

보도시점 2023. 12. 21.(목) 09:00 배포 2023. 12. 21.(목) 08:30

[2023년 특허청 정책 돋보기]

특허청, 세계 최초 초거대 특허전용 언어모델 구축

- 특허청-엘지(LG) 인공지능(AI)연구원, 특허 인공지능(AI) 개발 협력 성과 발표회 개최(12.21) -
- 특허전용 언어모델 기초로 내년부터 인공지능(AI) 심사체계 개발에 본격 착수 -

특허청(청장 이인실)은 엘지(LG) 인공지능(AI)연구원(원장 배경훈)과 12. 21.(목) 10시 정부대전청사(대전시 서구)에서 인공지능(AI) 개발에 대한 양 기관의 협력 성과를 공유하는 발표회를 개최한다.

발표회에는 이인실 특허청장, 배경훈 엘지(LG) 인공지능(AI)연구원장 등 관계자 13명이 참석하며, 올해 양 기관이 협력해 세계 최초로 구축한 초거대 특허전용 언어모델을 처음으로 공개할 예정이다.

<특허청, 엘지(LG) 인공지능(AI)연구원과 특허분야에 특화된 초거대 인공지능(AI) 언어모델 개발>

특허청은 지난 7월 14일 엘지(LG) 인공지능(AI)연구원과 특허행정을 혁신하기 위해 인공지능(AI) 기술을 특허행정에 적용하는 업무협약을 체결했고, 그 첫 번째 협력 과제로 초거대 특허전용 언어모델을 구축해 왔다.

이번 언어모델은 엘지(LG) 인공지능(AI)연구원이 보유하고 있던 초거대 인공지능(AI) ‘엑사원(EXAONE)’에 특허청이 보유하고 있는 특허공보 등 특허행정과 관련된 7종의 정보*(1.78TB)를 학습해 구축됐다. 또한 특허청의 환경 및 활용목적 등을 고려해 88억 파라미터* 규모로 설계됐으며, 2차에 걸친 학습 과정을 통해 성능을 향상시켰다.

* 7종의 정보: 국/영문 특허공보, 통지서, 기계번역·시피시(CPC)분류·기계독해 정보, 상담사례집

* 파라미터: 사용자가 원하는 방식으로 자료가 처리되도록 하기 위해 명령어를 입력할 때 추가하거나 변경하는 수치 정보

이에 따라 성능평가 결과, 문장 요약, 번역 등 전체 평가 업무에서 평균 이상의 결과를 보였으며, 분야별 튜닝이 이루어지면 보다 더 향상된 성능을 발휘할 것으로 기대된다.

<특허전용 언어모델 기초로 내년부터 인공지능(AI) 심사체계 개발에 본격 착수>

특허청은 이번에 구축된 특허전용 언어모델을 기초로 내년부터 특허검색, 분류 등 심사 업무를 혁신하기 위한 연구개발에 본격 착수할 계획이다.

이번 성과 발표회에서는 앞으로 양 기관이 인공지능(AI) 심사체계 개발을 위해 협력할 내용도 함께 논의될 예정이다.

이인실 특허청장은 “인공지능(AI) 영역에 강점이 있는 엘지(LG) 인공지능(AI)연구원과의 긴밀한 협력을 통해 특허전용 인공지능(AI)를 개발하게 된 것을 기쁘게 생각한다”면서 “앞으로도 특허행정 분야에 혁신적인 기술을 적극적으로 도입해 특허 체계의 발전을 이끌겠다”고 밝혔다.

엘지(LG) 인공지능(AI)연구원 배경훈 원장은 “엘지(LG) 인공지능(AI)연구원은 특허청의 특허 문헌을 바탕으로 학습한 엑사원(EXAONE)이 특허 분야 업무의 생산성을 혁신해, 우리나라 인공지능(AI) 산업 발전에 토대가 되기를 희망한다”면서 “앞으로도 특허청과 긴밀하게 협력해 전문가 인공지능(AI)의 세계 선진 활용 사례를 만들어 갈 것”이라고 말했다.

※ 붙임 1. 특허청-LG AI연구원 AI 개발협력 성과발표회 참고 이미지

2. MOU 이후 양 기관 협력 내용

3. 엑사원 유니버스(EXAONE Universe) 개요

※ 사진은 행사 후 배포 예정

담당 부서	산업재산정보국 산업재산정보시스템과	책임자	과 장	최일승 (042-481-5099)
		담당자	사무관	김영백 (042-481-8634)



□ 추진 경과

- 위탁연구사업 계약 체결 : 한국특허정보원 - LG AI연구원(7. 7)
- MOU 체결 : 특허청 - LG AI연구원(7. 14)
- 실무협의회를 통해 구체적 협력 내용 협의(8. 10)

□ 양 기관 협력 내용

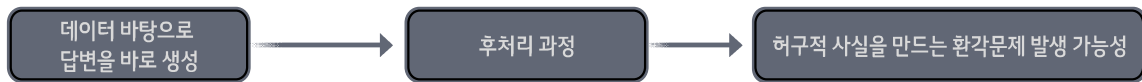
- (청 → LG) 특허 학습 데이터 제공('23.7월)
 - 국문 특허공보 학습 데이터
 - 영문 특허공보 학습 데이터
 - 의견제출통지서 학습 데이터
 - 한-영 기계번역 학습 데이터
 - CPC분류 학습 데이터
 - 특허상담 기계독해(MRC) 학습 데이터
 - 특허고객상담사례집 데이터
- (LG → 청) 초거대 언어모델 관련 기술 및 노하우 이전
 - 거대 언어모델 사전학습 환경 구성('23.7월)
 - 거대 언어모델 사전학습 실시('23.7월)
 - 거대 언어모델의 지식사업 적용 관련 PATINEX 행사 발표('23. 9월)
 - 기술 이전 세미나 실시('23.10월)
 - 거대 언어모델 성능평가('23.11월)
 - 상표 주요이슈(식별력 등) 관련 AI기술 적용방안 자문 제공('23.11월)

□ EXAONE Universe 소개

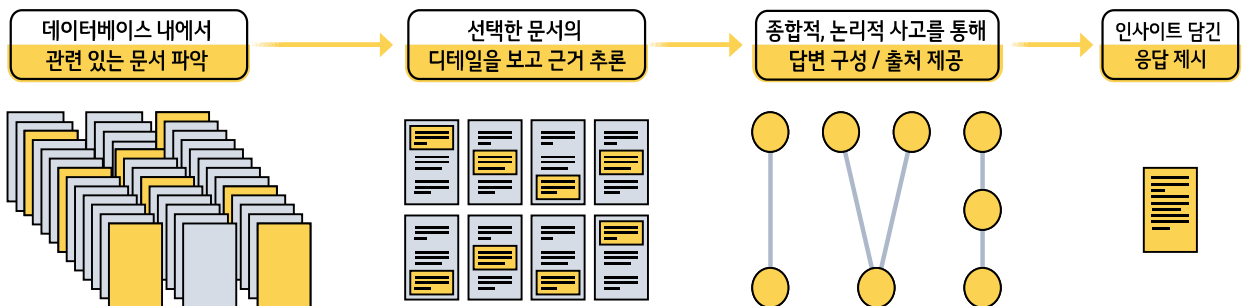
- 전문가에게 인사이트를 제공하기 위해, 정확한 근거를 기반으로 종합적이고 논리적인 사고를 통해 답변을 생성하고, 생성의 근거가 되는 자료를 레퍼런스로 제공

* 기존 AI 모델들의 단점 중 하나인 부정확한 답변 제공 문제 해결

기존 생성형 AI



UNIVERSE



□ EXAONE Universe 추론 과정

- 한국어와 영어를 지원하며, 논리적인 추론 과정을 거쳐 답변을 생성

- ① 문서 검색 단계에서, 질문과 관련된 문서 서치
- ② 검색된 문서에서 질문과 관련된 단락 서치
- ③ 문서의 관련 단락들의 내용을 기반, 요약 및 근거 답변 생성
- ④ 근거들을 재조합하여 최종 답변을 생성

* 엑사원(EXAONE) : EXpert + AI for everyONE