

# 오픈 소스 및 데이터를 활용한 설명가능한 의사결정 지원 인공지능 기술 연구

(4세부) 전문지식 대상 판단결과의 이유/근거를 설명가능한 전문가 의사결정 지원 인공지능 기술 개발

---

2024. 10. 15.

배 경 만

언어지능연구실 / 한국전자통신연구원



# CONTENTS

01 과제 개요 및 연구 결과

02 연구 내용 및 오픈소스 활용 현황





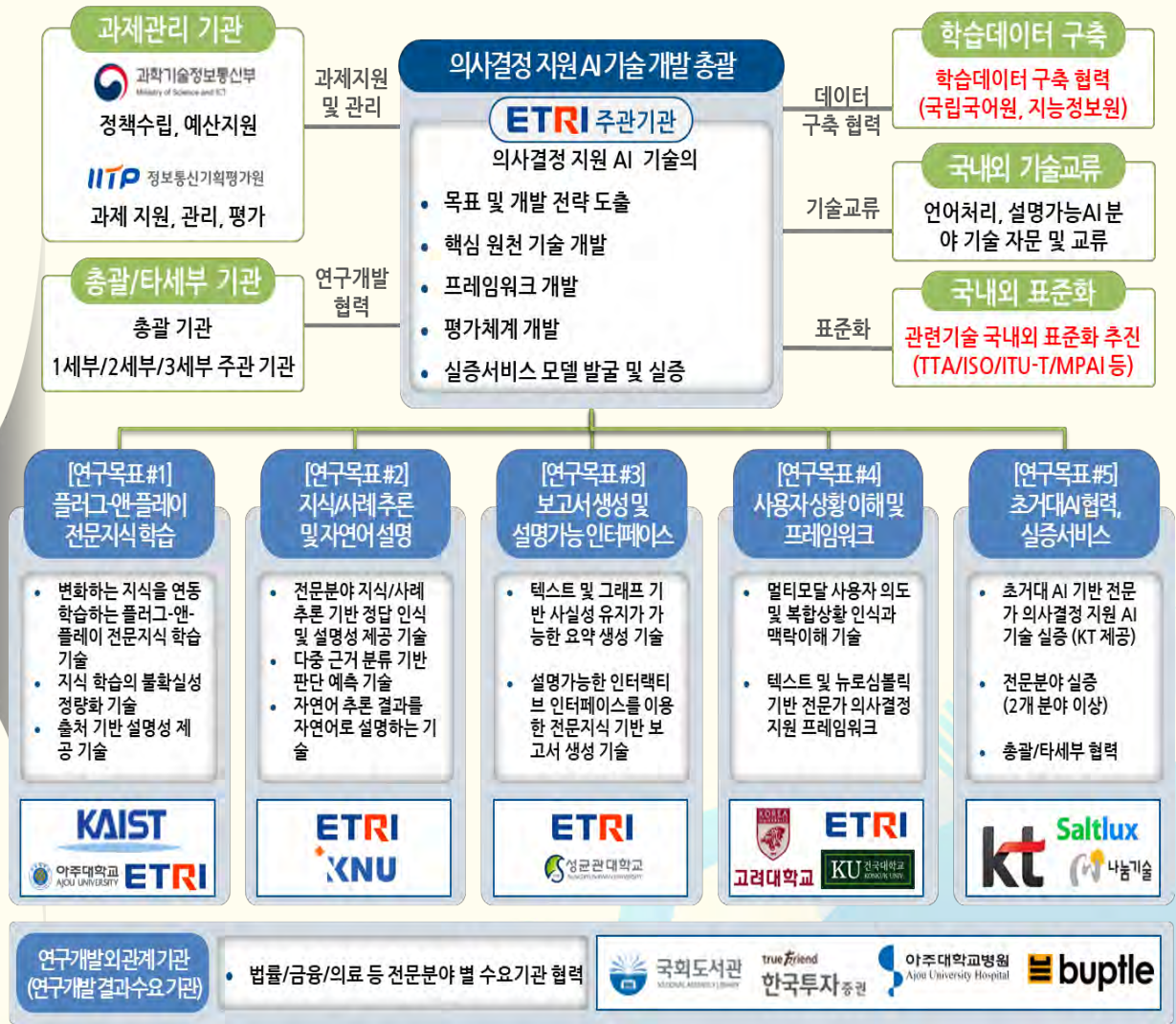
# 01

## 과제 개요 및 연구 결과

# 과제 개요

과제명	(4세부) 전문지식 대상 판단결과의 이유/근거를 설명가능한 전문가의사결정 지원 인공지능 기술 개발			
총괄 과제명	(총괄) 사용자 맞춤형 플러그앤플레이 방식의 설명가능성 제공 기술 개발			
기술 준비도 (TRL)	시작: TRL 3 → 종료: TRL 6 (RFP 요구 5단계 대비 목표 상향 설정)		표준 화 연 계	표준화 비대상 과제 이나 표준화 추진
			주관, 공동연구기관	
	<p>-주관기관: 한국전자통신연구원</p> <p>-공동/위탁 연구기관: 대학 6, 기업 3 KAIST, 고려대, 성균관대, 전국대, 아주대, 강원대, (주)KT, (주)솔트룩스, (주)나눔기술</p> <p>-연구개발 외 관계기관: 국회도서관, 한국투자증권, 아주대병원, (주)법틀</p>			
총연구기간 당해년도기간	2022. 4. 1. ~ 2026. 12. 31. (4년 9개월) 2024. 1. 1. ~ 2024. 12. 31. (12개월)		총괄책임자	배 경 만
투입재원	예산 (당해년도)	20.9억원	인력 (당해년도)	99명
	예산 (총연구기간)	110.9억원	인력 (총연구기간)	495명

[과제 개요]



[추진 체계]

# 과제 개념 예시

## 전문지식 기반 자연어 설명을 제공하는 인공지능 SW

다양한 인공지능 사용 ->  
mission  
critical topic



[자율주행차]



자율주행자동차  
정지

판례 연혁 위임행정규칙 규제

### 도로교통법

[시행 2022. 4. 20.] [법률 제18491호, 2021. 10. 19., 일부개정]

경찰청(교통기획계(법제총괄)), 02-3150-2251  
경찰청(교통운영계(신호, 안전표지, 도로공사신고)), 02-3150-2753  
경찰청(운전면허계(운전면허)), 02-3150-2253  
경찰청(교통안전계(안전, 단속, 어린이통학버스)), 02-3150-2252  
경찰청(교통조사계(교통사고조사)), 02-3150-2552

### 제1장 총칙

판례 연 제1조(목적) 이 법은 도로

**제27조(보행자의 보호)** ① 모든 차 또는 노면전차의 운전자는 보행자(제13조의2제6항에 따라 자전거등에서 내려서 자전거등을 끌거나 들고 통행하는 자전거 등의 운전자를 포함한다)가 횡단보도를 통행하고 있거나 통행하려고 하는 때에는 보행자의 횡단을 방해하거나 위험을 주지 아니하도록 그 횡단보도 앞(정지선이 설치되어 있는 곳에서는 그 정지선을 말한다)에서 일시정지하여야 한다. <개정 2018. 3. 27., 2020. 6. 9., 2022. 1. 11.>  
② 모든 차 또는 노면전차의 운전자는 교통정리를 하고 있는 교차로에서 좌회전이나 우회전을 하려는 경우에는 신호기 또는 경찰공무원등의 신호나 지시에 따라 도로를 횡단하는 보행자의 통행을 방해하여서는 아니 된다. <개정 2018. 3. 27.>



# 과제 배경

## 팩트체크 시스템에 설명가능한 AI 기술 적용 연구 실세계 지식 제공 부재로 전문가 업무능력 미향상 AI 모델은 현실 지식에 기반한 판단 결과 및 설명성 제공 필요



Volume 2, Issue 4  
Special Issue: DARPA's  
Explainable Artificial  
Intelligence (XAI) Program  
December 2021  
e49

### APPLIED AI LETTERS

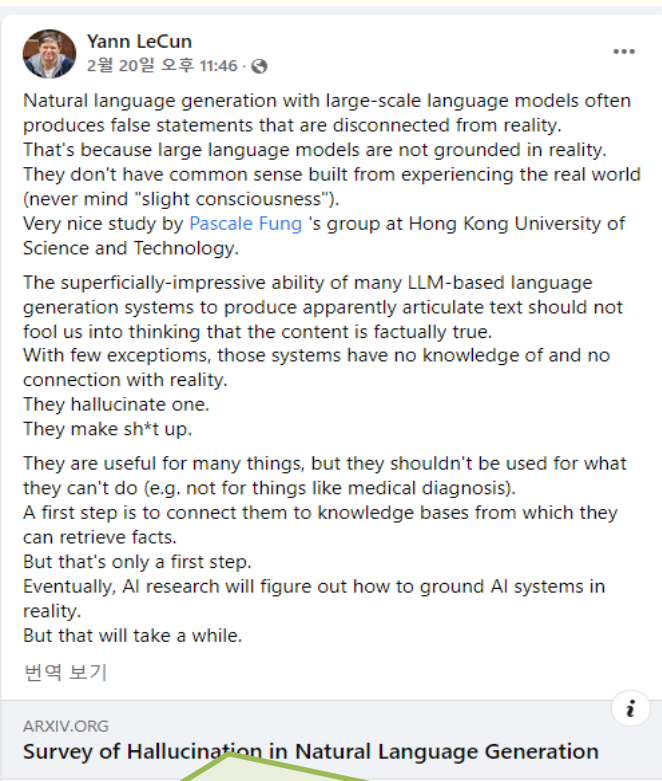
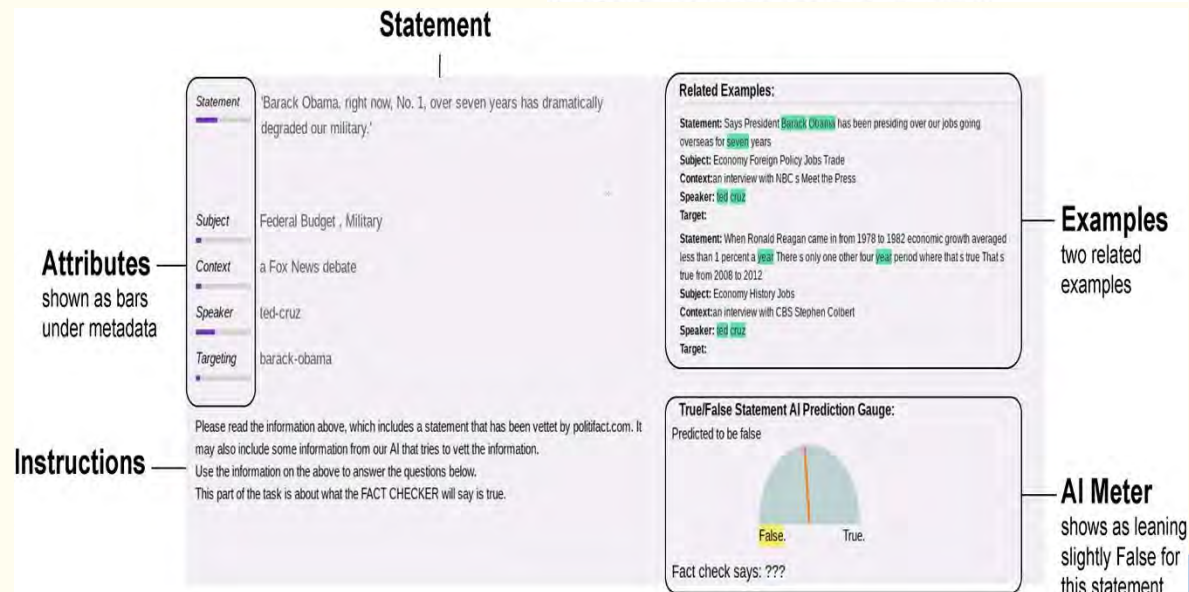
LETTER | Open Access | CC BY

How level of explanation detail affects human performance in interpretable intelligent systems: A study on explainable fact checking

Rhema Linder, Sina Mohseni, Fan Yang, Shiva K. Pentiyala, Eric D. Ragan, Xia Ben Hu

First published: 08 November 2021 | <https://doi.org/10.1002/aii2.49> | Citations: 1

Funding information: DARPA, Grant/Award Number: N66001-17-2-4031



AI모델의 출력 결과가 현실(reality)에 기반하지 않고 있다는 문제점 제기 (Yann LeCun, '22.02.20)

# 과제 배경



- KAIST 주관 XAI 과제('17-'21): 시계열·영상·의료 데이터 등에서 설명성 연구
  - 주요성과: 시계열 데이터 분석, 이미지 생성 경계 분석, 스마트 고로 등
- XAI 중요성에 따라 '사람 중심 AI 원천기술사업'으로 추진: **분야의 확장 필요성**



- DARPA 주관 XAI 과제('17-'21): 설명가능한 모델·인터페이스 정립(학습성능 유지)
  - 주요성과: XAI 개념 및 목표 발굴, 11개 팀에서 다양한 분야의 XAI 연구
- DARPA XAI Retrospective: **실세계 적용 필요성**
  - In order for explanations to improve user task performance, the task must be difficult enough that the AI explanation helps.



- EU 주관 AI4EU 과제('19-'20): 실세계 AI 플랫폼·생태계 구축(AI 주도권 확보)
  - 주요분야: Explainable AI, Verifiable AI, Integrative AI, Collaborative AI, Physical AI
- 설명 가능한 AI에서 나아가 검증 가능한 AI(Verifiable AI) 연구: **설명의 검증 필요성**
  - Important human decisions often require explanation, conferring, or second opinions, and are subject to appeals mechanisms, audits, and other accountability structures.

## 시사점

실세계 문제와 현실 지식에 기반한 판단 결과 및 대안의 설명 제공 필요  
XAI for Machine Learning(레벨1) → XAI for Machine Reasoning(레벨2)

# 연구목표 및 과제 개념

구글 딥마인드는 **AGI를 5단계로 구분**  
 챗GPT, 바드(Bard), 라마2(Llama) 등이 초기 AGI에 해당  
 AGI 수치가 판단 가능한 **벤치마크 'GAIA'**(메타, 허깅페이스 등)  
 사람과 같은 역량에 중점. Level 1~3 평균 **사람은 92%, GPT-4+plugins는 15% 수준**



\* "인공 일반지능(AGI)" 혹은 "강 인공지능(Strong AI)"이란 인간과 비슷한 수준의 능력 혹은 지식을 갖춘 인공지능을 의미함  
 자료: 카카오 엔터프라이즈(2023. 05) [IT TREND] AGI를 향한 기대와 우려

출처: 2023 생성형 AI 주요 이슈와 의미 보고서, KISDI(정보통신정책연구원)

**Position: Levels of AGI for Operationalizing Progress on the Path to AGI**

Meredith Ringel Morris<sup>1</sup> Jascha Sohl-Dickstein<sup>2</sup> Noah Fiedel<sup>2</sup> Tris Warkentin<sup>2</sup> Allan Dafoe<sup>3</sup>  
 Aleksandra Faust<sup>2</sup> Clement Farabet<sup>3</sup> Shane Legg<sup>3</sup>

Google DeepMind

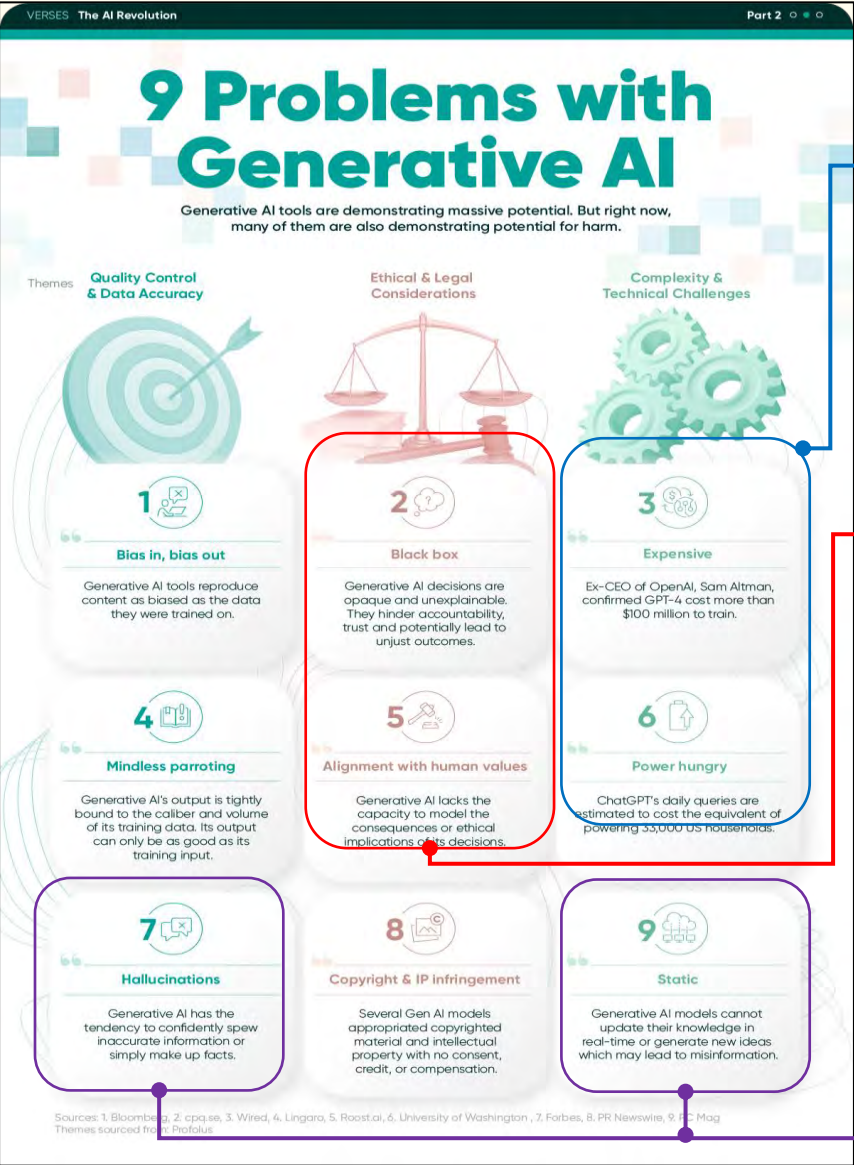
**AGI 성능 구분 5단계**

단계	특정기능 수행(협의의 AI)	범용기능 수행(AGI)
레벨1: 인공 (Artificial)	인도어 글자 가독 시스템 자율주행 차량 프로그래밍 SHARPLU	AGI 성능 50%
레벨2: 경쟁력 (Competent)	예측 시리, 이미지 캡션, 구글 어시스턴트 등 스마트 스피커 같은 현재의 특정, 간단한 응용 분야 SOTA LLM 특정 컴퓨터 프로그램	
레벨3: 전문가 (Expert)	문헌 일사기 Grammarly 이미지 캡션 등 간단한 이미지 캡션	
레벨4: 인간 (Human)	단일 작업 완료	
레벨5: 초인간 (Superhuman)	단일 작업 3D 구조를 생성하는 장악물도 바둑 세스 등 모든 알고리즘을 포함한 일차적으로 테스트 결과 등 가장 뛰어난 성능을 보일 수 있음	

출처: <https://www.epnc.co.kr/news/articleView.html?idxno=239045>



# 연구목표 및 과제 개념



### 이코리야

막대한 AI 비용에 골머리 앓는 빅테크

[이코리야] AI 챗봇의 막대한 운영 비용이 관련 산업 발전의 최대 걸림돌로...

지난 4월에는 오픈 AI가 챗 GPT를 운영하는 데 하루에 70만 달러(약 90억 원)를...

생성형 AI의 환경비용 급증, 공공연한 비밀이다

엄청난 물과 에너지 비용을 누가 감당하나

### nature

EDITORIAL | 25 July 2023

## ChatGPT is a black box: how AI research can break it open

Despite their wide use, large language models are still mysterious. Revealing their inner workings is urgent and important.

### FAST COMPANY

05-23-2024 | AI DECODED

## Anthropic takes a look into the 'black box' of AI models

Progress in mechanistic interpretability could lead to major advances in making large AI models safe and bias-free.

### OpenAI

Documentation API reference

## Gpt-4o hallucinates a location in Ringsted, Klosterparken Allé 10.

We found that gpt-4o is consistently hallucinating a location in Ringsted, Klosterparken Allé 10, Denmark.

### The JoongAng

"환각 줄인 생성AI, 누가 잘하나" 올거나이즈, 미니투데이 | 고석용 기자

2024.05.29 13:00

## 리더보드 공개

팀명	점수	평가	Average
0.95 (5/100)	0.916 (5/100)	0.833 (5/100)	0.899 (247/290)
0.97 (5/100)	0.893 (5/100)	0.833 (5/100)	0.899 (248/290)
0.911 (5/100)	0.893 (5/100)	0.718 (4/100)	0.864 (220/290)
0.898 (5/100)	0.833 (5/100)	0.833 (5/100)	0.854 (244/290)

- 고비용 문제(효율성)
  - Task 특화 모델
  - 디코더/인코더 기반 모델
- 블랙박스 문제(설명성)
  - 설명가능한 인공지능(XAI)
  - sLLM 기반 쉬운 설명 생성
- 환각 문제(사실성)
  - 사실검증(KorFactScore)
  - 검색증강생성(RAG)

# 연구목표(전문분야 - 법률/금융(AICC)/행정)



이미지 출처: <https://www.hankyung.com/article/2023110157611>

국내

△ 판례검색 △ 법률정보 제공

해외

△ 변호사 검색 △ 법률문서 자동화 △ 실무수습 관리  
△ 판례분석 및 전략수립 △ 법률정보 제공 △ 준법감시  
△ 법률교육 △ 온라인 분쟁해결 △ 전자증거개시 등



이미지 출처: <https://www.ajunews.com/view/20230101131021046>

상담 지원 AI 기술

상담 분류

상담 요약

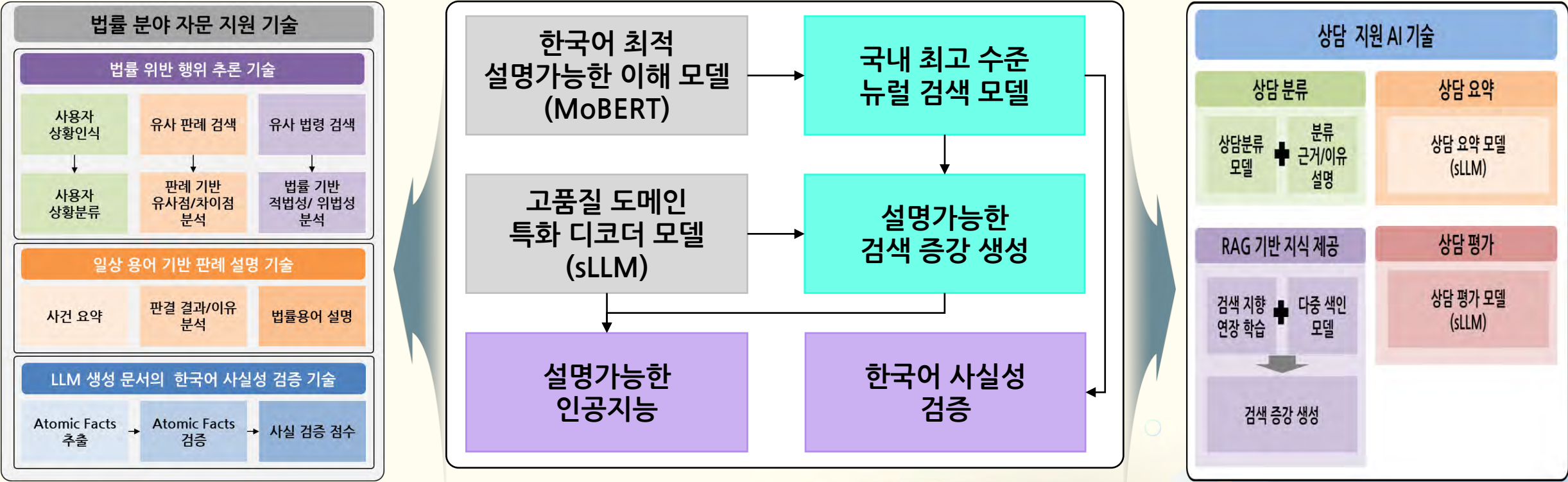
RAG 기반 지식 제공

상담 평가



# 연구목표(전문분야 - 법률/금융(AICC)/행정)

전문 분야 지식 기반 설명가능한 인공지능 기술 개발  
법률(법률 종사자 지원) / 금융(전문 상담사 지원)



# 연구목표(통합 개발 기술 - 법률 분야)

## 법률 분야 자문 지원 기술

### 법률 위반 행위 추론 기술

사용자  
상황인식

유사 판례 검색

유사 법령 검색

사용자  
상황분류

판례 기반 유사  
점/차이점 분석

법률 기반 적법  
성/ 위법성 분석

### 일상 용어 기반 판례 설명 기술

사건 요약

판결 결과/이유  
분석

법률용어 설명

### LLM 생성 문서의 한국어 사실성 검증 기술

Atomic Facts  
추출

Atomic Facts  
검증

사실 검증 점수

## 법률 위반 행위 추론 기술

### Question:

여자 친구와 교제하다가 다툼이 있어 헤어졌었다. 사과하기 위해 집을 찾아 갔습니다. 불러도 응답이 없어 이미 알고 있는 비밀번호로 사용하여 집에 들어갔습니다.

### Situation:

다들 후 헤어진 여자 친구를 사과하기 위해 집을 찾아 갔던 남자는, 비밀번호를 알고 있어 집에 들어갔습니다.

### Case Name:

주거침입

### Case IR / Law Result:

피고인은 2020. 9. 28. 02:00경 서울 중랑구 B, C동 앞에서 피해자 D(여, 32세)이 이별을 통보하자, 그녀를 만나기 위해 평소 알고 있던 공동출입문 비밀번호를 눌러 빌라 안으로 들어가, 피해자의 주거지인 E호 현관문 초인종을 눌렀으나, 피해자가 문을 열어주지 않자, 4층 복도 창문을 통해 난간을 밟고 E호 베란다 안으로 들어가 피해자의 주거에 침입하였다. (Score: 21.384014129638672)

관련 법령: 주거침입, 퇴거불응 (형법 제319조): ①사람의 주거, 관리하는 건조물, 선박이나 항공기 또는 점유하는 방실에 침입한 자는 3년 이하의 징역 또는 500만원 이하의 벌금에 처한다. <개정 1995. 12. 29.> ②전항의 장소에서 퇴거요구를 받고 응하지 아니한 자도 전항의 형과 같다.

피고인은 약 10개월 전에 피해자의 아내와 교제했던 사람이다. 피고인은 2020. 12. 20. 00:20경 피해자 B(남, 38세)이 거주하고 있는 안산시 상록구 C에 이르러 헤어진 여자 친구이자 피해자의 아내인 D를 만나기 위해, 이전에 D이 알려주어 알게 된 공동현관문 비밀번호를 누르고 건물 안으로 들어가, 옥상까지 올라가는 방법으로 피해자가 거주하고 있는 주거에 침입하였다. (Score: 21.367164611816406)

관련 법령: 주거침입, 퇴거불응 (형법 제319조): ①사람의 주거, 관리하는 건조물, 선박이나 항공기 또는 점유하는 방실에 침입한 자는 3년 이하의 징역 또는 500만원 이하의 벌금에 처한다. <개정 1995. 12. 29.> ②전항의 장소에서 퇴거요구를 받고 응하지 아니한 자도 전항의 형과 같다.

### Inference Result:

#### 판례 분석

- 판례 1: 여자가 이별을 통보한 후 주거에 침입한 사건
- 유사성: 둘 다 과거에 교제했던 사람을 만나기 위해 주거지에 방문했다는 점.
  - 주거지 비밀번호를 알고 있었고 이를 활용하여 침입했다는 점.
  - 차이점: 판례 1에서는 피해자가 문을 열어주지 않자 창문을 통해 주거에 침입한 점이 추가적으로 고려된다.
  - 사용자의 경우 창문을 사용하지 않고 비밀번호를 통해 바로 들어갔다는 점.

#### 판례 2: 헤어진 여자친구의 주거에 침입한 사건

- 유사성:
- 교제했던 사람을 만나기 위해 주거지에 방문했다는 점.
  - 비밀번호를 알고 있었고 이를 통해 건물이나 주거지에 침입한 점.
  - 차이점: 판례 2에서는 피해자의 남편이 있는 주거지로 들어갔다는 점과 옥상까지 올라가 침입했다는 점이 추가적으로 고려된다.
  - 사용자 상황에서는 옥상 등을 사용하지 않고 비밀번호로 직접 주거에 들어간 점.

#### 법률 분석

##### 판례 1: 형법 제319조 주거침입, 퇴거불응

- "조항 내용": 사람의 주거, 관리하는 건조물, 선박이나 항공기 또는 점유하는 방실에 침입한 자는 3년 이하의 징역 또는 500만원 이하의 벌금에 처한다.
- "분석": 사용자는 여자 친구의 주거지에 그녀의 허락 없이 비밀번호를 사용해 들어갔으며, 이는 "사람의 주거에 침입"한 행위로 해석될 수 있다. 따라서 주거침입죄가 성립할 수 있다.

##### 판례 2: 형법 제319조 주거침입, 퇴거불응

- "조항 내용": 사람의 주거, 관리하는 건조물, 선박이나 항공기 또는 점유하는 방실에 침입한 자는 3년 이하의 징역 또는 500만원 이하의 벌금에 처한다. 전항의 장소에서 퇴거요구를 받고 응하지 아니한 자도 전항의 형과 같다.
- "분석": 사용자가 퇴거요구를 받지 않았다면, 퇴거불응죄는 성립하지 않는다. 하지만 침입 자체만으로 주거침입죄가 성립될 수 있다.

#### 결론

사용자의 상황은 주거침입죄에 해당할 가능성이 높습니다. 비밀번호를 알고 있었다고 주거지에 허락 없이 들어간 행위는 형법 제319조에 따라 주거침입으로 간주될 수 있습니다. 판례와 비교해도 유사한 점이 많아 법적 책임을 면하기 어렵다고 판단됩니다.

## 판례

## 일상 용어 기반 판례 설명 기술

### 판례 설명:

사건 요약: 채권자가 채무자 조합장의 해임을 안건으로 하는 임시 총회 개최를 발의하고, 이에 따라 임시 총회가 개최되어 조합장 해임 결의가 이루어졌습니다. 채권자는 해임 결의의 효력 정지를 신청했고, 1심 법원은 이를 기각했으나, 항소심 법원에서 인용되었습니다.

### 판결 결과:

항소심 법원은 채무자 조합장 해임 결의가 무효임을 판단하고, 이에 따라 결의의 효력 정지 가처분 신청을 인용했습니다.

### 핵심 이유:

조합 총회 소집 절차 위반: 조합 총회 개최 7일 전에 조합원들에게 소집 통지를 하지 않은 사실이 인정되었습니다.

결의 방식 위반: 도시정비법에서 허용하는 서면 결의 방식이 아닌 전자적 결의 방식으로 결의가 이루어졌습니다.

### 주요 용어:

조합: 도시 또는 주거 환경 정비 사업을 수행하기 위해 설립되는 법인  
조합원: 조합에 가입한 사람  
임시 총회: 조합원의 3분의 1 이상 또는 조합 자본의 3분의 1 이상을 소유하는 조합원의 발의로 소집되는 총회  
안건: 총회에서 논의하고 결의할 사항  
의결권: 총회에서 의견을 표하고 결의에 참여할 권리  
서면 결의: 서면으로 의견을 표하고 결의에 참여하는 방식  
전자적 결의: 전자 문서를 통해 의견을 표하고 결의에 참여하는 방식  
무효: 법률상 효력이 없는 것  
효력 정지: 일정한 기간 동안 법률상 효력을 발생하지 못하도록 하는 것  
가처분: 소송 계속 중 분쟁의 현상을 유지하거나 회복하기 위해 법원이 신청인의 신청에 따라 내리는 임시적인 결정

### 참고:

이 판례 해석은 법률 전문가가 작성한 것이 아닌, 인공지능 기반 법률 솔루션에 의해 자동 생성된 것임을 알려드립니다.  
법률 문제는 사건마다 상황이 다를 수 있으므로, 정확한 판단을 위해서는 반드시 법률 전문가와 상담하시기 바랍니다.

## LLM 생성 문서의 한국어 사실성 검증 기술

### 생성 문서의 사실 비율(%)

85.19% (True: 23, False: 4)

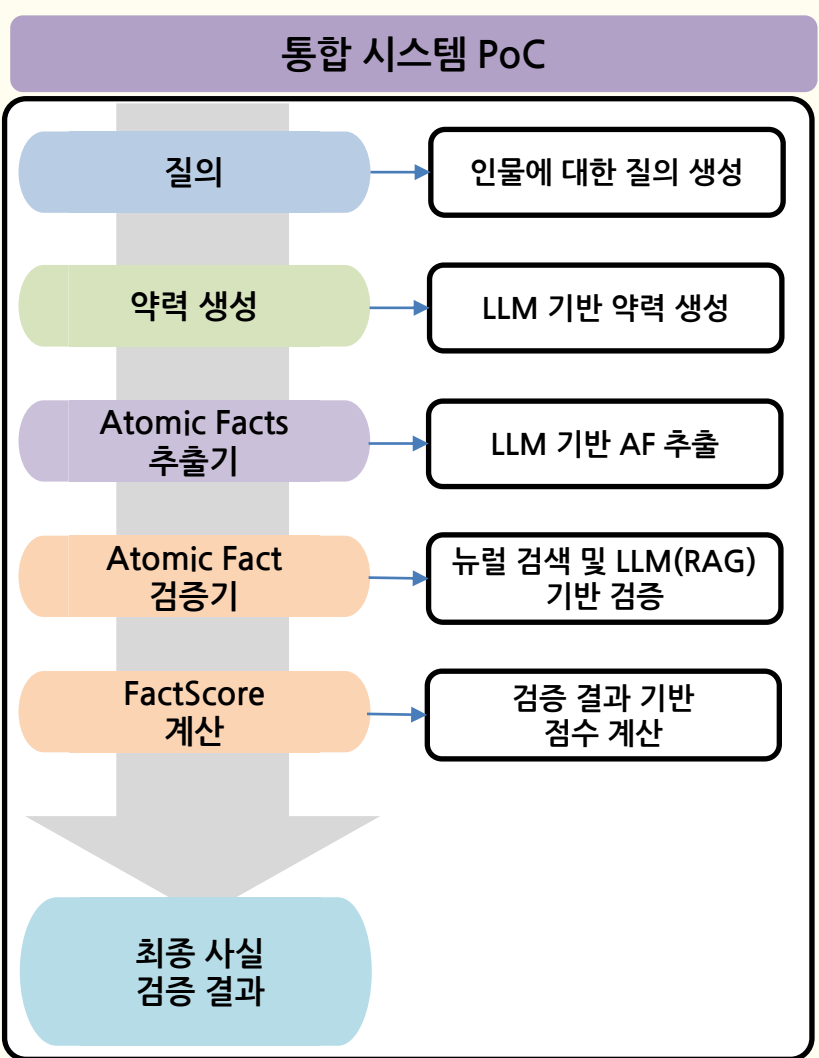
### Atomic Facts 검증 결과

- 장기하는 유자선입니다. TRUE
- 장기하는 가수입니다. TRUE
- 장기하는 작사가입니다. TRUE
- 장기하는 작곡가입니다. TRUE
- 장기하는 장기하와 얼굴들이라는 밴드의 전 리더입니다. TRUE
- 장기하는 장기하와 얼굴들의 전 보컬입니다. TRUE
- 장기하는 대한민국의 유자선입니다. TRUE
- 장기하와 얼굴들은 밴드입니다. TRUE
- 그는 1982년 2월 20일에 태어났습니다. TRUE
- 그는 음악 활동 외에도 다방면으로 활발하게 활동하고 있습니다. TRUE
- 장기하는 '프로젝트 밴드 헬로루키' 대회에서 최종 우승했습니다. FALSE
- '프로젝트 밴드 헬로루키' 대회는 예술의 전당 한가람디자인미술관에서 개최되었습니다. FALSE
- 장기하는 대회 우승을 통해 처음으로 대중 앞에 등장했습니다. FALSE
- 그 후, '장기하와 얼굴들'이라는 밴드를 결성했습니다. TRUE
- '장기하와 얼굴들'은 2008년 '싸구려 커피'를 발표했습니다. TRUE
- '싸구려 커피' 발표로 큰 주목을 받았습니다. TRUE
- 이 곡은 특유의 가사와 멜로디로 대중적 인기를 얻었다. TRUE
- 장기하와 얼굴들은 '2011 대한민국 대중문화예술상'에서 대중명상을 수상했다. FALSE
- 장기하와 얼굴들은 많은 성공을 거뒀다. TRUE



# LLM 생성 문서의 한국어 사실검증 기술

## LLM이 생성한 인물 약력에 대해 환각 문제 확인 및 사실 검증



### LLM 생성 한국어 문서 사실 검증 기술

한국전자통신연구원(ETRI)

**인물**

장기하

LLM이 생성한 약력

장기하는 대한민국의 뮤지션, 가수, 작사가, 작곡가이며, 장기하와 얼굴들이라는 밴드의 전 리더이자 보컬로 잘 알려져 있습니다. 그는 1982년 2월 20일에 태어났으며, 음악 활동 외에도 다방면으로 활발하게 활동하고 있습니다. 장기하는 예술의 전당 한가람디자인미술관에서 개최된 '프로젝트 밴드 헬로우키' 대회에서 최종 우승하며 처음으로 대중 앞에 등장했습니다. 그 후, '장기하와 얼굴들'이라는 밴드를 결성해, 2008년 '싸구려 커피'를 발표하며 큰 주목을 받았습니다. 이 곡은 특유의 가사와 멜로디로 대중적 인기를 얻었고, 장기하와 얼굴들은 '2011 대한민국 대중문화예술상'에서 대통령상을 수상하는 등 많은 성공을 거뒀습니다. 대중문화에 영향을 미친 장기하는 그의 독특하고 개성 넘치는 음악 스타일과 솔직하고 위트 있는 가사로 많은 팬을 사로잡았습니다. 그는 대한민국 인디 음악계에서 중요한 인물로 인정받으며, 다양한 장르를 시도하고 실험적인 음악을 추구합니다. 음악 활동 외에도, 장기하는 방송과 광고 등에서도 활약하며 자신만의 색깔을 드러내고 있습니다. '장기하의 대단한 라디오'를 진행하며 청취자들과 소통하기도 했으며, 여러 프로그램에 출연하여 다양한 매력을 발산했습니다. 그의 특징 중 하나는 일상에서 영감을 받아 재치 있고 유머러스한 가사를 쓴다는 것입니다. 장기하는 기존의 K-pop과는 다른 새로운 스타일로 대중 음악계에 새로운 바람을 일으켰으며, 한국 대중음악계에 큰 발자취를 남겼습니다. 작곡가, 가수로서의 장기하의 약력을 간략히 살펴보면, 그는 실력과 개성을 겸비한 아티스트로서 꾸준히 사랑받으며, 음악적으로 끊임없이 성장하고 있는 중입니다.

**생성 문서의 사실 비율(%)**

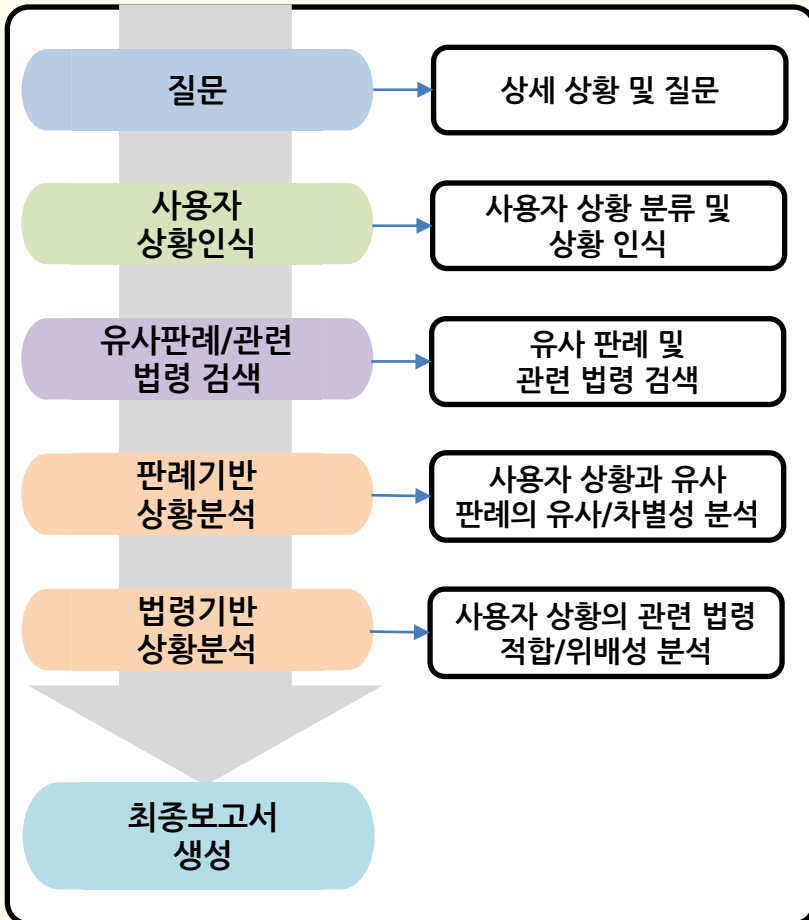
**Atomic Facts 검증 결과**

**사실 검증**

# 법률 위반 행위 추론 기술

자신의 **상황과 유사한 판례와 관련 법률 분석**을 통하여  
**상황에 따른 법률 위반 행위 추론**

## 통합 시스템 PoC

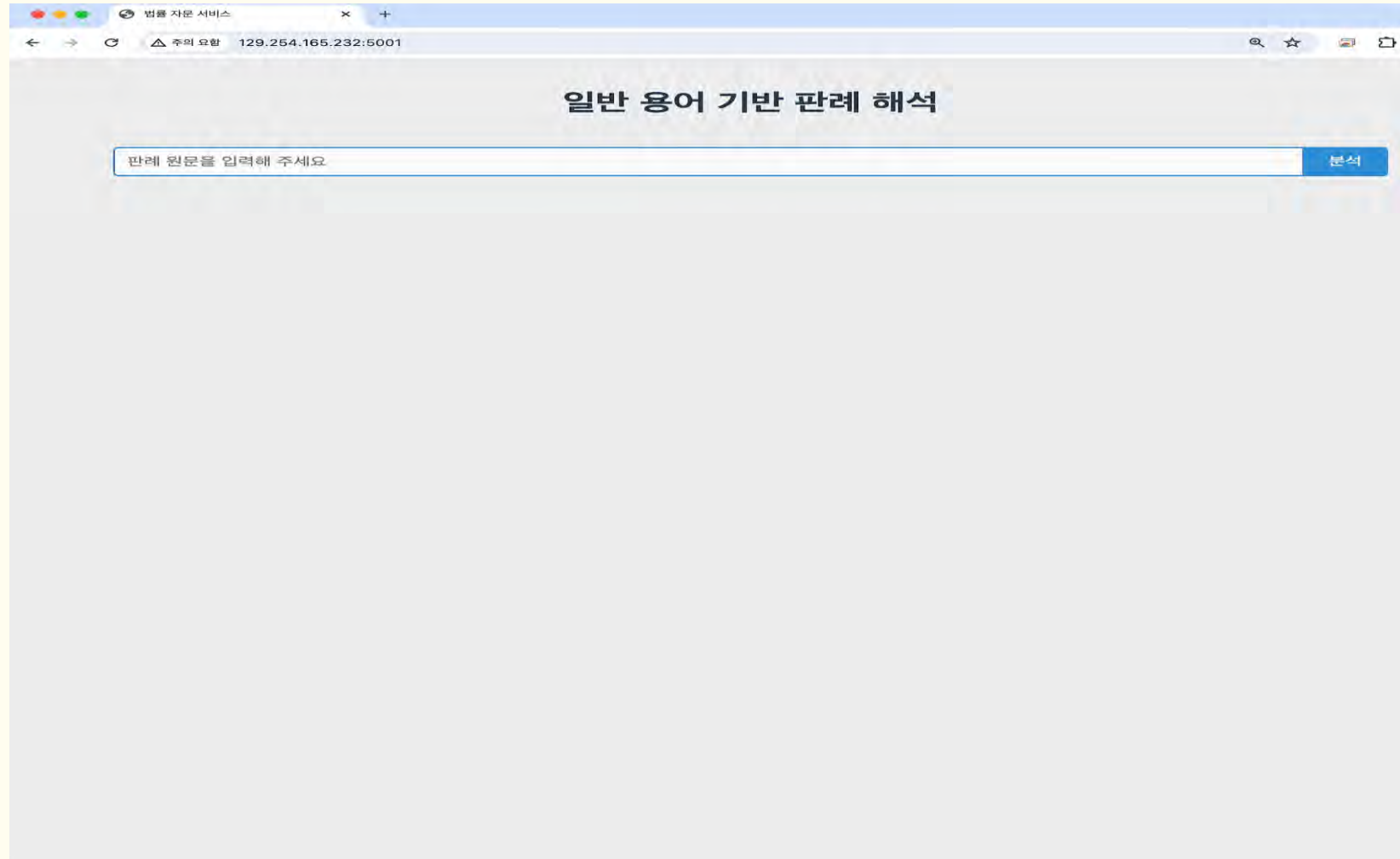
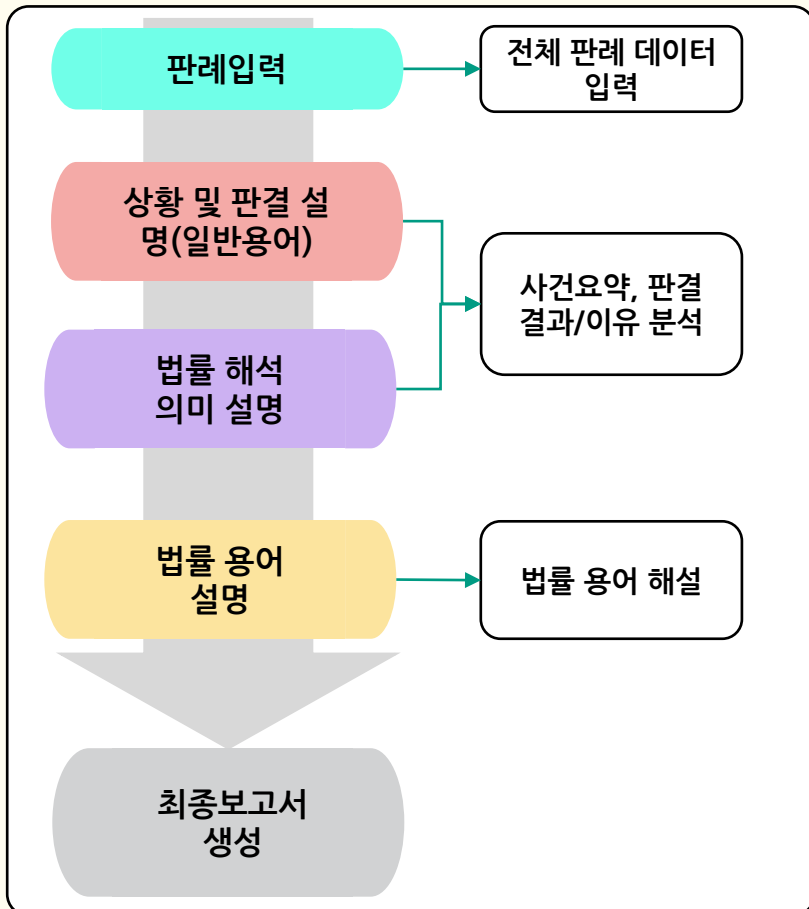


The screenshot shows a web browser window with the title '법률 자문 서비스'. The address bar shows the URL '129.254.165.232:5000'. The main content area has a title '법률 자문 서비스' and a text input field with the placeholder text '법률 자문을 위해 상황 및 질의 내용을 상세히 입력하세요'. Below the input field is a blue button labeled '분석'.

# 일상 용어 기반 판례 설명 기술

일반인이 이해하기 어려운 법률 용어 기반의 판례를  
이해하기 쉬운 일상 용어로 해설 및 용어 설명

## 통합 시스템 PoC



# 과제 개요(1단계 성과 요약)

연구개발/사업화/생태계 조성면에서 **우수한 성과 달성!**

## \* 한국어AI 유니콘 기업 육성(기술 출자 4.16억원+a)

- (주)\*팩은 기술출자를 통한 AICC 사업화 추진(기술가치 약 4.16억)
- (주)\*\*\*\*랩스는 기술 출자를 통해 교육분야 사업화 추진  
(연구소 기업 추진 중)

## \* 기술이전 2건 / 7천만원

- 한국어 특화 이해모델 기반 뉴럴 단락 검색 기술('24.02)  
. 자연어처리 모델을 이용한 IT 서비스 자동화 모델 개발에 활용
- 한국어에 최적화된 언어이해 모델 (MoBERT) 및 응용 기술('24.04)  
. 공정위 실증사업(법률 분류)에 활용 예정

## \* 공개SW를 통한 생태계 확산

- 연구를 통해 개발된 기술들은 Github를 통해 공개(32개 저장소)  
. Star 965 / Commit 661 / Fork 240 / issued(closed) 50 / pull request(closed) 7
- sLLM 학습 및 모델 공개(16개 모델 65,990건 다운로드)

## \* IP 확보를 통한 기술 확산

- 국제 표준 기고서 12건 제출 10건 채택 2건 승인
- 국제 특허 3건 출원 / 국내 특허 22건 출원 / SW 등록 11건

연구개발

기술이전.  
사업화

생태계  
조성

## \* 한국어 최고 수준 사업화 가능 기술 개발

고성능 뉴럴 검색 기반  
설명가능한 RAG 기술

고품질 도메인 특화  
디코더 모델(sLLM)

국내 최초 공개 한국어  
사실검증 기술

한국어 최적 설명가능한  
인코더 모델(MoBERT)

사용자가 이해하기  
쉬운 설명제공 기술

## \* 실증을 통한 기술 검증

- 사람과 교감하면서 대화가 가능한 로봇(ETRI)
- 법률 자문 서비스 / 일반 용어 판례 해설 서비스(자체)
- 행정 분야 설명가능한 문서 배부 RPA(나눔기술)
- AICC 상담사 지원 서비스(주)에프앤유신용정보)

## \* 세계 최고 원천 기술 연구(SOTA 기술)

- 국내외 학술대회 49편/학술지 19편
- Top-tier 10편, SCIE 14편(Top 7 mrrF 평균 96.49%)

뉴로-심볼릭 지식  
그래프 기술

다중 문서 지식  
추출 기술

Prompting 기반  
자연어 추론 기술

근거기반 불확실성  
추론 기술

설명가능한 오픈  
도메인 질의응답 기술

사용자 상황 이해를  
위한 근거 추론 기술





# 02

## 연구 내용 및 오픈 소스 활용 현황

# 한국어 최고 수준 사업화 가능 기술

고성능 뉴럴 검색 기반  
설명가능한 RAG 기술

고품질 도메인 특화  
디코더 모델(sLLM)

국내 최초 공개 한국어  
사실검증 기술

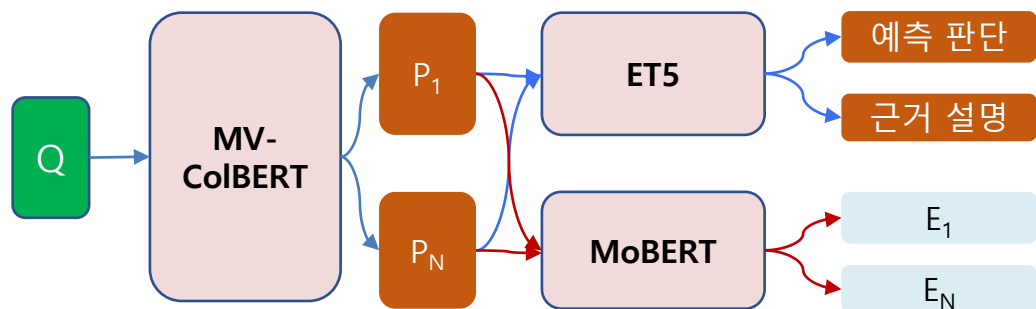
한국어 최적 설명가능한  
인코더 모델(MoBERT)

사용자가 이해하기  
쉬운 설명제공 기술



## 예측 판단과 판단의 이유 설명 모델 개발(ET5/sLLM)

- (1) ET-5모델(330M)을 이용, 예측 판단 및 판단 이유 설명, 근거인식(MoBERT)
- (2) 한국어 sLLM을 이용한 **법률 자문 및 판례 해설 보고서 생성**



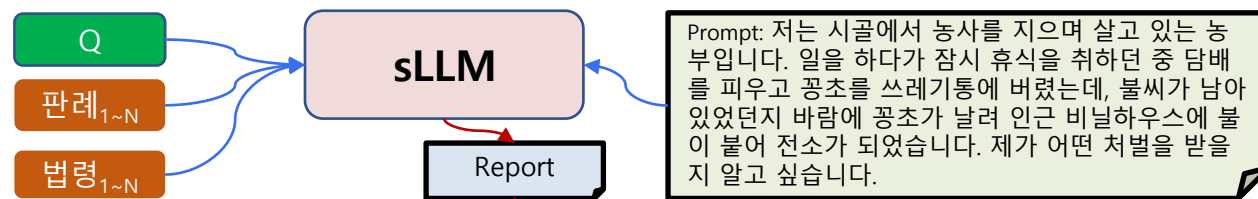
**Query:**  
도요토미 히데쓰구와 이마가와 요시모토 두 사람 다 장남으로 태어났는가?

**예측 판단:** 아니오, 도요토미 히데쓰구는 장남으로 태어났지만, 이마가와 요시모토는 장남으로 태어나지 않았습니다.

**이유 설명:** 도요토미 히데쓰구는 하시바 히데요시의 누나·도모(とも)와 미요시 요시하루(三好吉房)의 장남으로 태어났지만, 이마가와 요시모토는 이마가와씨 제 7대 당주에 있던 이마가와 우지치카(今川氏親)의 다섯째 아들로 태어났습니다.

### 성능현황(한국어 멀티홉 1,000 set)

	F1 Score	ROUGE-1	BLUE
예측 판단	81.72%	26.60%	46.83%
이유 설명	82.17%	64.67%	43.64%
근거인식	81.26% (Precision@Top10)		



질문자의 상황과 두 유사 판례 간의 유사점과 차이점을 분석해 보겠습니다.

#### 유사점:

과실에 의한 화재 발생: 세 경우 모두 피고인(농부 포함)이 담배를 피우는 과정에서 발생한 과실로 인해 화재가 발생했습니다. 불씨 관리의 소홀: 불씨를 제대로 끄지 않아 화재로 이어진 점에서 피고인들은 유사한 부주의를 보였습니다. 주변 환경 요인: 바람과 같은 외부 요인으로 인해 불씨가 다른 곳으로 이동하여 화재가 확산된 점이 공통됩니다. 재산상 손해 발생: 모든 경우에 타인의 재산에 상당한 손해가 발생했습니다.

#### 차이점:

발생 장소의 차이: 질문자의 경우는 자신의 농장 근처에서 사고가 발생했으며, 유사 판례 1은 상가 뒤편, 유사 판례 2는 피고인의 집 마당에서 발생했습니다. 화재가 확산된 범위와 종류: 질문자의 경우는 비닐하우스가 전소된 것에 반해, 유사 판례 1에서는 상가 벽면과 집기가, 유사 판례 2에서는 창고, 비닐하우스, 주택으로 화재가 확산되었습니다. 화재에 옮겨 붙은 대상: 질문자의 경우는 비닐하우스에 불이 옮겨 붙었고, 유사 판례 1에서는 파렛트, 유사 판례 2에서는 소각로의 생활쓰레기로 화재가 시작되었습니다.

(법적 측면에서 세 사례 모두 과실로 인한 재산 피해를 일으켰으나, 구체적인 사건의 세부 사항(화재 발생 장소, 피해 범위 및 소홀히 한 주의 의무의 정도 등)에 따라 법적 책임의 정도가 달라질 수 있습니다. 질문자의 경우에는 과실로 인한 재산손괴최가 적용될 수 있으며, 실제로 어떤 형사 책임을 지게 될지는 사건의 모든 상황을 종합적으로 고려하여 법원이 결정하게 됩니다.)

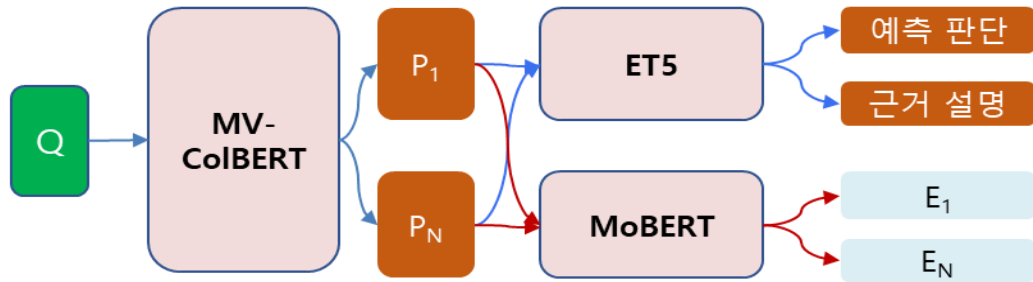
### 성능현황(법률 자문1,500 set, 판례 해설 500 set)

	ROUGE-1	ROUGE-2	ROUGE-L
법률 자문	70.50%	47.44%	60.31%
판례 해설	69.94%	48.12%	62.12%

# 예측 판단/이유 설명을 위한 근거 인식 및 보고서 생성 기술

## 예측 판단과 판단의 이유 설명 모델 개발(ET5/sLLM)

- (1) ET-5모델(330M)을 이용, 예측 판단 및 판단 이유 설명, 근거인식(MoBERT)
- (2) 한국어 sLLM을 이용한 **법률 자문 및 판례 해설 보고서 생성**



### Query:

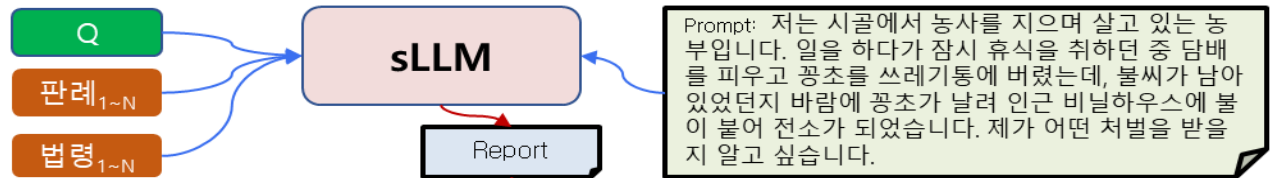
도요토미 히데쓰구와 이마가와 요시모토 두 사람 다 장남으로 태어났는가?

**예측 판단:** 아니오, 도요토미 히데쓰구는 장남으로 태어났지만, 이마가와 요시모토는 장남으로 태어나지 않았습니다.

**이유 설명:** 도요토미 히데쓰구는 하시바 히데요시의 누나·도모(とも)와 미요시 요시하루(三好吉房)의 장남으로 태어났지만, 이마가와 요시모토는 이마가와씨 제7대 당주에 있던 이마가와 우지치카(今川氏親)의 다섯째 아들로 태어났습니다.

### 성능현황(한국어 멀티홉 1,000 set)

	F1 Score	ROUGE-1	BLUE
예측 판단	81.72%	26.60%	46.83%
이유 설명	82.17%	64.67%	43.64%
근거인식	81.26% (Precision@Top10)		



질문자의 상황과 두 유사 판례 간의 유사점과 차이점을 분석해 보겠습니다.

### 유사점:

과실에 의한 화재 발생: 세 경우 모두 피고인(농부 포함)이 담배를 피우는 과정에서 발생한 과실로 인해 화재가 발생했습니다. 불씨 관리의 소홀: 불씨를 제대로 끄지 않아 화재로 이어진 점에서 피고인들은 유사한 부주의를 보였습니다. 주변 환경 요인: 바람과 같은 외부 요인으로 인해 불씨가 다른 곳으로 이동하여 화재가 확산된 점이 공통됩니다. 재산상 손해 발생: 모든 경우에 타인의 재산에 상당한 손해가 발생했습니다.

### 차이점:

발생 장소의 차이: 질문자의 경우는 자신의 농장 근처에서 사고가 발생했으며, 유사 판례 1은 상가 뒤편, 유사 판례 2는 피고인의 집 마당에서 발생했습니다. 화재가 확산된 범위와 종류: 질문자의 경우는 비닐하우스가 전소된 것에 반해, 유사 판례 1에서는 상가 벽면과 집기가, 유사 판례 2에서는 창고, 비닐하우스, 주택으로 화재가 확산되었습니다. 화재에 옮겨 붙은 대상: 질문자의 경우는 비닐하우스에 불이 옮겨 붙었고, 유사 판례 1에서는 파렛트, 유사 판례 2에서는 소각로의 생활쓰레기로 화재가 시작되었습니다.

(법적 측면에서 세 사례 모두 과실로 인한 재산 피해를 일으켰으나, 구체적인 사건의 세부 사항(화재 발생 장소, 피해 범위 및 소홀한 주의 의무의 정도 등)에 따라 법적 책임의 정도가 달라질 수 있습니다. 질문자의 경우에는 과실로 인한 재산손괴죄가 적용될 수 있으며, 실제로 어떤 형사 책임을 지게 될지는 사건의 모든 상황을 종합적으로 고려하여 법원이 결정하게 됩니다.)

### 성능현황(법률 자문1,500 set, 판례 해설 500 set)

	ROUGE-1	ROUGE-2	ROUGE-L
법률 자문	70.50%	47.44%	60.31%
판례 해설	69.94%	48.12%	62.12%



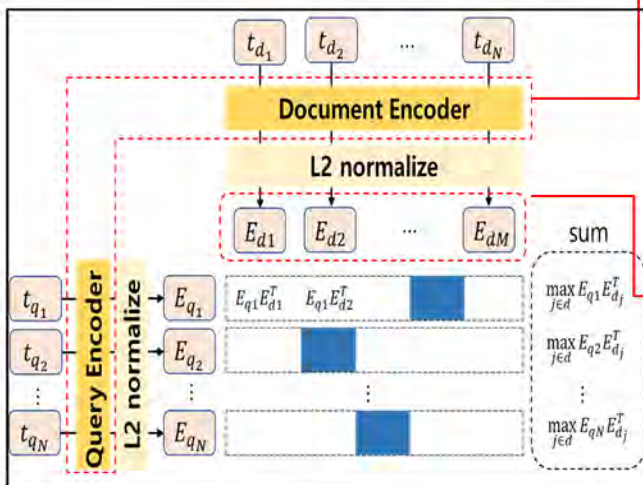
# 뉴럴 단락 검색 기술

All-to-All 토큰으로 늦은 상호작용 연산으로 검색하는 **ColBERT의 성능 고도화**

1. BoW기반 **검색 지향형 언어모델** 사후 연장학습 ← 검색 성능 고도화
2. 멀티 뷰 토큰 기반 **단락의 문맥 정보 저장** ← 색인 볼륨 축소, 실용성

## ColBERT

$$f(q, d) = \sum_{i \in q} \max_{j \in d} E_{q_i} \cdot E_{d_j}^T$$



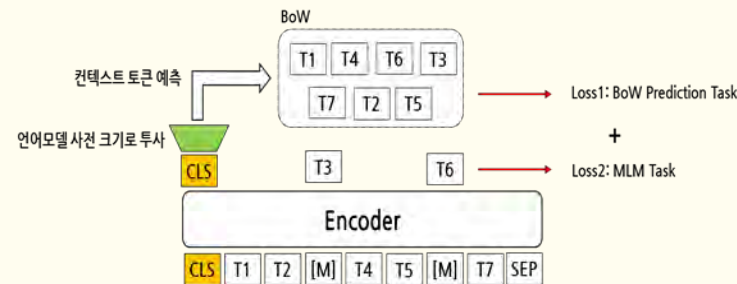
ColBERT 검색 유사도 계산

언어이해모델  
(MoBERT)

기반모델

## 개선대상1: 검색 지향형 이해모델(3차년도)

토큰의 문맥 표현 강화를 위한 **BoW Prediction 태스크 적용** 연장학습 기술 개발  
(법률도메인: **7.3%** BR@Top1 성능 향상)



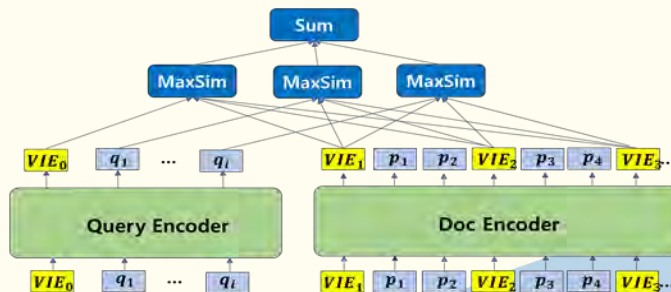
평가데이터

	컬렉션	질문
법률	79,359	1,000

Models	BERT-base		ROM(BERT-base)	
BR@TopN	Top1	Top5	Top1	Top5
법률	0.753	0.947	0.826 (+0.073)	0.969 (+0.022)

## 개선대상2: MV-ColBERT(2차년도)

효율적인 임베딩 색인을 위한 Multi-View 토큰을 이용한 Late Interaction 기반 뉴럴 검색 기술 개발  
(Top40 성능 복구율: **99%**, 색인 축소율: **77%**)



Models	BR@Top1	BR@Top10	BR@Top40	Embedding Size(GB)
BM25	0.599	0.848	0.909	0.6
DPR	0.679	0.920	0.976	0.3
ColBERT	<b>0.798</b>	<b>0.974</b>	<b>0.992</b>	<b>8.3</b>
MV-ColBERT(k=2)	0.782	0.966	0.991	4.7
MV-ColBERT(k=3)	0.783	0.964	0.987	3.2
MV-ColBERT(k=4)	0.770	0.962	0.986	2.4
MV-ColBERT(k=5)	<b>0.770</b>	<b>0.957</b>	<b>0.983</b>	<b>1.9</b>



# 고품질 한국어 slm 모델

한국어 토큰 추가 및 사전 학습을 이용한  
kor-llama2-13.2b 모델 개발  
(vocab size: 50,317 [llama2: 32,000])

## Motivation

- 영어 기반의 2 Trillion 토큰의 초대규모 사전 학습
- 한국어 기반의 fine-tuning (SFT, DPO) 기술 발전
- LLaMA2 모델의 파생 모델 다수, 상용 라이선스

## Concept

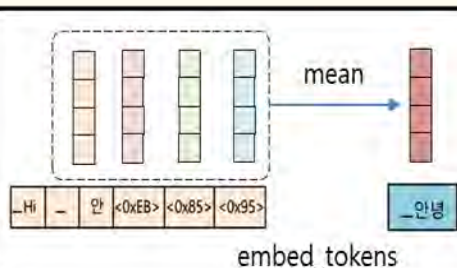
Llama2 tokenizer (32,000 vocab)   kor-sentencepiece tokenizer (20,000 vocab)

Hi   \_   안   <0xEB>   <0x85>   <0x95>

Hi   \_   안녕

kor-Llama2 tokenizer  
(50,317 vocab)

Hi   \_   안녕



## Procedure



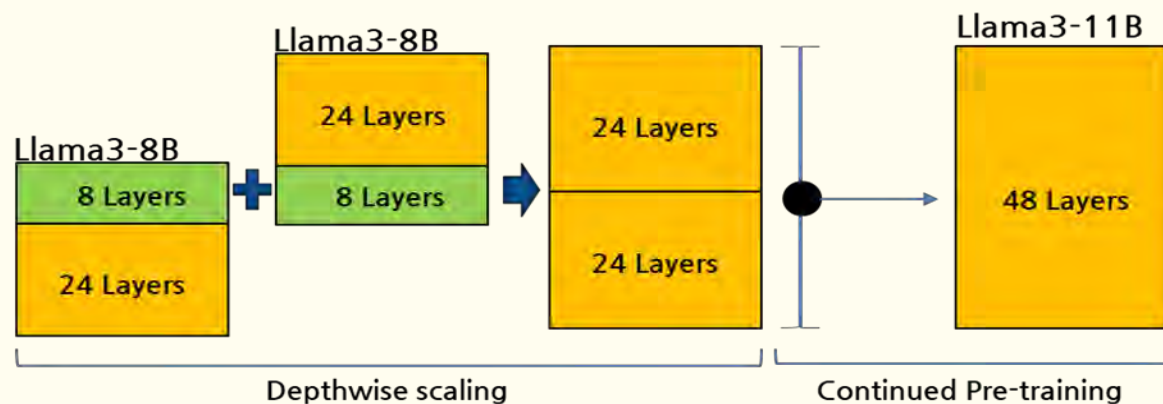
1. Sentence Piece 토큰 학습 및 모델(75G 한국어 코퍼스)
2. Fully Pretraining with Kor-sentencepiece tokenizer
  - a. 87.5G 코퍼스(한국어 52.5G, 영어 35G)를 이용하여 llama2 모델 continual learning (FSDP)
3. Fine-tuning 및 Alignment 학습(FSDP)
  - a. 1,822k개의 instruction-following/CoT 학습데이터
  - b. 222k개의 Preference data로 Direct Preference Optimization 학습 수행

## Performance

- 한국어 NLP Task 평가 결과(평가 measure: F1-Score)
- 학습데이터: 63,247, 평가데이터 각 500건

모델	Average(F1)	ko_hellaswag	ko_arc	ko_mmlu	ko_commeagen
GPT4-turbo	0.625	0.638	0.926	0.560	0.376
ChatGPT-turbo	0.418	0.344	0.658	0.294	0.374
llama2-13b(base)	0.415	0.442	0.504	0.238	0.474
llama2-13b-sft-dpo	0.446	0.442	0.628	0.231	0.484
kor-llama2-13.2b(8e-7, data1)	0.474	0.480	0.612	0.262	0.543
kor-llama2-13.2b(8e-5, data1)	0.491	0.476	0.646	0.296	0.545
kor-llama2-13.2b(8e-6, data1)	0.495	0.481	0.651	0.302	0.546
kor-llama2-13.2b(8e-6, data2)	0.615	0.542	0.804	0.492	0.608

사전학습 모델의 **깊이확장** 및 **연장학습**을 이용한 성능 고도화  
Llama3 모델 기반 한국어 및 다국어 연장 학습(0.1T tokens 목표 학습진행 중)



모델	AVG	EL	FA	NLI	QA	RC	klue (ner)	klue (re)	kmmlu	kobest (copa)	kobest (hs)	kobest (sn)	kobest (wic)	korea (cg)	kornli	Korsts (pears on)	Korsts (spear man)
mistral-7b	0.6275	0.6087	0.3173	0.8213	0.5020	0.8882	0.5273	0.6900	0.3600	0.8580	0.8340	0.8425	0.6440	0.3173	0.7720	0.9077	0.9145
mistral-7b-instruct-v0.2	0.6255	0.5440	0.4149	0.7413	0.5012	0.9261	0.5540	0.5340	0.4044	0.7800	0.7640	0.9825	0.5980	0.4149	0.6800	0.9023	0.8936
llama3-8b	0.3284	0.0337	0.0945	0.5373	0.4008	0.5755	0.0514	0.0160	0.3556	0.8640	0.7480	0.0000	0.4460	0.0945	0.0000	0.8654	0.8612
llama3-8b-instruction	0.6430	0.6090	0.3859	0.7807	0.5427	0.8969	0.4880	0.7300	0.4133	0.8720	0.8220	0.9800	0.6720	0.3859	0.6480	0.8592	0.8515
solar-10.7b-v1.0	0.7313	0.7000	0.4589	<b>0.9087</b>	<b>0.6379</b>	0.9509	0.5980	0.8020	<b>0.4978</b>	<b>0.9440</b>	0.9260	0.9925	0.7780	0.4589	<b>0.8560</b>	0.9300	0.9301
solar-10.7b-instruct-v1.0	0.7224	0.6830	0.4488	0.9040	0.6225	<b>0.9538</b>	0.5820	0.7840	0.4489	0.9320	<b>0.9360</b>	0.9925	0.7960	0.4488	0.8440	<b>0.9376</b>	<b>0.9312</b>
XAI-NLP/Llama3-11b	<b>0.7457</b>	<b>0.7690</b>	<b>0.4958</b>	0.8813	0.6287	0.9537	<b>0.7280</b>	<b>0.8100</b>	0.4533	0.9240	0.9160	<b>0.9950</b>	<b>0.8040</b>	<b>0.4958</b>	0.8040	0.9366	0.9294



# 생성형 언어모델의 한국어 사실성 검증 기술 개발: KorFaCTSCORE

## MoBERT 기반 검색기를 활용한 Atomic Fact 검증기 개발

입력: GPT-4 생성 문장들(=Atomic Facts), 출력: True/False

사람 검증 대비 에러율 약 3% (검색기 제외 시 에러율 12% 상승)

한글 문서에 대한 사실을 atomic level로 세밀하게 식별하여 정보의 신뢰성을 높임

<질의> 장기하의 약력에 대해 말해줘.



생성형 LM(언어모델)

### <생성 결과>

장기하는 예술의 전당 한가람디자인미술관에서 개최된 '프로젝트 밴드 헬로루키' 대회에서 최종 우승하며 처음으로 대중 앞에 등장했습니다. '장기하와 얼굴들'이라는 밴드를 결성해, 2008년 '싸구려 커피'를 발표하며 큰 주목을 받았습니다. 이 곡은 특유의 가사와 멜로디로 대중적 인기를 얻었고, 장기하와 얼굴들은 '2011 대한민국 대중문화예술상'에서 대통령상을 수상하는 등 많은 성공을 거뒀습니다.

[RQ] "LLM(언어모델) 생성 결과의 개별 정보 조각들이 얼마나 정확할까?"

→ KorFactScore: 한국어 사실성 검증 모델

### ① 정보 조각 : Atomic Facts

- 장기하는 '프로젝트 밴드 헬로루키' 대회에서 최종 우승했습니다. ✗
- '프로젝트 밴드 헬로루키' 대회는 예술의 전당 한가람디자인미술관에서 개최되었습니다. ✗
- 장기하는 대회 우승을 통해 처음으로 대중 앞에 등장했습니다. ✗
- '장기하와 얼굴들'이라는 밴드를 결성하였습니다. ✓
- '장기하와 얼굴들'은 2008년 '싸구려 커피'를 발표하였습니다. ✓
- 싸구려 커피 발표로 큰 주목을 받았습니다. ✓
- 이 곡은 특유의 가사와 멜로디로 대중적 인기를 얻었다. ✓
- 장기하와 얼굴들은 '2011 대한민국 대중문화예술상'에서 대통령상을 수상했다. ✗
- 장기하와 얼굴들은 많은 성공을 거뒀다. ✓

### ② 사실성 검증 : Fact Verification

외부지식 기반 검증(RAG)



③

KorFactScore  
89.5%

# 생성형 언어모델의 한국어 사실성 검증 기술 개발: KorFaCTSCORE

## System의 사실판단 능력

System 결과와 정답(ground truth)의 일치된 정도를 보여줌

Model	Size	한국인 BM25	Cross-Encoder	외국인 BM25	Cross-Encoder
GPT-3.5	-	82.4	85.4	75.1	77.2
GPT-4	-	92.4	94.2	91.6	92.8
Gemini-1.0	-	81.4	84.5	68.2	79.3
Gemini-1.5	-	81.9	57.6	70.3	74.2
EXAONE-7.8B	7.8B	89.9	90.8	86.6	86.0
LLaMa-2-chat	7B	61.1	64.5	60.1	62.8
	13B	70.6	71.5	80.6	80.7
	70B	82.3	82.9	76.3	79.7
LLaMa-3-Instruct	8B	84.7	86.2	83.9	86.5
	70B	89.1	90.7	86.5	86.9
LLaMa-3.1-Instruct	8B	78.6	82.0	69.6	70.9
	70B	89.7	90.3	89.9	90.2

## 생성문서의 사실성 점수 (FactScore)

Model	Size	한국인 BM25	Cross-Encoder	외국인 BM25	Cross-Encoder
GPT-3.5	-	45.3	47.4	62.9	65.4
GPT-4	-	53.8	53.1	79.6	81.1
Gemini-1.0	-	33.4	36.2	55.2	64.1
Gemini-1.5	-	52.0	70.0	81.0	85.3
EXAONE-7.8B	7.8B	50.7	51.2	76.3	76.1
LLaMa-2-chat	7B	39.6	37.3	57.6	59.6
	13B	64.8	67.3	79.0	80.3
	70B	48.9	49.8	67.1	69.9
LLaMa-3-Instruct	8B	48.9	50.9	73.8	77.4
	70B	47.9	47.3	74.5	74.3
LLaMa-3.1-Instruct	8B	27.9	31.1	52.8	54.6
	70B	54.7	54.5	82.0	82.7

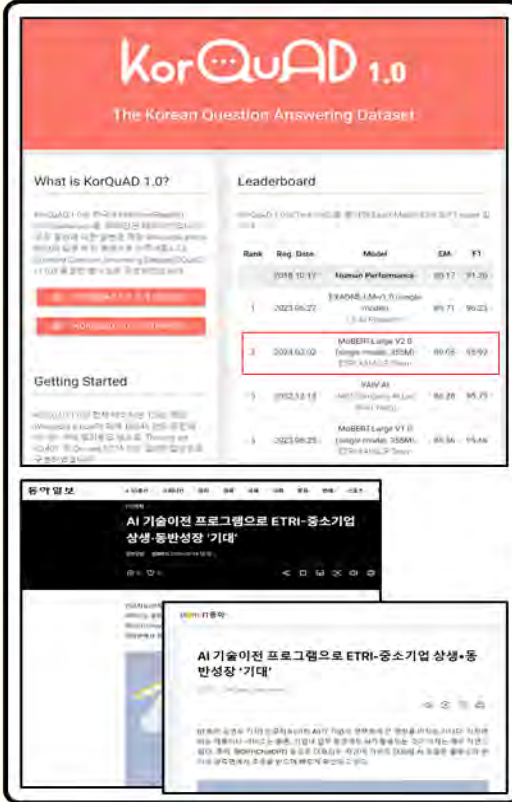
\*Cross-Encoder 검색기는 ETRI 자체 보유 기술  
\*Atomic Facts 분할 후 평가 대상 고정하여 진행



# 한국어 언어이해 모델(MoBERT) 및 응용 기술

- 소형 언어이해 모델로 초거대 LLM 수준의 한국어 이해 기술 개발
  - ✓ 한국어 최고 언어이해 모델 개발 (저작권 문제가 없는 고품질 데이터 사용)
    - 기계독해 챌린지 KorQuad 1.0 2위 (언어이해 모델로는 1위)
  - ✓ 초거대 LLM 수준의 한국어 이해 기술 개발 (텍스트 분류, 자연어 추론, 뉴럴 검색 등)
    - 한국어 이해 관련 다수 태스크(NSMC, KLUE)에서 최고 성능
  - ✓ 기술이전을 통한 산업계 보급 2건

기술이전명	이전 기업	이전 시기	기술료 (백만원)
한국어 특화 이해모델 기반 뉴럴 단락 검색 기술 v1.0	(주)인포플라	2024. 02.21	50
한국어에 최적화된 언어이해 모델 (MoBERT) 및 응용 기술 *일부: 1세부기술(MoBERT)	(주)딥모달	2024. 04.03	20



- 기계독해 태스크 성능 평가

KorQuad 1.0: 질문의 정답(정답의 시작/끝 범위)을 단락에서 추론

KLUE-MRC: 질문의 정답(정답의 시작/끝 범위, 또는 정답 없음)을 단락에서 추론

Model	KLUE-MRC	
	EM	ROUGE
KLUE-RoBERTa-Large	75.58	80.59
MoBERT-Large	82.66	85.89

- 한국어 이해 태스크 성능 평가

KLUE-YNAT: 문장의 주제를 7개 범주로 분류

KLUE-STs: 두 문장 간의 의미적 유사성을 실수(0.0~5.0 사이)로 추론

KLUE-NLI: 두 문장 간의 의미적 유사성을 함의, 중립, 모순 관계로 분류

KLUE-RE: 문장의 주어, 목적어 객체 간의 관계(40개) 분류

Model	YNAT	KLUE-STs		KLUE-NLI	KLUE-RE	
	F1	R <sup>P</sup>	F1	ACC	F1 <sub>mic</sub>	AUC
KLUE-RoBERTa-Large	85.69	93.35	86.63	89.17	71.13	72.98
MoBERT-Large	88.09	93.76	87.53	93.77	75.76	80.51

- 교착어인 한국어의 특성을 반영한 학습과 추론
  - 의미 최소 단위로 언어의 구조와 의미를 보다 정확히 학습 가능 (형태소의 품사 태그 및 원형 복원 기능 제외)
  - 원문을 표현하는 토큰 개수 감소로 더 많은 문맥의 처리 가능
- 한국어에 최적화된 트랜스포머 인코더 구조 연구
  - 긴 학습(1백만 스텝), 마스킹 비율 상향(15%→20%) 등을 통한 성능개선
  - disentangled attention, 절대·상대 위치정보의 결합 등은 성능개선 효과 미미
- 한국어 이해 태스크는 이해모델이 성능·비용 측면에서 생성모델보다 효과적임

MoBERT 기반 응용 기술(fine-tuning)

기계 독해

텍스트 분류

패러프레이즈 인식

단락 재순위화

정답 근거 인식

한국어에 최적화된 언어모델 (MoBERT 언어모델)

사전학습 (pretraining)

Knowledge Source

# 오픈 소스 활용 현황



Open LLM 기반  
연장 학습 및 파인튜닝



# 오픈 소스 활용 현황

OpenSource AI Full Stack v1.0

robin.hwang

AI Agent

Agent

AutoGPT

Javis

MetaGPT

AgentVerse

Apply

ChatDEV

Leon

PR Agent

GPT Pilot

Web UI

Chat

Chatbot UI

LibreChat

Quivr

h2o

Anything

RolePlay

Next.js AI Chatbot

Langchain-Chatchat

Other

ComfyUI

SoM

SD WebUI

Productivity

Automation

Tldraw

ShellGPT

JupyterAI

PandasAI

Open Interpreter

Development

keploy

Sweep

Continue

Screenshot to Code

Prompt

Improve

PromptFlow

MemGPT

NeMo Guardrails

Convert

{guidance}

Lingua

TypeChat

LLM Framework

Framework

LangChain

Spring AI

Haystack

Data Chain

LlamaIndex

Data-Juicer

Pipeline

Flowise

Langflow

Evaluation

Evaluation

OpenAI Evals

DeepEval

UpTrain

promptfoo

Performance

OpenCompass

ragas

프레임워크

AI Hub

AI 데이터찾기

AI 허브소개

참여하기

커뮤니티

AI 개발지원

고객지원

로그인

회원가입

데이터 찾기

분야별 보기

☐ 한국어
 ☐ 영상이미지
 ☐ 헬스케어
 ☐ 교통물류
 ☐ 재난안전환경
 ☐ 농축수산
 ☐ 문화관광
 ☐ 스포츠

데이터 유형 보기

☐ 이미지
 ☐ 비디오
 ☒ 텍스트
 ☐ 오디오
 ☐ 3D
 ☐ 센서

구축년도별 보기

☐ 2017년
 ☐ 2018년
 ☐ 2019년
 ☐ 2020년
 ☐ 2021년
 ☐ 2022년
 ☐ 2023년

조회

선택 초기화

검색어를 입력해주세요

Q

데이터셋 (5건)

※ 데이터 다운로드 PC에서만 가능합니다.

전체

최신순

20

조회

BETA

다운로드

법률

법률/규정 텍스트 분석 데이터 (고도화)- 상황에 ...

287.66 MB

BETA

다운로드

법률

법률 지식기반 관계 데이터

243.96 MB

BETA

다운로드

법률

국내 법률 대상 다국어 번역 데이터

559.10 MB

NEW

다운로드

법률

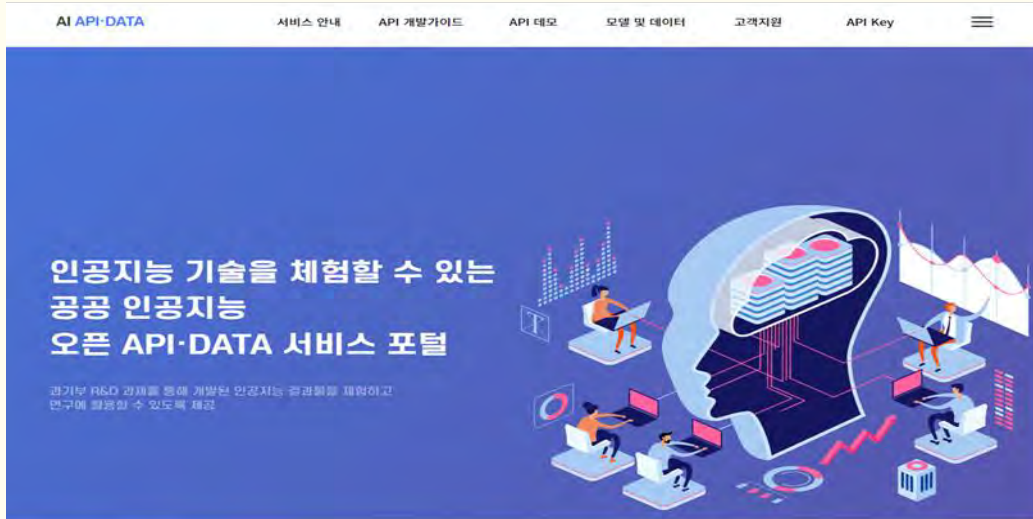
금융, 법률 문서 기계독해 데이터

125.87 MB

출처: <https://revf.tistory.com/303>

<https://www.aihub.or.kr/>

# 오픈 소스 활용 현황(기술 및 데이터 공개)



2,349개 기관에서 활용

엑소브레인 OpenAPI  
누적 6,145만건 사용  
(23년 2월 기준)

행정문서QA, 패러프레이즈  
인식 API 공개 ('21.09)

개체연결, 상호참조해결  
공개 ('22.10)

→ 총 18종 기술

질의응답 학습말뭉치  
3만건 공개('21.10)  
4만건 공개('22.10)

→ 총 14종 데이터

언어 모델 2종 공개  
(KorBERT, ET5)





# 오픈 소스 활용 현황(기술 및 데이터 공개)

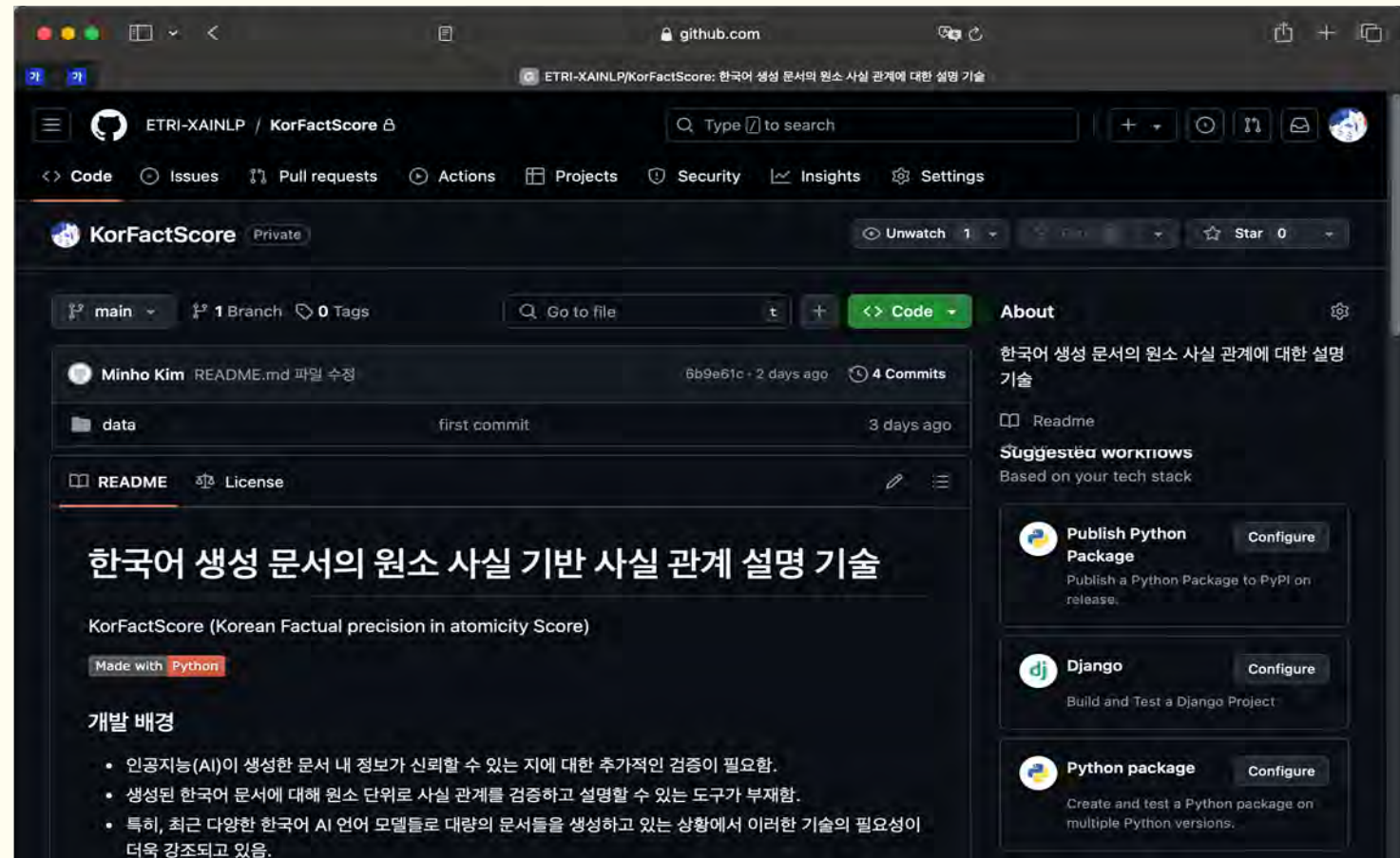
## 주요 특징

- 특정 언어 모델에 제한되지 않으며(Model Agnostic), 평가용 언어모델을 교체하여 모델의 성능을 평가할 수 있음
- 생성문서의 사실 관계 설명을 위한 평가용 데이터 제공
- 문서의 사실성 평가와 언어모델의 사실 판단 능력 평가를 위해 정답셋(Ground Truth Set) 제공

## • 활용 사례

- 생성된 문서의 사실성 평가에 활용
- 자체 개발된 언어모델의 환각 생성 가능성 평가에 활용
- 자체 개발된 언어모델의 사실판단 능력 평가에 활용
- 개선된 기술 개발을 위한 기초 프레임워크로 활용 가능

한국어 생성 문서의 원소 사실 기반  
사실 관계 설명 기술





# 감사합니다.

오픈 소스 및 데이터 활용한  
설명가능한 의사결정 지원 인공지능 기술 연구

