

정책연구용역 결과 평가서

※ []에는 해당되는 곳에 √표를 합니다.

정책연구과제명	손해배상 사건에서의 인공지능(AI) 활용방안 (후속연구)		
과제 담당관	소속(직급)	사법정책연구원(연구위원)	성 명 이영창
연구자	서울대학교 산학협력단 (책임연구원 : 박성철)		
연구기간	2023. 5. 18. ~ 2023. 11. 17. (6개월)		
연구금액	금 일천만원정 (부가가치세 포함)		
계약방법	■ 수의계약		
연구결과	<p>○ RoBERTa 모델(KLUE-RoBERTa-Large)의 도메인 적응 및 미세조정을 통한 예측 - 한국어 사전학습 언어모델인 KLUE-RoBERTa-Large 모델에 2015년~2020년간 자동차 사고 1심 판결문 등 다수의 법률 코퍼스를 추가 투입하여 도메인 적응적 사전훈련을 하여 특화된 모델을 학습시킨 후 과실상계 예측이라는 하류작업에 맞추어 미세조정하는 방식으로 어느 정도의 성능향상이 있는지를 측정 → 3분위 분류 정확도: 보행자 사고 75.86%, 전체 사고 70.18%</p> <p>○ 도메인 적응된 RoBERTa 모델의 임베딩과 코사인 유사도를 활용한 예측 및 애플리케이션의 가능성 검토 - 2015년~2020년간 자동차 사고 1심 판결문 중 과실상계 판단 부분에 현출된 사실관계를 위와 같이 도메인 적응된 RoBERTa 모델의 임베딩에 배치하여 벡터공간에서의 위치를 구한 후, 이를 근거로 각각의 판결문과 다른 판결문의 과실상계 판단 부분의 위치와의 코사인 유사도를 계산하여 유사 판결을 검색하고 상위 검색 판결들의 과실상계 비율을 평균하거나 다수결하는 방식으로 과실상계 비율을 예측한 후 성과를 측정 → 보행자 사고 평균제곱오차의 제곱근은 16.6%, 예측정확도는 65.9%. 전체 사고 평균제곱오차의 제곱근은 18.1%로 측정. 예측정확도는 67.5%</p>		
평가항목	상	중	하
연구목적과의 부합성	○		
내용의 완결성		○	
구성, 체제의 적정성		○	
참고문헌의 충실도		○	
학술적, 실무적 가치		○	
제출기간 준수	○		
용역수행자의 성실성	○		
연구결과 활용가능성	○		
평과 결과 총평	별지와 같음		
공개 여부	<input checked="" type="checkbox"/> 공개 <input type="checkbox"/> 비공개		
비공개 사유	공공기관의 정보공개에 관한 법률 제9조 제1항 <input type="checkbox"/> 1호 <input type="checkbox"/> 2호 <input type="checkbox"/> 3호 <input type="checkbox"/> 4호 <input type="checkbox"/> 5호 <input type="checkbox"/> 6호 <input type="checkbox"/> 7호 <input type="checkbox"/> 8호		
2023. 11. 20.			
법원행정처 정책연구용역심의위원회			

평가 결과 총평

- ◎ 본 정책연구용역은 서울대학교 산학협력단이 2020. 12. 17. ~ 2021. 9. 30. 수행한 “손해배상 사건에서의 인공지능(AI) 활용방안” 연구의 후속과제(교통사고 판결 과실상계 비율 예측에 관한 추가연구)에 관한 것임 → 선행연구과제 완료 이후 약 2년간의 GPT, BERT 등 Transformer 기반 사전학습 언어모델이 급격하게 발달하였기에 이들을 반영한 새로운 모델로 재검증을 해 볼 필요성이 생겼기 때문임
- ◎ 이에 본 정책연구용역이 이루어졌으나, 애초 다음과 같은 한계들이 있었기에 본 정책연구용역의 결과물이 재판실무에 바로 사용될 수 있을 정도의 정확도를 보일 것으로 기대하기는 곤란하였음
 - 1) 본 연구의 예산은 1,000만 원, 연구기간은 6개월로 단기간 내에 소규모의 인력이 투입될 수밖에 없었음
 - 2) 사법부에서 선행연구에 제공한 비실명화 처리를 한 판결문들은, 손해배상(자) 4,336건, 구상금 1,652건, 보험금 248건, 채무부존재확인 659건에 불과하였음 - 통상의 인공지능 모델의 훈련, 검증에 사용되는 데이터 양과는 비교가 되지 않을 정도의 적은 양이었음 → 본 연구에도 같은 데이터들이 사용됨
 - 3) 이에 연구자는 입수 가능한 데이터들(손해보험협회 자동차사고 과실비율분쟁 심의사례집, 엘박스 판결문 데이터, 도로교통법 전문, 찾기 쉬운 생활법령정보 백문백답, LegalQA데이터)을 추가하였으나, 위 데이터들 역시 소량이고, 인공지능 활용 용도로 만들어진 것들이 아니었으며, 질이나 정확도가 검증되지도 않았음
 - 4) 최근의 언어모델의 발전을 이끄는 폐쇄모델 GPT-4를 사용할 경우, 상당한 성능향상이 기대됨에도, 위 미공개 판결문들을 GPT-4 API에 투입할 경우 OpenAI 소유의 해외 서버에 미공개 판결문이 전송되는 결과가 초래되기에, 개방모델인 RoBERTa를 연구수행기관 시스템에 장착하여 이용할 수밖에 없었음
- ◎ 본 정책연구용역의 결과가, 선행연구결과와 마찬가지로, 법관의 과실상계 비율에 대한 판단을 신뢰성 있게 보조할 수 있는 수준에는 도달하지 못하였고, 판결서 작성 시 판결이유에 바로 인용할 수 있을 정도의 예시 문장을 제시해주는 수준이 아

님은 위 분류 정확도나 예측 정확도의 수치를 보면 알 수 있음(연구자도 인정하고 있음)

- ◎ 그러나 선행연구에서 BERT 모델을 사용하여 보행자 판결문에 대해서만 69.9%를 얻은 것에 비해, 본 정책연구용역에서 개발한 언어모델은 상당한 성능향상을 보이고 있고, 비록 선행연구에서 매뉴얼한 레이블링을 했을 때 얻은 알고리즘 별 75.4 ~ 80.8%의 예측정확도보다는 약간 떨어지나, 레이블링 작업을 전혀 하지 않고 그에 비견할 만한 예측정확도를 확보한 것은, 위 한계들을 고려할 때 대단히 고무적인 성과라고 평가할 수 있음 → 새로운 모델 개발의 가능성, 잠재력을 충분히 인정할 수 있음
- ◎ 본 정책연구용역은, 미공개 하급심 판결문들을 이 연구에 투입된 것보다 다량 투입할 경우 예측모델의 성능이 크게 향상될 수 있을 것이고, 판결문에 정리된 사실 관계로부터 판결의 결과를 예측하기보다는 당사자가 제출한 서면으로부터 판결의 결과를 예측하는 모델의 개발로 이행이 이루어지고, 이미지와 텍스트 등 다양한 포맷의 데이터를 처리하는 멀티모달(multi-modal) 기법을 도입할 경우 예측력이 더욱 향상될 것이며, 나아가 법관의 과중한 업무를 경감할 수 있는 “생성” 모델의 개발에 본 연구가 기여할 수 있을 것임을 지적하고 있는데, 모두 타당한 지적으로 보임(다만 법원이 자체 서버와 GPU 등 장비를 구축하여 언어모델을 훈련시키는 방안은 비용효율적이지 않은 측면이 있으므로, 판결문 공개 범위의 조정 에 대한 논의가 필요할 수 있다는 지적에 대하여는 추가 검토가 필요함)
- ◎ 이러한 점들 및 본 정책연구용역 보고서에 언어모델 제작에 필요한 코딩(Python 코드)까지 포함되었고 그 응용방법까지 기술되어 있는 점 등을 종합적으로 살펴볼 때, 본 정책연구용역은 신속하고 적절한 재판의 구현을 위한 수단 중 하나로 ‘재판업무에 인공지능을 활용하는 실제 방안’의 기초자료로서 활용가치가 매우 높다고 볼 수 있음