

GIST, 정확도 23% 높은 AI 법률 서비스 개발

모르는 것은 AI 스스로 묻고 찾아

답변 신뢰성 높이는 RAG (검색 증강 생성) 기술

- 전기전자컴퓨터공학과 이흥노 교수팀, 법률 분야에 최적화된 RAG 프레임워크 기술 개발... 기존 RAG 대비 검색·응답 정확도 23% 향상, 파인 튜닝된 LLM 대비 성능 14% 높아
- 법률 실무 활용은 물론 법률상담 서비스를 받기 어려운 취약계층에도 도움 기대
- 국제학술지 《IEEE Access》 게재



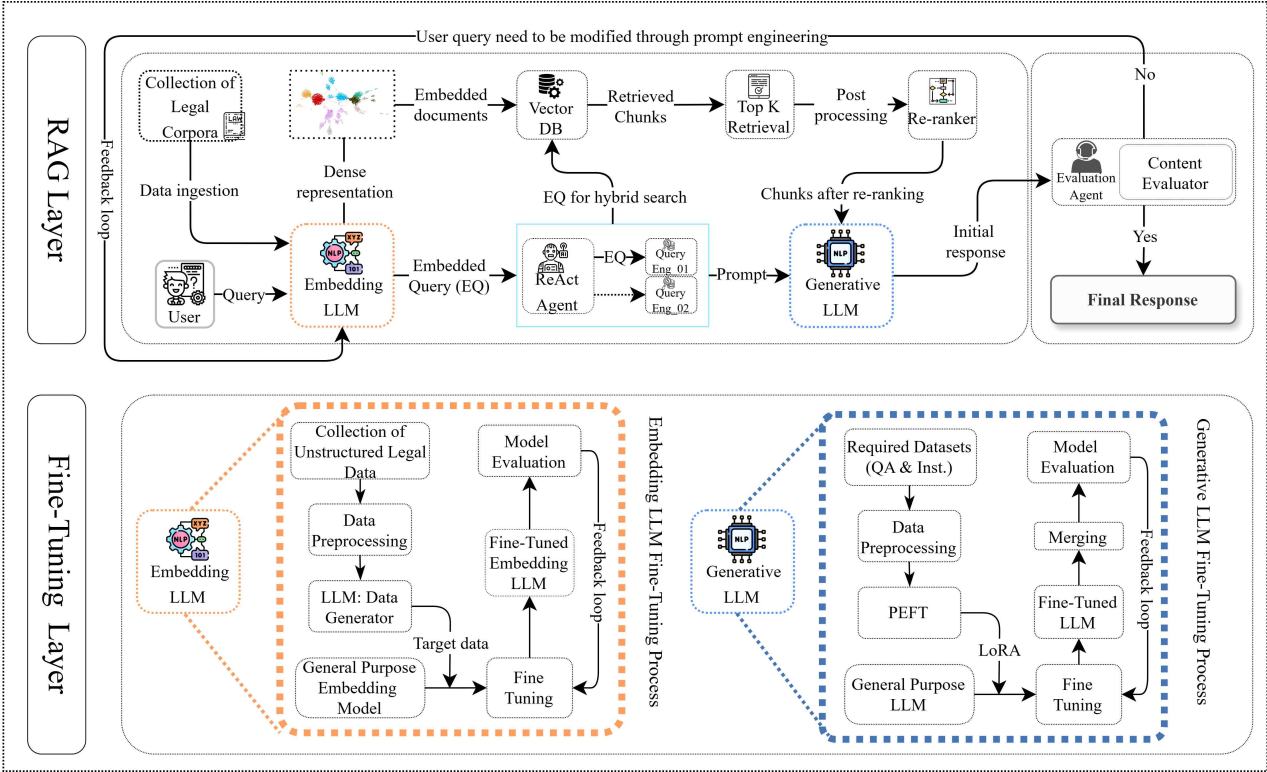
▲ (왼쪽부터) 전기전자컴퓨터공학과 이흥노 교수, RAHMAN S M WAHIDUR 학생

최근 인공지능(AI) 기술의 급격한 발전은 사회·경제 전반에 혁신적인 변화를 일으키며, 법률 분야에서도 빠르게 확산되고 있다.

그러나, 학습된 데이터를 기반으로 답변을 생성하는 대규모 언어 모델(LLM)은 법률 분야의 복잡한 질의를 정확히 처리하는 데 한계가 있다. 특히 해석이 중요한 판례 분석이나 계약서 작성에 있어 AI가 잘못된 정보를 제공할 경우, 심각한 문제가 발생할 수 있다.

광주과학기술원(GIST, 총장 임기철)은 전기전자컴퓨터공학과 이흥노 교수(ITRC 블록체인 지능융합센터장) 연구팀이 법률 분야에 특화된 '검색 증강 생성 (Retrieval-Augmented Generation, RAG)' 프레임워크 기술을 개발했다고 밝혔다. 모르는 것은 AI 스스로 묻고 찾아봄으로써 답변의 신뢰성과 정확도를 높인다는 점이 가장 큰 특징이다.

이 기술은 기존 AI 기반 법률 추론에서 발생하는 ‘할루시네이션(hallucination: 잘못된 정보, 환각)’ 문제를 획기적으로 줄이며, 정확도와 투명성, 신뢰성을 높여 **취약 계층을 위한 법률 지원뿐 아니라 법률 실무 전반에 폭넓게 활용될 수 있을 것**으로 기대된다.



[그림] 제안된 LQ-RAG 시스템의 개략적인 설계. 본 시스템은 파인 튜닝 레이어(Fine-Tuning Layer)와 RAG 레이어(RAG Layer)의 두 가지 주요 구성 요소로 구분된다.

- ▶ **파인 튜닝 레이어(Fine-Tuning Layer):** 임베딩 LLM과 생성 LLM의 성능을 최적화하기 위한 파인 튜닝을 수행하는 계층이다.
- ▶ **RAG 레이어(RAG Layer):** 고급 RAG 모듈, 평가 에이전트(evaluation agent), 프롬프트 엔지니어링 에이전트(Prompt Engineering Agent), 그리고 피드백 메커니즘을 포함하여 시스템의 응답 생성 과정에서 정확성과 신뢰성을 보장하는 역할을 수행한다.

본 설계를 통해 LQ-RAG 시스템은 법률 도메인에서 보다 정밀하고 신뢰할 수 있는 정보 검색 및 응답 생성을 가능하게 하며, 궁극적으로 법률 AI 시스템의 성능을 향상시키는 데 기여한다.

기존의 LLM 기반 법률 AI 시스템은 58~82%의 할루시네이션 발생률을 보이는 것으로 보고되었다. 이에 따라 AI가 정확히 검색한 법률 정보를 반영하는 RAG* 기술이 주목받고 있지만, 기존 RAG 방식도 정보 검색의 한계와 법률 문맥에 대한 적용력 부족이라는 문제를 안고 있다.

* **RAG:** 사전에 설정된 데이터베이스나 문서에서 필요한 정보를 검색하여, 기존 LLM이 반영하지 못하는 최신 데이터를 기반으로 답변을 생성할 수 있다. RAG는 대규모 모델을 다시 학습시키는 대신 외부 데이터를 검색해 사용하기 때문에 비용과 시간을 절약할 수 있다는 장점도 있다.

이를 해결하기 위해 연구팀은 **법률 데이터를 효율적으로 검색하고 활용하는 동시에 답변의 신뢰성과 정확도를 높인 법률 분야에 최적화된 ‘Legal Query RAG(LQ-RAG)’ 프레임워크를 개발**하였다.

LQ-RAG 모델은 광범위한 법률 텍스트를 활용해 임베딩* 생성 LLM과 응답 생성 LLM을 각각 파인 튜닝* 했다. 방대한 판례와 법령 자료를 학습함으로써 전문 법률 용어와 문서 구조를 심층적으로 이해할 수 있도록 했으며, 실제 법률 질의응답 데이터를 기반으로 생성 모델을 재학습해 더욱 정교한 답변 능력을 확보했다.

* **임베딩(Embedding)** : 문서를 벡터로 변환해 의미를 수치화한 표현

* **파인 튜닝(fine tuning)**: 기존에 학습된 모델을 특정 목적이나 데이터세트에 맞게 추가 학습하는 과정을 의미한다. 이미 학습된 모델을 가져와 특정 작업에 적합하도록 최적화하는 것으로 기본 모델을 처음부터 학습하는 것보다 빠르고 효율적이다.

LQ-RAG는 ▲**맞춤형 평가 에이전트(Evaluation Agent)** ▲**응답 생성 LLM(Response Generator LLM)** ▲**프롬프트 엔지니어링 에이전트(Prompt Engineering Agent)** ▲**임베딩 생성 LLM(Embedding Generator LLM)** 등 네 가지 핵심 요소를 통합했다.

이 구조는 할루시네이션을 효과적으로 최대한 줄이고, 도메인별 정확도를 개선하며, 복잡한 질문에도 명확하고 수준 높은 답변을 제공한다. 또한 재귀적 피드백 과정을 통해 성능을 지속적으로 개선할 수 있도록 했다.

* **재귀적 피드백 과정**: 생성된 응답의 평가 기반의 검색 및 생성 단계를 반복적으로 개선하여 더 정확하고 관련성 높은 답변을 유도한 메커니즘

LQ-RAG는 추론 과정에서 에이전트 기반의 반복적 개선 메커니즘을 명시적으로 적용해 최적의 답변을 도출한다. 생성된 답변은 평가 에이전트를 통해 맥락의 적절성과 사실적 정확성을 기준으로 평가된다.

LQ-RAG는 AI가 생성한 답변을 지속적으로 개선하는 재귀적 피드백 메커니즘을 통해 신뢰성을 높였으며, 연구팀은 법률 문서를 체계적으로 임베딩하여 고차원 벡터로 변환하고, 이를 바탕으로 **AI가 보다 정확한 법률 정보를 제공할 수 있도록** 했다.

중국의 생성형 AI '딥시크(DeepSeek)-R1'과 연구팀의 LQ-RAG를 비교하면 두 모델 모두 더 정확하고 정교한 답변을 제공하기 위해 고유한 기법을 사용하지만, **응답을 개선하는 방식에서 차이를** 보인다.

딥시크-R1은 강화 학습(RL, Reinforcement Learning)을 기반으로 사고 능력을 발전시키며, 연쇄적 추론 과정을 통해 답변의 질을 향상시킨다. 스스로 생성한 답변을 검토하고 개선하는 기능도 수행한다.

딥시크-R1이 단일 모델 내에서 재귀적으로 응답을 개선하는 반면, **LQ-RAG는 다중 에이전트 협력 방식을 통해** 답변을 개선한다. 두 모델 모두 사람이 직접 피드백을 제공하지 않고, 자체적인 최적화 과정을 거친다는 공통점이 있다.

LQ-RAG를 적용한 결과, 기존 RAG 시스템 대비 **법률 문서 검색 및 응답 정확도가 23% 향상**되었으며, 파인 튜닝된 LLM과 비교해도 14% 높은 성능을 기록했다.

이는 고급 RAG 모듈과 피드백 메커니즘이 결합된 도메인 특화 LLM이 법률 실무에서 AI의 신뢰성과 성능을 크게 높일 수 있음을 보여 준다.

연구팀은 **계약서 작성과 준법 감시 등 법률 업무의 효율성을 높이기 위한 에이전트 기반 법률 워크플로우(workflow)**를 개발해 법률 전문가들이 핵심 업무에 집중할 수 있도록 지원할 계획이다.

이흥노 교수는 “이 기술을 **법률 문서 분석, 계약서 작성 자동화, 준법 감시 등 다양한 법률 업무에 적용**할 계획”이라며, “검색 증강 생성(RAG)과 다중 에이전트 협업 기술을 결합해 신뢰성 높은 법률 AI 시스템을 구축하고, 보다 정확한 **법률 분석과 신뢰할 수 있는 AI 기반 법률 솔루션을 제공할 것**”이라고 밝혔다.

GIST 전기전자컴퓨터공학과 이흥노 교수가 지도하고 Rahman, S M Wahidur 통합과정생이 수행한 이번 연구는 과학기술정보통신부와 정보통신기획평가원(IITP)의 대학ICT연구센터(ITRC) 지원을 받았다. 연구 결과는 국제학술지 《IEEE ACCESS》에 2025년 2월 14일 온라인 게재되었다.

논문의 주요 정보

1. 논문명, 저자정보

- 저널명 : IEEE ACCESS (IF: 3.4, 2023년 기준)
- 논문명 : Legal Query RAG
- 저자 정보 : RAHMAN S M WAHIDUR(제1저자, EECS), 이흥노(교신저자, EECS)