

# 2026년 한이음 드림업 프로젝트 수행계획서

## I. 프로젝트 개요

프로젝트명	설명가능한 가짜뉴스 탐지 AI Agent 플랫폼
주제영역	<input checked="" type="checkbox"/> 생활 <input type="checkbox"/> 업무 <input checked="" type="checkbox"/> 공공/교통 <input type="checkbox"/> 금융/핀테크 <input type="checkbox"/> 의료 <input type="checkbox"/> 교육 <input type="checkbox"/> 유통/쇼핑 <input type="checkbox"/> 엔터테인먼트
기술분야	<input type="checkbox"/> 소프트웨어 <input checked="" type="checkbox"/> 인공지능(AI) <input type="checkbox"/> 스마트디바이스 <input type="checkbox"/> 방송·콘텐츠 <input type="checkbox"/> 디지털융합 <input type="checkbox"/> 차세대보안 <input type="checkbox"/> 사이버보안
성과목표	<input checked="" type="checkbox"/> 특허출원 <input checked="" type="checkbox"/> 논문게재 <input checked="" type="checkbox"/> 앱등록 <input checked="" type="checkbox"/> 프로그램등록 <input type="checkbox"/> 기술이전 <input checked="" type="checkbox"/> 실용화 <input checked="" type="checkbox"/> 공모전 (공모전명: 한이음 엑스포) <input type="checkbox"/> 기타 ( )
수행기간	2026. 4. 1. ~ 2026. 10. 30.
프로젝트 소개 및 제안배경	<p><b>[프로젝트 소개]</b></p> <p>본 프로젝트는 뉴스 기사 URL 입력 및 키워드를 기반으로 AI 에이전트가 자동으로 기사를 수집 및 분석하여 가짜뉴스 여부를 판정하고, 그 판단 근거를 설명 가능한 형태로 제시하는 가짜뉴스 검증 플랫폼 개발을 목표로 한다. 데이터 수집부터 모델 학습, 서비스 배포까지 AI 개발의 전 과정을 아우르는 프로젝트이다. 단순한 진위 판별을 넘어 사용자에게 논리적인 판단 근거를 제공함으로써, AI 판별 결과에 대한 신뢰성을 확보하는 것을 핵심으로 한다.</p> <p><b>[제안 배경]</b></p> <p>온라인과 SNS를 통해 정보가 전파되는 환경에서 검증되지 않거나 사실이 아닌 가짜뉴스가 사회적 혼란과 불신을 유발하는 심각한 문제로 나타나고 있다. 문제의 중요성이 커에도 불구하고 일반 사용자가 정보의 진위를 직접 검증할 수 있는 기술적 방법은 매우 제한적인 상황이다.</p> <p>이를 해결하기 위해, 뉴스 데이터 기반의 이슈 분석 기술과 AI를 접목하여 올바른 정보 소비를 돕는 서비스를 기획하게 되었다. AI 기술이 사회적으로 중요한 문제를 해결하는 데 기여해야 한다는 목적의식을 기반으로 잘못된 정보로 인한 개인과 사회의 피해를 예방하고자 한다. 본 프로젝트를 통해 사용자가 사실에 기반한 올바른 의사결정을 내리고 객관적으로 검증된 정보를 접할 수 있는 환경을 구축하고자 한다.</p>
주요기능	<p><b>[뉴스 자동 수집, 전처리]</b></p> <ul style="list-style-type: none"> <li>• 뉴스 URL 입력 시 기사 원문 자동 수집 및 AI 분석용 전처리 수행</li> </ul> <p><b>[가짜뉴스 판별 AI]</b></p> <ul style="list-style-type: none"> <li>• Transformer · LLM 기반으로 뉴스의 가짜뉴스 여부를 판별</li> </ul>

	<p><b>[판별 근거 제공(XAI)]</b></p> <ul style="list-style-type: none"> <li>• 중요 문장 및 판단 근거를 설명 가능한 형태로 제공</li> </ul> <p><b>[웹, 모바일 플랫폼]</b></p> <ul style="list-style-type: none"> <li>• 대국민 가짜뉴스 검증 웹 서비스 및 모바일 앱 구현</li> </ul>
<p>적용 기술</p>	<p><b>[인프라 및 모니터링]</b></p> <ul style="list-style-type: none"> <li>• 클라우드 GPU 서버: 고성능 GPU를 활용하여 모델의 학습을 위한 인프라 제공, 서비스 제공을 위한 서버 운영</li> <li>• Grafana: 수집한 메트릭 데이터를 대시보드 형태로 시각화</li> <li>• Docker: 컨테이너 패키징을 통한 배포 및 운영 효율성 증대</li> </ul> <p><b>[AI 모델 학습 및 추론]</b></p> <ul style="list-style-type: none"> <li>• Python: 데이터 전처리, 모델 학습, 추론 파이프라인 구축에 필요한 프로그래밍 언어</li> <li>• PyTorch: KoELECTRA 모델의 파인튜닝과 Attention Rollout 기반의 XAI 알고리즘 구현에 활용</li> <li>• HuggingFace: 사전 학습된 모델을 편리하게 로드하고 관리하기 위한 라이브러리</li> </ul> <p><b>[뉴스·가짜뉴스 공개 데이터셋]</b></p> <ul style="list-style-type: none"> <li>• Alhub: 한국어 뉴스 데이터 학습을 위한 데이터셋</li> <li>• SNU팩트체크: 언론사 팩트체크 결과가 아카이빙된 플랫폼</li> <li>• COVID-19 Fake News Dataset: 보건/의료 분야의 특정 도메인 내 허위 정보 패턴을 학습하기 위한 데이터셋</li> </ul> <p><b>[웹·모바일 서비스 개발]</b></p> <ul style="list-style-type: none"> <li>• Flutter: 프론트엔드 개발을 위한 프레임워크</li> <li>• Spring boot: 백엔드 개발을 위한 프레임워크</li> <li>• PostgreSQL: 서비스시 필요한 데이터를 저장할 데이터베이스</li> </ul> <p><b>[형상관리 및 협업 도구]</b></p> <ul style="list-style-type: none"> <li>• GitHub: 소스 코드의 버전 관리 및 GitHub Actions를 통한 CI/CD</li> <li>• Notion: 프로젝트 일정 관리, 기술 문서화</li> <li>• Discord: 실시간 질의응답, 공지 전달, 회의 진행</li> </ul>
<p>예상 결과물</p>	<p><b>가짜뉴스 판별 서비스: 사용자 여정 및 인터랙션 플로우</b></p> <p><b>1. 서비스 접속 및 이슈 탐색</b></p> <p>메인 화면 확인하고 싶은 기사 URL을 입력하세요 URL 입력창 검색창에 URL을 입력하여 특정 기사 직접 분석</p> <p><b>2. 정보 선택 및 요약 확인</b></p> <p>정보 기사 &amp; SNS 게시물 AI 핵심 요약 (병대한 원문 요약) AI가 분석한 이슈 맥락: ... AI가 분석한 이슈 맥락: ... AI 생성 핵심 요약으로 빠른 이슈 파악</p> <p><b>3. 심층 검증 및 근거 확인 (XAI)</b></p> <p>기짜뉴스 의심 주석 (Annotation): ... 구체적인 반박 근거 CSIS 보고서 장부 보자료 설명 가능한 AI (XAI)로 근거 확인</p> <p><b>4. 피드백 및 제보</b></p> <p>피드백 보내기 오분서 이용 오분서 제보 API가 표시된 정보들 피드백 사항 사용자 피드백을 통한 모델 성능 개선 참여</p>

<b>기대효과 및 활용 분야</b>	사회적 기대효과: 가짜뉴스 확산 방지 및 대국민 정보 신뢰성 향상 교육·연구 활용: AI·데이터 기반 연구 성과 창출 서비스 활용분야: 대국민 뉴스 검증 서비스, 공공·언론·교육기관 활용
-------------------------	--

## II. 프로젝트 수행계획

### 1. 프로젝트 개요

#### 가. 프로젝트 소개

해당 프로젝트는 사용자가 특정 뉴스 URL을 입력하거나 키워드 검색을 통해 자동 수집된 뉴스 기사 중에서 원하는 기사 콘텐츠를 클릭하면 AI 에이전트가 기사의 원문을 자동으로 분석하여 가짜뉴스 여부를 판정하고, 그 결과와 함께 판단 근거를 사용자가 이해할 수 있는 설명 가능한 형태(XAI)로 즉각 제공하는 시스템이다.

또한 이는 온라인상에서 무분별하게 확산되는 조작된 허위 정보를 구별하기 어려운 문제를 해결하며, 단순한 진위 판별을 넘어 논리적이고 명확한 근거를 제시함으로써 AI의 판별 결과에 대한 사용자의 신뢰성 확보를 목표로 하는 미디어 검증 솔루션이다.

#### 나. 추진배경 및 필요성

##### 1. 생성형 AI 발전으로 인한 가짜뉴스의 증가

현재의 미디어 환경은 전문가가 아니더라도 LLM을 이용해 정교한 허위 정보를 무한대로 생성할 수 있는 단계에 진입했다.

- 과거에는 뉴스 생산에 취재·검증·편집 등 많은 인력과 시간이 필요했지만, 이제는 비전문가도 생성형 AI와 간단한 프롬프트만으로 기사 형식의 텍스트를 빠르게 만들어낼 수 있음
- 이러한 기술은 실제 기사와 유사한 정교한 표현으로 인해 식별이 어려운 고도화된 가짜 뉴스를 대량으로 생성·유포할 수 있게 만들며, 검증되지 않거나 의도적으로 조작된 정보의 확산을 가속화함
- AI의 발전으로 인해 가짜 뉴스의 증가가 촉진되고 있음

또한, 연구 결과에 따르면 사람과 AI 모두 LLM이 만든 가짜뉴스를 판별하는 데 어려움을 겪고 있다.

- 대학생 참가자들이 LLM으로 제작한 252개의 가짜뉴스 기사를 대상으로 실험한 결과, 사람과 AI의 탐지 정확도는 모두 약 60% 수준에 그침<sup>1</sup>
- LLM은 실제 뉴스를 판별하는 능력은 사람보다 약 68% 높게 나타나지만, 자신이 직접 생성한 가짜뉴스를 식별하는 데는 명확한 한계를 보임<sup>2</sup>
- 이러한 통계는 누구나 정교한 가짜뉴스를 대량 생산할 수 있는 시대에 진입했음을 의미하며, 기존 탐지 시스템의 보완이 시급함을 보여줌

##### 2. 가짜뉴스 확산에 따른 막대한 사회적·경제적 피해 심화

<sup>1</sup> <https://arxiv.org/abs/2410.19250>

<sup>2</sup> <https://arxiv.org/abs/2410.19250>

가짜뉴스는 단순한 정보 왜곡을 넘어 국가와 기업의 경제 지표에 실질적인 타격을 주는 경제적 위협 요소이다.

- 볼티모어 대학교와 사이버 보안 기업 CHEQ의 조사에 따르면, 가짜뉴스로 인한 연간 글로벌 경제 비용은 약 782억 달러로 추산됨<sup>3</sup>
- 거시경제 관점의 연구는 가짜뉴스 충격이 실업률 증가, 생산 감소, 경기 변동성 확대를 나타내며, 특히 부정적 가짜뉴스가 경기 위축을 더 크게 유발한다고 분석함
- 영국에서 진행된 「The Cost of Lies」 연구에 따르면 코로나 관련 마스크·방역 가짜뉴스로 인해 2020년 2·3분기에만 영국 GDP 손실 약 36억 파운드(3.6 billion GBP, 약 46억 달러)가 발생했다고 추정함<sup>4</sup>

### 3. 기존 팩트체크 시스템의 한계 노출

기존의 팩트체크 시스템은 가짜뉴스의 확산 속도와 양을 감당하지 못하는 구조적 한계가 존재하여 일반 시민들이 일상에서 정보를 거르는 데 어려움을 겪고 있다.

- 이미 허위정보의 양과 속도가 수작업 팩트체크 역량을 구조적으로 초과했다고 지적하며, 수작업 팩트체크는 개별 주장 검토, 자료 수집, 기사 작성에 시간이 많이 들기 때문에, 이미 퍼진 가짜뉴스의 초기 파급을 막기 전에 ‘뒤따라가는 구조’에 머물게 됨<sup>5</sup>
- 팩트체크 사이트는 “간혹 이용하는 수단”으로는 꽤 언급되지만, 브랜드 인지도·도달 범위가 대형 뉴스매체에 비해 매우 제한적이라, 일상적인 검증 채널로 기능하지 못하는 것으로 평가됨<sup>6</sup>
- 팩트체크 기관들은 단순 기사 검증을 넘어 청소년·시니어·일반 시민 대상 미디어 리터러시 교육을 강화하며 “스스로 검증하는 시민”을 늘리려 하고 있지만 인터넷 접근성·예산·인력 한계 탓에 전국민을 포괄하는 것으로 확장되기에는 어려움이 존재함

### 4. 플랫폼의 책임성 결여 및 자율규제 체계의 실효성 한계

현재의 플랫폼 규제는 AI 기반 가짜뉴스의 확산 속도와 기술적 정교함을 제어하기에는 법적·제도적 구속력이 부족한 실정이다.

- 국내 규제 체계는 방송통신위원회와 방송통신심의위원회 중심의 사후 행정 규제와 플랫폼의 선의에 기댄 자율규제에 편중되어 있어, 생성형 AI를 활용한 실시간 대량 유포에 즉각 대응하기 어려운 구조적 한계를 지님
- AI 기술 발달로 가짜뉴스 생성 및 전파 비용이 급감했음에도 불구하고, 플랫폼의 기술적 조치 이행 여부를 감시할 투명성 보고 의무나 위반 시 실질적인 과징금·형사 책임 등의 제재 수단이 정교화되지 않아 민간 대응망의 공백이 심화됨<sup>7</sup>
- 자율규제 가이드라인의 참여와 이행이 사업자의 선택에 맡겨져 있고 모니터링 및 평가 메커니즘이 약해, 대형 플랫폼이 가짜뉴스 방역을 위해 적극적인 투자와 기술 개선을 할 유인이 낮다는 분석임<sup>8</sup>

<sup>3</sup> <https://www.weforum.org/stories/2025/07/financial-impact-of-disinformation-on-corporations/>

<sup>4</sup> [https://londoneconomics.co.uk/wp-content/uploads/2021/01/The-Cost-of-Lies\\_clean\\_2.2.21.pdf](https://londoneconomics.co.uk/wp-content/uploads/2021/01/The-Cost-of-Lies_clean_2.2.21.pdf)

<sup>5</sup> <https://www.asc.upenn.edu/research/centers/milton-wolf-seminar-media-and-diplomacy-20>

<sup>6</sup> <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/how-public-checks-information-it-thinks-might-be-wrong>

<sup>7</sup> <https://www.korea.kr/news/policyNewsView.do?newsId=148938576>

<sup>8</sup> <https://www.korea.kr/news/policyNewsView.do?newsId=148938576>

## 다. 선행 연구

### 1. 주장(Claim) 추출 및 복합 추론의 정형화

팩트체크의 첫 단계는 검증 가능한 단위를 정의하는 것이다. 초기 연구들이 기사 전체의 성격을 규명하려 했다면, 최근 연구는 이를 세부 주장으로 분해하여 다단계로 추론하는 데 집중한다.

- **FEVER (2018)**<sup>9</sup>: 자동 팩트체킹의 표준 과제인 '주장 - 근거 - 판정'의 3요소 구조를 처음 만들고, 단순히 판정 결과만 맞는 것이 아니라, 이를 뒷받침하는 정확한 문장을 찾아내야만 성능을 인정하는 지표를 넣음
- **HoVer (2020)**<sup>10</sup>: 단일 문서 내에서 답을 찾을 수 없는 복합 주장을 검증하기 위해 멀티홉 (Many-hop) 추론을 도입하여 여러 문서를 연결해야만 진위 확인이 가능한 주장을 다루며, 검색 단계에서 정보가 누락될 경우 전체 시스템 성능이 급격히 저하됨을 수치적으로 입증함

### 2. 비정형 및 구조화 데이터 기반 근거 검색 기술

신뢰할 수 있는 정보를 어디서, 어떻게 가져올 것인가에 대한 연구이다. 특히 텍스트 이외의 구조화된 정보를 다루는 능력이 강조된다.

- **FEVEROUS (2021)**<sup>11</sup>: 수치 데이터나 통계 정보가 포함된 표를 논리적으로 해석하여 문장과 대조하는 기술을 제안하였으며, 구조화된 정보의 정밀한 추출이 팩트체크의 신뢰성을 높이는 핵심 요소임을 증명함
- **FactIR (2025)**<sup>12</sup>: 판정 모델의 성능보다 실제 웹 환경에서 공신력 있는 문서를 선별해오는 검색 모델의 품질이 전체 시스템의 성패를 좌우한다는 점을 강조하며, 이를 위한 전용 벤치마크를 구축하여 검색 고도화의 방향성을 제시함

### 3. 설명 가능한 AI(XAI)와 원자 단위 평가

결과에 대한 이유를 투명하게 밝히는 연구이다. 사용자가 결과를 수용할 수 있도록 논리적 근거를 제시하는 방식을 주장한다.

- **SciFact (2020)**<sup>13</sup>: 모델이 답변을 생성할 때 임의의 문장을 나열하는 것이 아니라, 반드시 판정의 직접적인 이유가 되는 문장만을 선별하여 보여주도록 강제하는 최소 문장 집합인 'rationale'의 필요성을 말함
- **FactScore (2023)**<sup>14</sup>: 긴 문장의 설명을 하나의 단위로 보지 않고, 'Atomic Fact (원자 사실)' 단위로 분해하여 개별적으로 검증하는 체계를 제안하여 문장 내에서 어느 부분은 사실이고 오류인지를 세밀하게 판별할 수 있게 하고, 사용자에게 더욱 정교한 피드백을 제공할 수 있게 함

<sup>9</sup> <https://aclanthology.org/N18-1074/>

<sup>10</sup> <https://aclanthology.org/2020.findings-emnlp.309/>

<sup>11</sup> <https://arxiv.org/abs/2106.05707v1>

<sup>12</sup> <https://arxiv.org/html/2502.06006v1>

<sup>13</sup> <https://aclanthology.org/2020.emnlp-main.609/>

<sup>14</sup> <https://aclanthology.org/2023.emnlp-main.741/>

#### 4. LLM 및 RAG 기반 추론 아키텍처

LLM(Large Language Model)을 활용하되, 모델 내부의 지식에만 의존하지 않고 외부 근거를 활용해 환각 현상을 억제하는 방법이다.

- **ClaimCheck (2025)<sup>15</sup>**: 소형 언어 모델을 활용하더라도 웹 검색 쿼리 계획, 증거 요약, 재검색 단계를 체계적으로 구성함으로써 거대 모델을 뛰어넘는 실시간 팩트체크 성능을 확보할 수 있음을 보여줌

---

<sup>15</sup> <https://arxiv.org/abs/2510.01226>

## 라. 선행 서비스

### 1. 개방형 팩트체크 플랫폼 (팩트체크넷)

해당 플랫폼은 AI를 통한 판별보다는 시민, 전문가, 언론사가 유기적으로 협력할 수 있는 콘텐츠 관리 및 허브 구조 구현에 집중하고 있다.

#### [해당 서비스의 특징]

- 역할 분리형 허브 구조: 시청자미디어재단과 민간 IT 단체가 분리 운영하며, 검증 결과는 각 언론사 원문 링크를 요약·수집하는 저장소 형태임
- 검증 요청 기반 프로세스: 시민의 검증 요청 → 전문/시민 팩트체커 협업 → 언론사 최종 책임 및 발행으로 이어지는 단계별 워크플로우로 구성됨
- 계층적 권한 관리: 일반 사용자(제보·평가), 시민 팩트체커(조사 보조), 언론사 기자(최종 판정 및 발행)로 권한을 세분화하여 보도의 신뢰성과 책임성을 부여함

#### [사용자 수]<sup>16 17</sup>

극히 일부의 이용자만 활용하고 있으며, 하루 수만 건씩 생성되는 소셜 미디어상의 허위정보 절대량을 커버하기에는 연간 수백 건의 검증 처리량이 부족한 것을 알 수 있다.

초기 이용자 통계 (2021년 기준)	장기 운영 지표 (2022년 9월 기준 누적)
누적 이용 시민: 약 <b>12</b> 만 명 총 이용(조회·참여) 횟수: 약 <b>67</b> 만 회 검증 제안 건수: <b>263</b> 건	검증 제안: <b>781</b> 건 검증 완료: <b>315</b> 건 (연간 약 100~150건)

#### [운영 실태]

공공적 성격의 허위정보 사례 데이터베이스(DB)를 구축하고 시민 참여형 검증 문화를 확산하는 데 기여하였으나 예산 투입 대비 기술적 고도화 측면에서 한계를 보이고 있다.

- 사업 예산의 급팽창: '인터넷 환경 신뢰도 기반 조성' 예산이 2020년 5.5억 원에서 2021년 27.4억 원으로 약 **393.7%** 급증
- 성과 지연: 대규모 예산 투입에도 불구하고 핵심 과제인 'AI 활용 자동화 시스템'은 사업 2년 차에도 초기 데이터셋 구축 단계에 머물러 있으며, 실질적인 서비스 모델 구현이 지연되고 있음
- 처리 용량의 한계: 연간 검증 완료 건수가 약 **100~150**건 수준으로, 하루 수만 건씩 생성되는 생성형 AI 기반 가짜뉴스의 양적 폭발을 감당할 수 있는 처리 용량이 부족함

### 2. SNU 팩트체크 센터

다수 언론사의 검증 결과를 한데 모으고, 동일 이슈에 대한 서로 다른 판정 결과를 대조할

<sup>16</sup> <https://www.khan.co.kr/article/202510140600101>

<sup>17</sup> <https://press.ngonews.kr/newsRead.php?no=934630>

수 있는 비교 검증 플랫폼의 구조를 띄고 있다. 서로 다른 언론사에 속한 기자들이 해당 기사에 대해 수동으로 팩트체크를 하는 시스템이다.

**[해당 서비스의 특징]**

- 표준 판정 척도: '전혀 사실 아님'부터 '사실'까지 **5~6**단계의 표준화된 척도를 사용하여 서로 다른 언론사의 판정 결과를 객관적으로 비교할 수 있는 인터페이스를 제공함
- 교차 검증 시스템: 동일한 주장에 대해 여러 언론사가 각자 검증을 수행할 수 있으며, 플랫폼은 이를 병렬 배치하여 정보의 다각적 해석을 지원함
- 공적 인물 주장 **DB**화: 정치인 및 공적 인물의 발언을 데이터베이스화하여 특정 인물의 과거 발언과 현재 주장의 일치 여부를 추적할 수 있는 아카이브 기능을 갖추고 있음

**[영향력]**

네이버 뉴스 서비스와 연동되었기에 공공 플랫폼 대비 압도적인 노출량과 인지도를 보유하고 있었다.

- KBS, MBC, SBS, JTBC, 조선일보, 중앙일보, 동아일보 등 국내 **31**개 주요 언론사 참여함
- 'SNU 팩트체크 상' 시상 등을 통해 언론계 전체의 팩트체크 수준을 상향 평준화하는데 기여한 것으로 평가받음

**[한계점]**

- 운영 상태: **2024년 8월**에 운영이 중단되었고<sup>18</sup>, 현재는 기존 자료를 읽기 전용으로만 유지되고 있음
- 수작업 기반의 속도 한계: 개별 주장 검토 및 자료 수집에 막대한 인력과 시간이 소요되는 수작업 중심 구조임. 이미 확산된 가짜뉴스의 파급을 막지 못하고 사후에 대응하는 '뒤따라가는 구조'를 탈피하지 못함.

---

<sup>18</sup> [https://www.ohmynews.com/NWS\\_Web/View/at\\_pg.aspx?CNTN\\_CD=A0003055180](https://www.ohmynews.com/NWS_Web/View/at_pg.aspx?CNTN_CD=A0003055180)

## 2. 프로젝트 내용

### 가. 주요 기능

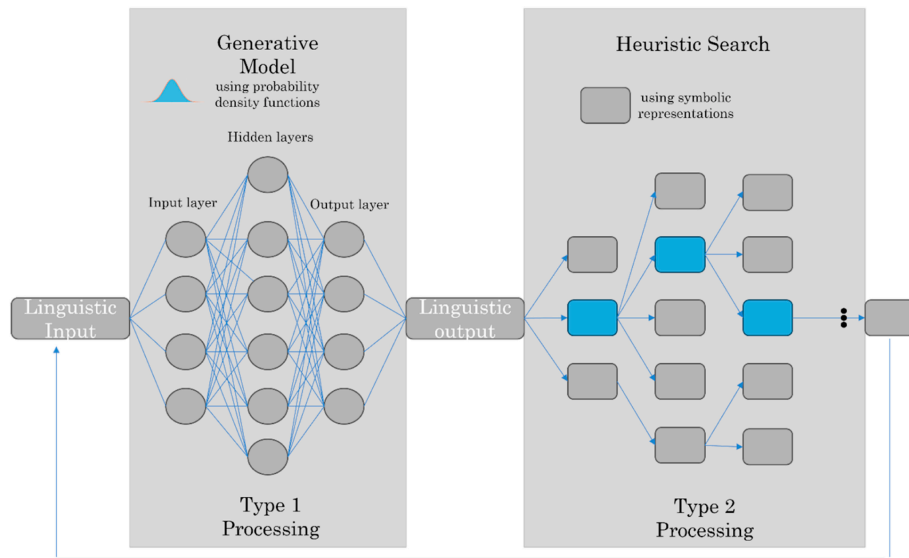
구분		기능	설명
S/W	수집 및 전처리	실시간 트렌드 크롤링	<ul style="list-style-type: none"> <li>- 실시간 트렌드를 파악하기 위해 <b>pytrends</b> 라이브러리를 사용하여 <b>Google Trends</b>의 한국 지역 인기 키워드를 상시 수집하며, <b>X(Twitter) Search API v2</b>를 통해 특정 키워드가 포함된 트윗 데이터를 확보한다.</li> <li>- 공식 <b>API</b>가 제공되지 않는 온라인 커뮤니티의 경우 <b>Python</b> 기반의 크롤러를 별도로 구축하여 본문을 추출하며, 모든 수집 과정은 <b>robots.txt</b> 정책을 엄격히 준수하여 진행한다.</li> <li>- 이렇게 수집된 대규모 원시 데이터는 <b>GCS</b>에 객체 형태로 보관한다. 데이터 조회 및 관리 효율성을 극대화하기 위해 핵심 속성 정보인 메타데이터만을 선별하여, <b>RDBMS</b>에 저장하는 계층형 저장 아키텍처를 사용한다.</li> </ul>
		데이터 정제	<b>HTML</b> 태그 제거 및 인코딩 통일 후, 문단·문장 단위로 분리하고, 특수문자 정리를 수행해 텍스트 형태로 정제한다.
	분석 및 판별	문맥 논리 구조 분석	<b>Transformer</b> 기반 <b>KoELECTRA</b> 모델에 <b>Attention Rollout</b> 기법을 적용한다. 기사 내에서 각 문장이 최종 판별에 미치는 영향력을 레이어별로 누적 추적하고, 중요도가 높은 문장들을 추출해 논리적 일관성 및 선동적 표현을 탐지한다,
		교차 검증	선별된 문장에서 <b>LLM</b> 이 검증할 수 있는 요소들을 추출한다. <b>RAG</b> 기반으로 정부 보도자료· <b>CSIS</b> 등 사전에 정의된 신뢰할 수 있는 출처로부터 수집한 문서와 대조하여, 사실에 얼마나 부합하는지를 <b>0~100%</b> 의 수치로 산출하고 판단 근거와 참조 출처를 함께 생성한다.
	시각화 및 설명	근거 기반 주석 생성	<ul style="list-style-type: none"> <li>- 가짜뉴스로 판별된 문단과 수집된 근거 문장을 결합하여 <b>LLM</b> 기반의 최종 요약 문장을 생성한다.</li> <li>- 생성된 문장에는 위키피디아 양식의 주석 체계를 적용 생성한 요약 문장을 달고, 사용자가 직접 사실 관계를 검증할 수 있도록 원본 출처 링크를 함께 첨부한다.</li> </ul>
서비스	인터랙티브 워드클라우드	수집된 키워드 빈도를 집계해 인터랙티브	

	제공	워드클라우드로 시각화하고, 사용자가 단어를 클릭하면 해당 이슈 관련 기사 목록과 판별 결과를 제공한다.
	사용자 피드백 시스템	잘못 분석된 글을 제보하여 모델을 재학습한다.

## 나. 적용 기술

### 1. 하이브리드 AI 추론 아키텍처

인간의 인지 구조에서 착안한 이중 추론 메커니즘을 기반으로 AI 추론 아키텍처를 설계한다. 심리학의 이중 처리 이론<sup>19</sup>에서 제시하는 속도가 빠르고 직관적인 ‘시스템 1’과, 느리지만 분석적이고 논리적인 ‘시스템 2’의 개념을 가짜뉴스 판별 파이프라인에 적용한다.



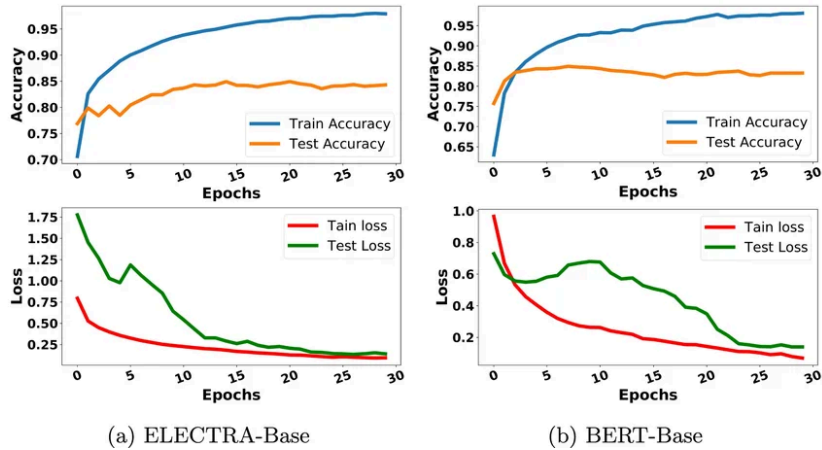
(그림1) 이중 처리 이론을 접목한 LLM 모델의 구성도

[출처] S. C. Bellini-Leite (2024). Dual Process Theory for Large Language Models. *Adaptive Behavior*, 32(4)

시스템 1: KoELECTRA 기반 패턴 인식 모듈로, 뉴스 본문을 입력으로 받아 가짜뉴스의 언어적 패턴, 문체, 구조적 특징을 학습된 가중치를 통해 빠르게 탐지한다.

- **ELECTRA** 모델: 기존 BERT 계열 모델의 마스킹 기반 사전학습(MLM) 방식과 달리, 생성기(Generator)가 삽입한 가짜 토큰을 판별기(Discriminator)가 구별하는 RTD(Replaced Token Detection) 방식으로 사전학습을 수행한다. 이러한 차이로 인해 분류 성능 및 학습 효율 면에서 BERT 계열 대비 우수한 성능을 보인다. 본 프로젝트에서는 KoELECTRA-Base-v3 모델을 한국어 가짜뉴스 데이터셋으로 파인튜닝(Fine-tuning)하여 활용한다.

<sup>19</sup>[https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/individual-differences-in-reasoning-implications-for-the-rationality-debate/2906AEF620B36C10018DD291F790BE97?utm\\_campaign=shareaholic&utm\\_medium=copy\\_link&utm\\_source=bookmark](https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/individual-differences-in-reasoning-implications-for-the-rationality-debate/2906AEF620B36C10018DD291F790BE97?utm_campaign=shareaholic&utm_medium=copy_link&utm_source=bookmark)



(그림2) 에폭(Epoch)에 따른 ELECTRA 및 BERT 모델의 정확도(Accuracy)와 손실(Loss) 변화 추이  
 [출처] ResearchGate, *Abusive Bangla Comments Detection on Facebook* (2022)

시스템 2: LLM 기반 사실 검증 모듈로, RAG(Retrieval-Augmented Generation) 아키텍처를 활용하여 구현한다. ‘시스템 1’이 탐지한 의심 신호를 바탕으로 실제 사실 관계를 외부 데이터와 대조해 검증하며, LLM을 단독 사용할 때 발생할 수 있는 환각(Hallucination) 문제를 최소화한다. ‘시스템 2’는 다음과 같은 순서로 동작한다

- **Claim 추출:** LLM이 Attention Rollout으로 선별된 중요 문장을 분석하여, 구체적인 수치·사건·행위 등의 사실 여부를 외부 데이터와 검증할 Claim을 선별하고, 최적화된 검색 쿼리를 자동 생성. 검증 불가능한 의견성 진술은 이 단계에서 제외하여 검증 정밀도를 높임.
- **근거 수집:** 생성된 쿼리를 기반으로 정부기관(질병관리청, 식품의약품안전처 등), 공신력 있는 언론사와 같이 신뢰할 수 있는 출처로부터 문서를 검색·수집함. LLM이 검증 과정에서 사용할 수 있는 출처를 사전에 정의하여 구조적인 신뢰성을 높임
- **최종 검증:** 수집된 문서를 컨텍스트로 제공받은 LLM이 각 검증 대상의 사실 여부를 사전에 정의된 신뢰할 수 있는 출처로부터 수집한 문서와 대조하여 사실에 얼마나 부합하는지를 0~100%의 수치로 산출함. 이 결과는 최종 판별 근거의 핵심 구성 요소가 됨.

## 2. Attention Rollout 기반 내재적 XAI

사후적으로 설명을 끼워 넣는 전통적 방식(Post-hoc XAI)과 달리, 파이프라인 내부에서 자연스럽게 설명이 생성되는 내재적 XAI(Intrinsic XAI) 구조를 채택한다. LIME, SHAP 등 기존 XAI 기법이 모델 외부에서 근사 해석을 시도하는 것과 달리, Attention Rollout은 모델 내부의 실제 연산 흐름을 직접 추적하여 해석의 신뢰성을 높일 수 있다.

### [Multi-head Self-Attention 구조]

- KoELECTRA는 12개의 Multi-head Self-Attention Layer로 구성되며, 각 레이어에서 토큰 간 상호 영향 관계가 Attention 가중치 행렬로 표현됨.
- 12개의 헤드는 각각 문법 구조, 의미 관계, 지시어 해소 등 서로 다른 언어적 역할을 분담하여 포착함.
- 단일 레이어의 Attention만으로는 이전 레이어의 누적 영향이 반영되지 않아 해석 왜곡이 발생하며, 이를 해결하기 위해 Attention Rollout을 적용함.

### [Attention Rollout 알고리즘]

- 각 레이어의 **Attention** 행렬에 항등행렬(I)을 더해 **Residual Connection**을 반영하고, 정규화 후 레이어 순서대로 순차적으로 행렬 곱하여 누적 영향을 추적함.

$$R_l = \hat{A}_l \cdot R_{l-1}, \quad \hat{A}_l = \frac{A_l + I}{\Sigma(A_l + I)}$$

#### [문장 단위 중요도 산출]

- 분류에 활용되는 [CLS] 토큰의 Rollout 값을 기준으로 각 토큰의 누적 영향력을 수치화하고, 문장 분리 라이브러리(kss)로 분리된 문장 단위로 집계하여 중요도 점수를 산출함.
- 중요도 상위 N개 문장을 선별하여 사용자에게 판단 근거로 시각화하는 동시에, System 2의 Claim 추출 단계 입력으로 전달함.

### 3. 데이터 수집 및 저장 아키텍처

수집부터 저장까지 전 과정을 역할에 따라 분리하여 설계하였다. 수집은 서버리스 기반으로 비용을 최소화하고, 저장은 데이터 성격에 따라 오브젝트 스토리지와 RDBMS로 이원화하여 성능과 비용을 동시에 최적화한다.

#### [서버리스 컴퓨팅]

- 서버리스 컴퓨팅을 활용한 아키텍처를 도입하여 뉴스 기사 크롤링 및 텍스트 전처리(정제, 정규화)를 수행함
- 상시 구동 서버 없이 이벤트가 발생할 때만 자원을 할당하므로 유휴 자원으로 인한 운영 비용을 절감함

#### [스토리지 서비스 기반 원시 데이터 저장]

- 크롤링을 통해 수집된 대용량 뉴스 원문, 비평 데이터는 스토리지 서비스에 객체 형태로 원본 그대로 저장함
- 스토리지 서비스의 높은 확장성과 저비용 스토리지 특성을 활용하여 지속적으로 누적되는 대규모 텍스트 데이터를 경제적으로 관리함

#### [RDBMS 스키마 최적화 및 메타데이터 관리]

- 관계형 데이터베이스에는 무거운 텍스트 데이터를 배제하고 메타데이터만 저장하여 스토리지 부하를 스토리지 서비스로 분산시킴
- 대규모 텍스트 처리가 필요한 AI 모델은 RDBMS를 거치지 않고 스토리지 서비스에서 직접 데이터를 호출하므로 데이터베이스 I/O 병목 현상을 차단하고 전체 시스템의 처리 속도가 향상됨

### 4. 통합 모니터링 및 로그 관리

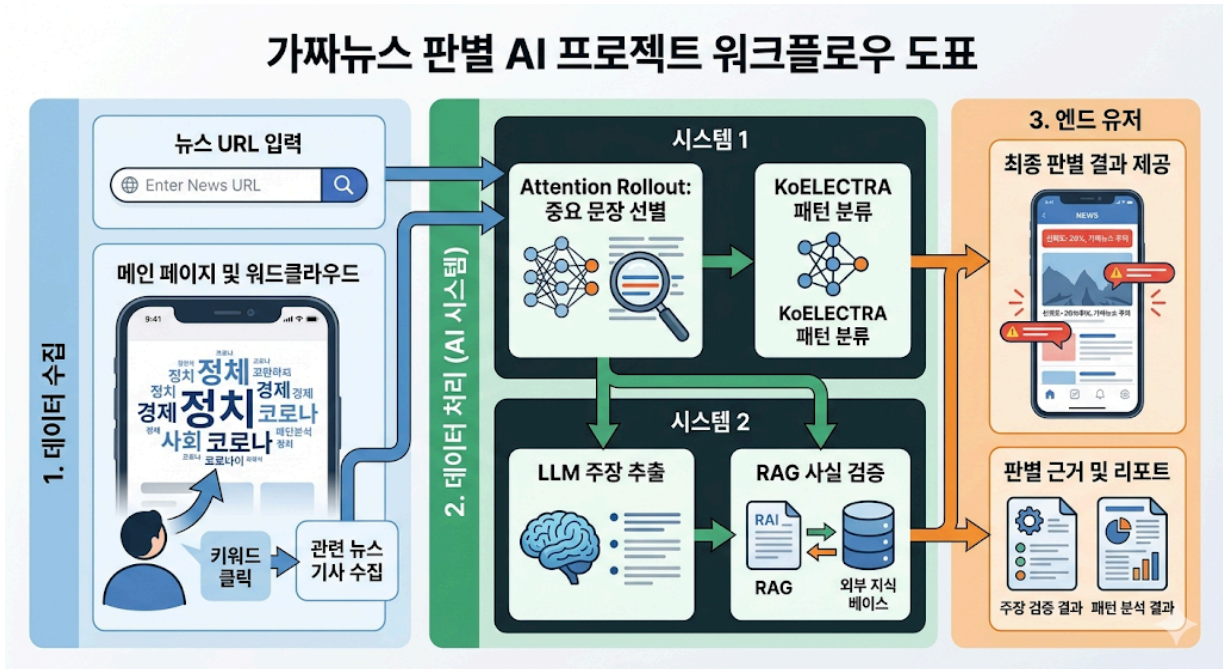
#### [실시간 리소스 모니터링 (Prometheus, Grafana)]

- Prometheus와 Grafana를 연동하여 CPU, Memory 등 서버 자원 및 AI 판별 모델이 동작하는 GPU의 실시간 상태를 모니터링함
- GPU 메모리 부족이나 서버 응답 지연 발생 시 Alertmanager를 통해 Slack 등으로 즉각적인 알림을 전송하여 신속하게 대응

**[중앙 집중식 로그 관리]**

- 로그를 중앙으로 수집하고 메트릭 데이터와 교차 검증하여 AI 모델의 판별 실패 원인이나 서버 오류를 체계적으로 분석

**가짜뉴스 판별 AI 프로젝트 워크플로우 도표**

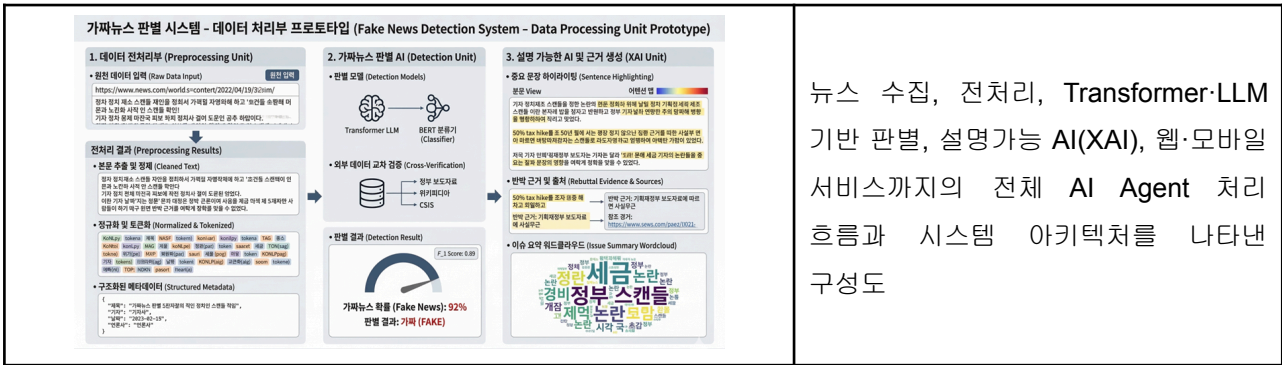


**다. 필요 기자재(기자재/장비)**

품목	활용계획
GPU 클라우드 서버	Transformer·LLM 모델 학습 및 추론 환경 구축
클라우드 스토리지	뉴스 데이터, 학습 데이터셋 및 모델 파일 저장
웹 서버 및 DB 서버	서비스 운영 및 사용자 요청 처리
개발용 노트북/PC	모델 개발, 서비스 연동 및 코드 관리
테스트용 모바일 기기	웹·모바일 서비스 기능 및 UI 테스트

**라. 예상 결과물**

예상 결과물 이미지	설명
<p><b>가짜뉴스 판별 서비스: 사용자 여정 및 인터랙션 플로우</b></p>	<p>사용자가 뉴스 URL을 입력하면 AI가 자동으로 기사 내용을 분석하고, 가짜뉴스 가능성과 신뢰도 지표, 판별 근거 문장을 시각적으로 제공하는 웹 서비스 화면</p>



뉴스 수집, 전처리, Transformer-LLM 기반 판별, 설명가능 AI(XAI), 웹·모바일 서비스까지의 전체 AI Agent 처리 흐름과 시스템 아키텍처를 나타낸 구성도

### 마. 성과목표

<b>성과목표</b>	<ul style="list-style-type: none"> <li>■ 특허출원</li> <li>■ 논문게재</li> <li>■ 앱등록</li> <li>■ 프로그램등록</li> <li>□ 기술이전</li> <li>■ 실용화</li> <li>■ 공모전 (공모전명 : 한이음 엑스포)</li> <li>□ 기타 ( )</li> </ul>
-------------	--

- 특허출원: 한이음 특허출원 컨설팅을 통한 특허출원
- 논문게재: 한국정보처리학회 추계학술대회 논문 투고
- 앱등록, 프로그램등록: 한이음 비용 지원을 통한 앱 등록 및 프로그램 저작권 등록
- 실용화: 대국민 가짜뉴스탐지 서비스 플랫폼 오픈
- 공모전: 한이음 엑스포 출전
- **AI 모델: KoELECTRA 파인튜닝 모델 F1-Score 0.75 이상 달성**

### 3. 프로젝트 수행방법

#### 가. 프로젝트 추진일정

구분	추진내용	추진일정							
		3월	4월	5월	6월	7월	8월	9월	10월
계획	프로젝트 주제 확정 및 역할 분담, 개발 환경 구축	■							
분석	가짜뉴스 데이터셋 조사 및 요구사항 분석	■	■						
설계	시스템 구조 설계		■	■					
	AI 모델 설계		■	■					
	데이터베이스 설계		■	■					
개발	뉴스 자동 수집/전처리 모듈 개발			■	■				
	가짜뉴스 판별 AI 모델 학습 및 개선			■	■	■			
	웹/모바일 서비스 UI 및 API 개발				■	■	■		
테스트	서비스 테스트 및 모델 성능 평가			■	■	■	■		
종료	프로젝트 결과 정리 및 발표 준비							■	
오프라인 미팅계획	중간 점검					■			
	최종점검						■		
	멘토링, 기술 리뷰	■	■	■	■	■	■	■	

#### 나. 의사소통방법

##### 1. 팀원 간 커뮤니케이션 방법

- **Slack**을 통한 과업별 채널 개설 및 의사소통  
 목표 : 파트별로 소통 채널을 분리하여 논의 내용을 체계적으로 관리하고, 필요한 정보를 빠르게 공유  
 내용 : 기능 개발, 디자인, 문서 작성, 이슈 대응 등 과업별 채널을 개설하여 관련된 팀원들이 중심이 되어 소통하고, 공지사항 및 진행 상황을 공유
- 정기 온라인 회의 : **Discord** 활용  
 목표 : 주간 진행 상황을 공유하고 주요 이슈를 논의하며 차주 작업 계획을 수립  
 내용 : 온라인 회의를 통해 구현 현황, 일정 점검, 문제점 및 해결 방향을 함께 논의

- 비동기 진행 상황 공유 : **Jira, Confluence, Notion** 활용  
 목표 : 팀원별 개발 진행 상황과 문서화된 내용을 비동기로 확인하고 협업 효율을 높임  
 내용 : **Jira**를 통해 이슈 및 작업 상태를 확인하고, **Confluence**를 통해 회의록과 개발 문서를 공유하며, **Notion**을 통해 전체 일정과 주요 마일스톤을 관리

## 2. 협업 툴 및 프로젝트 수행방법

- 프로젝트 관리 및 이슈 추적 : **Jira**  
 내용 : 기능 단위 작업을 이슈로 등록하고, 담당자 배정, 우선순위 설정, 진행 상태 관리 수행  
 활용 : 백로그 관리, 작업 분배, 칸반 보드를 통한 진행 현황 시각화 및 이슈 관리
- 문서 작성 및 지식 공유 : **Confluence**  
 내용 : 회의록, 요구사항 정리, 기능 명세, 설계 문서, 협업 규칙 등을 문서화하여 공유  
 활용 : 프로젝트 진행 과정에서의 의사결정 사항과 개발 관련 자료를 체계적으로 축적 및 관리
- 일정 관리 : **Notion**  
 내용 : 전체 개발 일정, 주차별 목표, 마일스톤, 역할 분담 내용을 정리 및 공유  
 활용 : 프로젝트 일정 관리, 팀원별 해야 할 일 확인, 주요 일정 변경 사항 반영
- 실시간 커뮤니케이션 : **Discord**  
 내용 : 실시간 질의응답, 공지 전달, 회의 진행 및 빠른 피드백 공유  
 활용 : 과업별 채널을 통한 주제별 소통과 활용한 온라인 회의 진행
- 프로젝트 수행방법 : **Agile** 방식 기반 협업  
 내용 : 우선순위가 높은 작업을 선정하고 팀원별 담당 업무를 배정  
 활용 : 완료된 작업 결과를 확인하고 구현 상태를 점검
- 칸반 방식 보완 적용 : **Jira Board** 활용  
 진행 상태 : **TODO, In Progress, Review, Done**  
 활용 : 각 작업의 현재 상태를 한눈에 파악하고 병목 구간을 빠르게 확인하여 일정 관리를 효율화

## 다. 프로젝트 **Ground Rule** (기본 원칙)

### 1. 프로젝트 수행 기본 원칙

- 주 단위 스프린트 운영  
 매주 스프린트 목표를 설정하고, 기간 내 완료할 작업을 팀 단위로 정리한다.  
 스프린트 종료 시에는 완료 결과와 미진한 사항을 함께 점검하고, 다음 주 우선순위를 다시 조정한다.
- 데일리 스크럼 운영  
 짧은 일일 공유를 통해 각자의 진행 상황을 확인한다.

어제 진행한 내용, 오늘 진행할 작업, 현재 막히는 부분을 중심으로 간단히 공유한다.

- 주간 진행 상황 점검  
주 1회 정기 회의를 통해 전체 진행 흐름을 확인한다.  
주요 작업 결과, 일정 지연 여부, 해결이 필요한 이슈, 다음 주 계획을 함께 정리한다.
- 의사결정 원칙  
가능한 경우 충분한 논의를 거쳐 공동으로 방향을 정한다.  
선택지가 나뉘는 경우에는 팀 내 합의된 기준에 따라 우선순위를 정한다.  
일정이나 역할 조정이 필요한 상황에서는 팀장이 최종적으로 정리한다.

## 2. 커뮤니케이션 및 협업 원칙

- 협업 도구 사용 원칙
  1. 업무 관련 논의는 정해진 협업 채널을 중심으로 진행하여 기록이 남도록 한다.
  2. 회의 결과, 일정 변경, 주요 결정 사항은 문서로 정리하여 팀원 모두가 확인한다.
  3. 진행 중인 문제는 늦지 않게 공유하여 일정 차질을 줄인다.
  4. 구두로 합의한 내용도 이후 문서 또는 이슈로 남겨 혼선을 방지한다.
- **JIRA** 이슈 관리 원칙
  1. 작업은 **JIRA** 이슈 단위로 등록하고 담당자와 진행 상태를 명확히 한다.
  2. 개발, 수정, 개선 사항은 모두 이슈와 연결하여 관리한다.
  3. 작업 흐름과 일정 추적은 **JIRA**를 기준으로 통합 관리한다.
- **GitHub Pull Request** 운영 원칙
  1. 코드 변경 사항은 **Pull Request**를 통해 공유하고 검토 후 반영한다.
  2. PR에는 관련 **JIRA** 이슈를 연결하여 어떤 작업과 연관된 변경인지 알 수 있도록 한다.
  3. 필요한 경우 실제 작업 내역을 기록한 이슈를 함께 남겨 추적 가능하도록 한다.
  4. 리뷰 과정에서 확인된 수정 사항은 반영한 뒤 병합한다.
- 피드백 원칙
  1. 피드백은 단순 지적보다 작업 품질을 높이는 방향으로 주고받는다.
  2. 문제가 발견되면 왜 문제가 되는지와 함께 수정 방향을 같이 제안한다.
  3. 서로 다른 의견이 있을 때에는 개인적인 판단보다 근거와 목적을 기준으로 논의한다.

## 3. 프로젝트 산출물 관리

- 산출물 관리 원칙
  1. 회의록, 기획 내용, 작업 규칙 등 문서 산출물은 **Confluence**에 정리하여 관리한다.
  2. 코드와 개발 결과물은 **GitHub** 저장소를 기준으로 버전 이력을 유지한다.
  3. 이슈, PR, 문서 기록이 서로 연결되도록 관리하여 작업 흐름을 추적할 수 있게 한다.
  4. 주요 산출물은 주기적으로 점검하고 정리하여 누락을 줄인다.

## 라. 역할 분배

이름	분야	담당 역할
----	----	-------

김선우	백엔드 + 인프라	클라우드, 백엔드, 아키텍처 설계
김경태	프론트엔드	앱 및 웹 프론트엔드
김채영	AI + 백엔드	데이터 수집 및 전처리, 백엔드
심명선	AI	데이터 수집 및 전처리, 모델 훈련
윤여찬	인프라	클라우드, 아키텍처 설계, CI/CD

### III. 기대효과 및 활용분야

#### 1. 기대효과

##### 가. 작품의 기대효과

- 허위정보로 인한 사회적 혼란과 정보 불신 완화

온라인과 SNS를 통해 검증되지 않은 정보가 빠르게 확산되면서 사회적 혼란과 불신이 심화되고 있으므로, 본 프로젝트는 일반 사용자가 뉴스의 진위를 보다 손쉽게 확인할 수 있는 환경을 제공함으로써 허위정보 확산으로 인한 부정적 영향을 완화하는 것을 목표로 한다. 특히 단순한 진위 판별을 넘어 판단 근거를 함께 제공하는 방식은 정보 수용 과정에서의 막연한 불안과 불신을 줄이고, 보다 책임 있는 정보 소비 문화 형성에 기여할 것으로 기대된다.

- 기존 수작업 중심 팩트체킹의 한계를 보완하는 대국민 검증 접근성 확대

기존 인력 중심 팩트체킹은 허위정보의 양과 확산 속도를 실시간으로 따라가기 어렵고, 이미 확산된 이후에 대응하는 구조적 한계를 가진다. 본 프로젝트는 이러한 한계를 보완하여 일반 시민이 일상적으로 활용할 수 있는 뉴스 검증 접근성을 높이고, 전문가 중심의 제한적 검증 체계를 대중 친화적 검증 환경으로 확장하는 것을 목표로 한다. 이를 통해 팩트체킹이 특정 기관이나 전문가 집단에 머무르지 않고, 보다 폭넓은 이용자에게 실질적인 도움을 주는 방향으로 활용될 수 있도록 하고자 한다.

- 경제·정책·사회 이슈 분야의 허위정보 피해 예방 기반 마련

가짜뉴스는 단순한 정보 왜곡을 넘어 사회적·경제적 피해를 유발하며, 특히 주식시장이나 정책 이슈처럼 신속한 판단이 필요한 영역에서는 그 파급력이 더욱 크게 나타날 수 있다. 본 프로젝트는 이러한 분야에서 허위정보로 인한 판단 왜곡과 피해를 줄일 수 있는 검증 기반을 마련하는 것을 성과목표로 한다. 향후에는 대국민 뉴스 검증 서비스를 바탕으로 공공정책, 사회 현안, 경제 이슈 등 다양한 영역으로 활용 범위를 확대하여 실질적 사회 기여도를 높이고자 한다.

##### 나. 참여 멘티의 교육적 기대효과

- 최신 언어모델(LLM) 및 설명가능한 AI(XAI) 실무 적용 역량 확보

단순한 텍스트 분류를 넘어, PyTorch, HuggingFace 등 핵심 프레임워크를 활용해 오픈소스 LLM을 Fine-tuning하고, 뉴스 데이터의 문맥과 논리적 모순을 심층 분석하는 고도화된 AI 모델링 역량을 키울 수 있다. XAI 기술을 적용해 AI의 판단 근거를 시각적으로 설명하는 알고리즘을 구현한다. 이를 통해 사용자가 납득할 수 있는 신뢰성 높은 AI 서비스를 설계하는 최신 연구 및 트렌드를 체득한다.

- 대규모 데이터 파이프라인 구축 및 MLOps 기반 실전 아키텍처 경험

실시간으로 쏟아지는 뉴스 기사와 웹 데이터를 크롤링하고, 정규표현식 및 형태소 분석을 통해 노이즈를 제거하여 AI가 학습할 수 있는 최적의 형태로 정제하는 데이터 엔지니어링

기술을 학습한다. 로컬 환경에서의 단순 모델 테스트에 그치지 않고, GPU 기반 클라우드 환경에 Docker 등을 활용해 AI 모델을 배포한다. 이를 웹/모바일 플랫폼 및 API와 연동하는 MLOps 파이프라인을 구축하며, 대용량 트래픽을 고려한 백엔드 시스템 아키텍처 설계 능력을 기른다.

- 일상 속 사회적 문제(가짜뉴스)를 기술로 해결하는 실증(PoC) 경험

누구나 일상적으로 노출되고 피해를 입을 수 있는 '가짜뉴스'라는 심각한 사회 현안을 AI 기술을 통해 직접 분석하고 방어해 내는 실증 경험을 한다. 단순히 정해진 코딩 기술을 배우는 데 그치지 않고, 대국민 웹/모바일 검증 서비스로 직접 기획하고 구현해냄으로써 기술의 사회적 가치를 창출하는 문제 해결형 개발자로 성장한다. 실제 사용자 관점에서 UX/UI를 고민하고, 판별 결과를 직관적으로 전달하는 서비스 운영 감각을 익힐 수 있다.

## 2. 활용분야

- 정부 부처 및 지자체의 재난/위기 대응 모니터링 시스템

국가적 재난 상황이나 중대한 공공 정책 발표 시에는 대중의 불안 심리를 자극하는 유언비어와 가짜뉴스가 급증하여 막대한 사회적 혼란과 비용을 초래한다. 본 프로젝트의 실시간 트렌드 기반 가짜뉴스 판별 기술은 정부 부처 및 지자체가 악성 루머의 확산 조짐을 선제적으로 감지하고 대응하기 위한 핵심 모니터링 도구로 활용될 수 있다.

(예시) 질병관리청이나 재난안전대책본부의 위기관리 상황실, 혹은 지자체의 정책 대시보드에 본 시스템이 도입된다면 SNS와 커뮤니티에서 확산되는 미확인 재난 정보를 조기에 포착할 수 있다. 나아가 AI가 RAG를 통해 찾아낸 기존 팩트와 논리적 모순을 바탕으로 신속하고 정확한 해명 보도자료를 배포함으로써, 대국민 불안을 조기에 진화하고 불필요한 행정력 낭비를 막는 강력한 실무적 효과를 기대할 수 있다.

- 대형 포털 및 언론사의 뉴스 검증 자동화 및 신뢰성 강화

포털 사이트나 언론사처럼 하루에도 수만 건의 기사와 정보가 끊임없이 쏟아지는 환경에서는 사람이 일일이 그 진위를 검증하는 것이 물리적으로 불가능에 가깝다. 본 프로젝트의 AI 기반 가짜뉴스 판별 기술은 이러한 대형 플랫폼의 뉴스 어뷰징 모니터링 및 팩트체크를 위한 핵심 솔루션으로 활용될 수 있다.

(예시) 네이버·카카오 등 대형 포털 사이트의 모니터링 시스템이나 언론사 팩트체크 전담팀의 초기 검증 보조 툴로 본 서비스가 도입된다면, 실시간 트렌드 기사 중 논리적 모순이 있는 데이터를 1차로 자동 필터링할 수 있다. 결과적으로 판별된 의심 확률과 반박 근거를 시스템 상에 경고 라벨 형태로 제공함으로써, 플랫폼 운영자는 모니터링에 소요되는 막대한 인건비와 시간을 획기적으로 절감하고, 궁극적으로 플랫폼 자체의 정보 신뢰성을 구축하는 데 실질적인 도움을 줄 수 있다.

- 금융·증권 시장의 정보 왜곡 방지 및 투자자 보호

주식 시장이나 가상화폐 시장처럼 정보의 확산 속도가 즉각적인 가격 변동으로 이어지는 분야에서는 검증되지 않은 찌라시나 가짜뉴스가 치명적인 경제적 손실을 유발한다. 본

프로젝트의 실시간 자동 수집 및 RAG 기반 팩트체크 기술은 이러한 금융 정보 서비스 환경에 핵심적인 보조 지표로 활용될 수 있다.

(예시) 증권사의 트레이딩 시스템 내 시황 분석 탭이나 금융 정보 커뮤니티에 본 서비스가 도입된다면, 특정 기업이나 거시 경제에 대한 급상승 키워드를 실시간으로 모니터링하여 허위 정보를 신속하게 걸러낼 수 있다. 결과적으로 투자자들이 가짜뉴스로 인한 패닉 셀링 등 비합리적인 결정을 내리는 것을 방지하고, 객관적 데이터에 기반한 안전한 투자 환경을 조성하는 데 도움을 줄 수 있다.