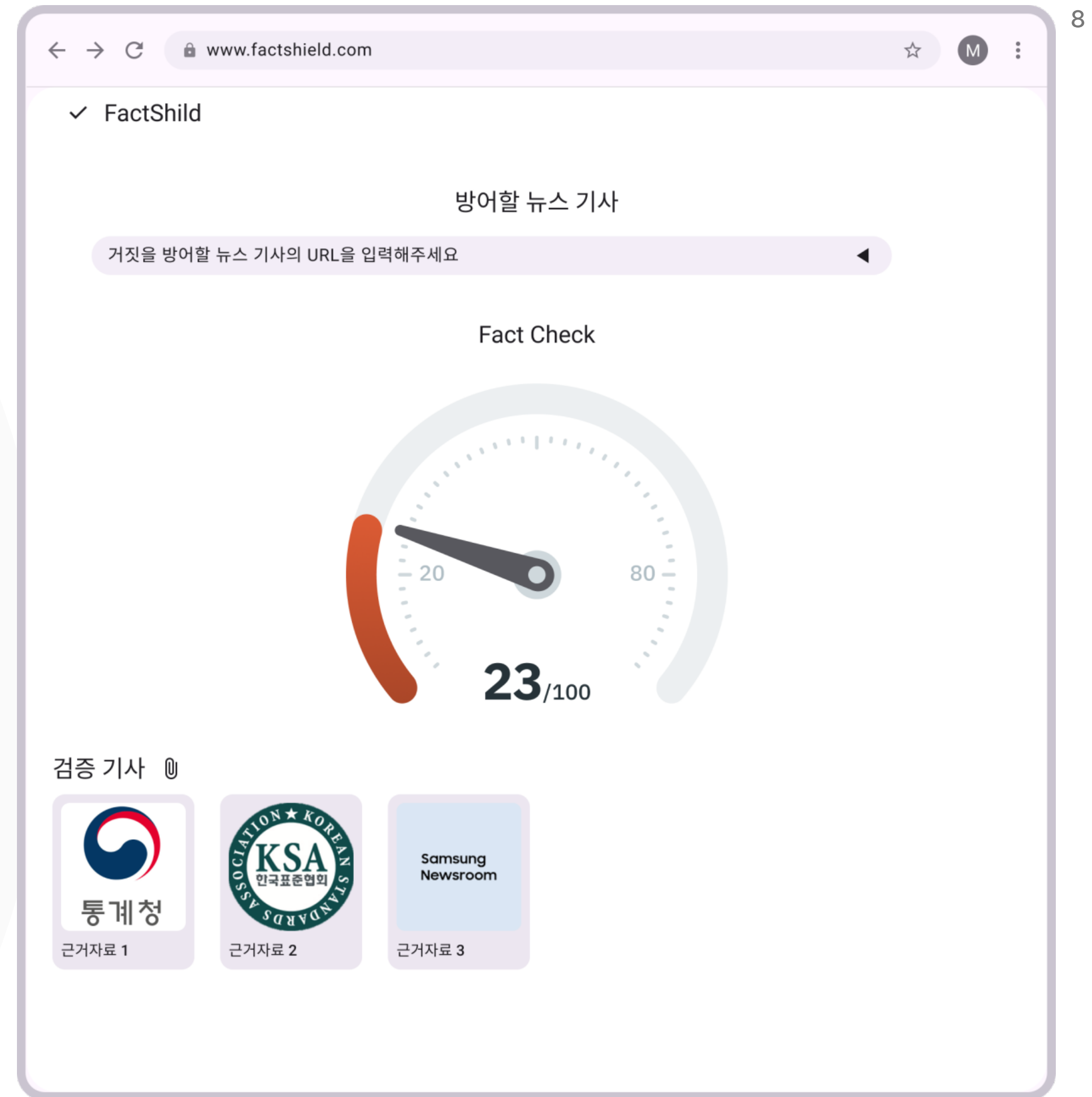
(kobert_env) C:\Users\win\Desktop\kobert_news_project>
  
```

전체 정확도 65% 수준

UI 디자인

User Interface Design

- URL 입력 시 AI가 신뢰 점수를 분석해 시각화하고 근거 기사까지 함께 제공 하는 UI 구성
- 사용자 인터페이스는 React로 개발 예정



향후 계획

Future plans

향후 계획

Future plans

활동	중간보고	10주차	11주차	12주차	13주차	14주차	최종보고
모델 성능 비교 분석							
모델 튜닝 및 AI 모델 학습							
프론트엔드 개발							
API 호출 구현							
백엔드 개발							
모델 테스트 및 검증							
시스템 배포							
문서화							
산출물			프론트 코드	API 명세서	백엔드 코드 학습된 모델 파일		서비스형 웹사이트 (URL) 데모 자료

Classification
by Name

최지우

오주연

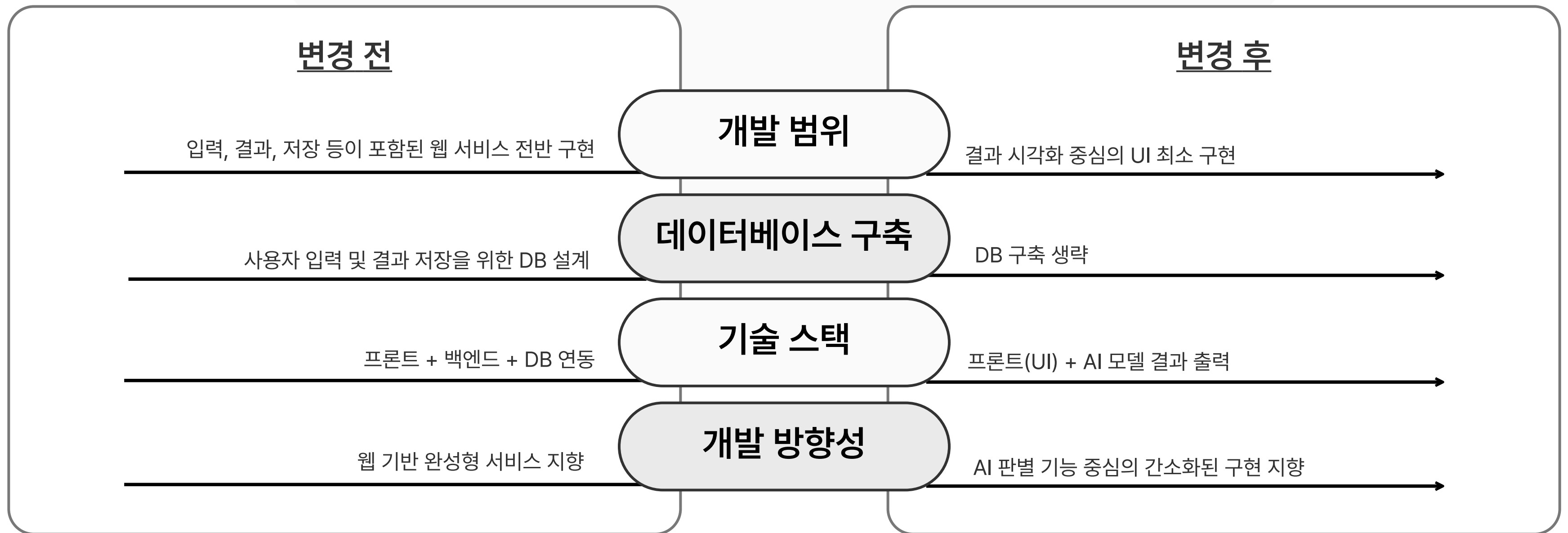
김봉민

변경 사항

Changes

변경 사항 요약

DB 구축을 생략하고, 모델 결과 시각화 중심으로 개발 범위를 조정



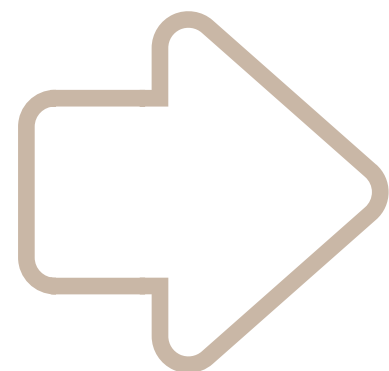
시연 계획

Demonstration

시연 계획



뉴스 URL 입력



AI 분석 결과 출력



근거 기사 확인

감사합니다