

대규모 언어모델을 활용한 리걸 마인드 개발의 현황과 전망*

경기룡** · 이상엽 · 김소연 · 인경석 · 지원석*** · 홍대식****

- I. 서론
- II. 자연어 모델의 발전 현황 및 현재의 법률 인공지능 서비스
- III. ChatGPT를 기반으로 한 리걸 마인드 형성 실험
- IV. 법률 인공지능의 발전 전망
- V. 결론

국문초록

본 논문은 대규모 언어모델이 리걸 마인드(Legal Mind)를 가질 수 있는지에 관하여, 특히 변호사 시험 문제를 해결하기 위한 능력을 가질 수 있는지에 대한 물음을 통해 그 현황을 살펴보고 전망을 제시한다. 본 연구는 먼저, 자연어 인공지능 모델의 발전사를 소개한다. 순환신경망(Recurrent Neural Network, RNN)부터 임베딩(Embedding)을 위한 단어의 벡터 전환(Word to Vector) 기술, 트랜스포머(Transformer) 구조를 활용한 BERT(Bidirectional Encoder Representations from Transformers)를 거쳐, 1,750억 개의 파라미터를 가진 GPT 같은 대규모 언어모델들의 발전 과정을 서술하고, 최근 법률 분야에 자연어 인공지능 모델의 활용이 화두가 되고 있는 이유를 설명한다.

다음으로 ChatGPT 프롬프터(Prompter)에 한국의 민법 규정과 대법원 판례를 입력함으로써 새로운 학습과정 없이도 ChatGPT에게 점진적으로 리걸 마인드를 형성시킬 수

* 이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2022-0-00621, 대화 기반 설명가능성을 멀티모달로 제공하는 인공지능 기술 개발)

** 서강대학교 법학전문대학원 법학전문석사과정, 제1저자

*** 이상 서강대학교 법학전문대학원 법학전문석사과정

**** 서강대학교 법학전문대학원 교수, 교신저자

있음을 보여준다.

마지막으로 법률 인공지능 서비스의 발전 전망을 소개하고, 리걸 마인드를 탑재한 대규모 언어모델이 법률 시장에 미치는 영향을 논의한다. 대규모 언어모델에 리걸 마인드를 형성시키는 연구는 법률 시장에 혁신적인 변화를 가져올 것으로 기대되며, 본 논문에서는 기술 발전에 따른 앞으로의 법률 시장 변화를 제안한다.

• 핵심어 대규모 언어모델, 리걸 마인드, 법률 인공지능 서비스, 리걸테크

I. 서론

1. 자연어 인공지능의 현재

2022년 11월에 공개된 OpenAI의 ChatGPT는 인공지능 자연어처리 분야에서 커다란 반향을 일으켰다. 사람처럼 대화를 생성하는 수준이 이전과 비교해서 뚜렷하게 좋아졌고, 긴 대화를 주고받음에 있어서도 이전의 대화형 모델에서 한층 나아간 모습을 보여주었다. 무엇보다 ChatGPT의 추론(reasoning)¹⁾ 능력이 크게 증가하였다는 점이 주목할 만한 점이다.

인공지능 모델의 추론 능력이 증가함에 따라, 인공지능 기술을 법률에 적용하고자 하는 시도가 증가하였다. 개선된 인공지능 기술은 기존의 모델들이 어려움을 겪었던 법률서 요약, 판례 분석 등에서 탁월한 성과를 보이고 있다. 인공지능 기술의 발전으로 인해 법률 분야에서도 높은 수준의 자동화와 효율성을 기대할 수 있게 된 것이다.

법률 분야는 끊임없이 발전하고 변화하는 분야 중 하나이다. 새로운 사례와 판례, 법률 규정 등이 지속적으로 업데이트되는 것은 물론이고, 사회가 변화함에 따라 발생하는 거의 대부분의 문제들이 법률 문제로 이어지기 때문이다. 따라서 법조인 개인에게도 점점 빨라지는 사회변화에 대응하는 데에 인공지능 기술 활용이 크게 도움이 될 것으로 예측된다.

1) 추론은 사고력의 이용을 통한 유추나 결론의 도출을 의미한다. 인공지능에서의 추론이란 추론을 제공하는 컴퓨터 모델(computational model)의 개발에 관한 문제이다. Ashley, Kevin D., Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age, Cambridge, 2017(이 책의 한국어 번역본은 오태운·정영수·조동관 역, 『법을 분석하는 인공지능』, 박영사, 2020), pp.32-34.

2. 법률 영역에서 자연어 인공지능 기술이 활용될 수 있는 영역

대규모 언어모델(Large Language Model, “LLM”)이 등장하면서 다양한 자연어 처리 문제들을 해결하고 있다. 법률분야에서도 어떤 문제를 풀기 위해서 자연어 인공지능 기술을 어떻게 적용할지가 중요한 연구 대상이다. 이하에서는 판례 검색 및 유사 판례 연구와 법률 형량 예측 연구를 중심으로 선행연구를 검토하겠다.

(1) 판례 검색 및 유사 판례 연구

쟁점이 되는 내용에 부합하는 또는 유사한 판례를 찾는 것은 법률 전문가뿐만 아니라 일반시민에게도 필요한 법률서비스이다. 여러 법률 데이터 중에서 판례가 법률적 판단이 가장 압축된 형태로서 판단 근거를 제공하기 때문이다.

해외에서는 변호사시험 등 정제된 문제가 아닌 일반적인 시민들의 질문을 활용하고, 영어가 아닌 각 국가의 모국어를 활용하는 연구가 제시되었다.²⁾ 예를 들어 ‘동산 압류’의 개념에 대한 질문을 한다면, 법률가들이 최초로 라벨링한 카테고리인 ‘금전-채무회복’의 범주에서 답안을 제공하는 것이다. 이 연구는 프랑스어 기반으로 데이터를 활용하여 프랑스어 법령조문검색 모델(Belgian Statutory Article Retrieval Dataset)을 구축하였다는 의의가 있으나, 벨기에의 32개 법률 조문만이 포함하였기 때문에 수많은 법령, 판례, 지침을 누락하여 보편적인 모델을 구축하지 못한 한계가 있다.

최근에는 LLM 내에서 단어(Word)를 벡터(Vector)로 변환하는 방법인 단어의 벡터 전환 방식(“Word to Vector”)³⁾이 많이 활용되고 있다. 단어의 벡터 전환 방식은 크게 Sparse 방식과 Dense 방식으로 나뉜다. Sparse 방식은 벡터에 특정 단어와의 일치 여부에 대한 수치를 출력하는 것이다. 반면 Dense 방식은 벡터에 단어와의 유사 여부를 수치로 출력한다. 판례 검색도 검색한 단어와의 정확한 일치보다는 검색한 단어와의 유사성이 더 중요하게 평가되며, Dense 방식을 적용하는 연구들이 있다.

그리고 ‘단어’가 아닌 ‘단락’(Paragraph)을 기본단위로 하는 연구들도 진행 중이다. 이는

2) Antoine Louis and Gerasimos Spanakis, “A Statutory Article Retrieval Dataset in French”, arXiv: 2108.11792 [cs.CL], 2021.

3) 단어의 벡터 전환(Word to Vector) 방식은 단어를 벡터로 표현하는 방법으로, 각 단어 간의 유사도를 벡터화하여 해당 단어의 의미를 수치화하는 알고리즘이다. 단어의 의미는 [0,1]의 디지털 방식으로 구현하기 어렵다. 따라서 ‘문장 내 비슷한 위치에 등장하는 단어는 비슷한 의미를 지닌다’는 가정을 기반으로, 단어 간 유사도를 벡터화하여 ‘단어의 의미’라는 정보를 담고자 한다. 더 자세한 설명은 아래 II. 4. (3) 참조.

절대 다수의 LLM 모델들이 입력 토큰에 대해서 제한을 두고 있기 때문이다. 기본 단위의 텍스트 규모를 단락으로 확대하고 Dense 방식을 적용할 경우, 방대한 법률 문서를 효과적으로 활용할 수 있게 된다.

실제로 Sparse 방식과 Dense 방식을 평가할 때, Dense 방식을 적용한 경우 더 높은 성능을 보였다. 특히, 단어가 아닌 구절을 기본 단위로 할 경우 검색할 때 주제와 유사한 문서를 도출하는 데 효과적이었다. 특정 단어보다 문장과 문서 내의 맥락을 기반으로 검색할 때 부합하는 관계가 도출되기 때문이다.⁴⁾ 또한, 검색 단계에서 문서 간 의미 수준의 유사성을 추가적으로 고려하여 검색하는 단위를 결정하는 방식('LeCut' 방식)도 제시되고 있다.⁵⁾ 다른 영역에서와 마찬가지로 법률 및 판례 검색 영역에서도 Word to Vector 방식이 원하는 검색결과를 포함할 확률이 유의미하게 높아졌다.

(2) 법률 형량 예측

형법은 법률 중에서도 자연어 처리 기술이 적용되기 쉬운 분야이다. 특정 사안에 대한 구성요건과 법정형이 법률에 명확히 규정되어 있기 때문이다. 많은 연구들은 사실관계 내에서 정확한 형량 예측을 목표로 하고 있다.

사례 내에서 세부적인 내용들을 분류하고, 각 내용들 간의 연계성을 강화하는 것이 효율적인 방법으로 제시된다. 기존의 사실관계 내에서 내용을 최대한 세분화해서 분석하는 모델에 비해서는 높은 효율을 보이고 있다. 위상학습⁶⁾ 방식을 기반으로 하위작업 간의 종속성을 적용한 방법도 유의미한 개선을 나타내었다.⁷⁾ 또한, 연계성을 강화하기 위한 구체적인 방법으로 사례 내에서 세부 내용을 계층화(예: 법률조항, 죄명, 형량)하고, 계층 간에

4) Sophia Althammer, Sebastian Hofstätter, Mete Sertkan, Suzan Verberne, Allan Hanbury, "PARM: A Paragraph Aggregation Retrieval Model for Dense Document-to-Document Retrieval", arXiv: 2201.01614 [cs.LG], 2022

5) Ma, Yixiao Ai, Qingyao Wu, Yueyue Shao, Yunqiu Liu, Yiqun Zhang, Min Ma, Shaoping, "Incorporating Retrieval Information into the Truncation of Ranking Lists for Better Legal Search", SIGIR '22: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022.

6) 위상학습(topological learning)은 기계 학습 및 인공지능 분야에서 사용되는 용어로, 데이터의 구조와 패턴에 대한 학습을 의미한다. 이는 이미지, 음성, 텍스트 등 다양한 형태의 데이터에서 특징을 추출하고 그것들 간의 관계를 학습하는 프로세스를 나타낸다.

7) Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, Maosong Sun, "Legal Judgment Prediction via Topological Learning", Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3540-3549, Brussels, Belgium, Association for Computational Linguistics, 2018.

모순이 발생하지 않도록 제약조건을 설정하는 방법도 제시되고 있다.⁸⁾

3. 변호사시험 문제 해결을 위한 리걸 마인드

(1) 변호사 실무와 리걸테크(Legal Tech) 분야

리걸 마인드(Legal Mind)가 법률 시장과 관련하여 어떠한 부분을 목적으로 하는 리걸테크 분야가 될 수 있는지를 명확하게 하기 위하여 우선 변호사 실무의 일반적인 구조를 간략히 설명한다. 통상적인 변호사들의 업무는 법적 쟁점의 해결이라는 측면에서 바라볼 때 4단계로 이루어진다. 가장 먼저 서면 등을 포함한 기록을 읽고(이하 [1]이라 한다), 사안에서 불필요한 부분을 제외한 사실관계를 구성한다(이하 [2]라 한다). 이후 사실관계에서 법적으로 문제가 될 수 있는 법적 쟁점을 추출하며(이하 [3]이라 한다), 각각 쟁점에 맞는 법령과 판례를 찾아 결론을 도출한다(이하 [4]라 한다).

리걸테크 분야는 법률 시장에 편의성을 가져다주면서 경제적 이득을 취하는 것을 본질로 하기에, 그 분류에 있어 위와 같은 변호사 업무의 부분들과 직접 비교하는 것이 가장 보편적이라고 할 수 있다. 가령 국내 리걸테크 분야의 대부분을 차지하는 법률검색 모델(투스론이터의 Westlaw Edge, Westlaw Precision, 엘박스의 검색시스템 등)은 [3]단계까지 변호사가 업무를 마친 후 [4]단계의 업무를 돕는 것을 목적으로 한다고 평가할 수 있다.

(2) 리걸 마인드의 정의 및 목적

리걸 마인드라는 표현은 1933년에 미국의 풀러(Fuller)가 발표한 논문에서 유래하는데,⁹⁾ 법적 사고력 또는 법적 사고방식을 뜻하는 말로 사용되는 용어이다. 특히 우리나라에서

8) Yi Feng, Chuanyi Li, Vincent Ng, "Legal Judgment Prediction via Event Extraction with Constraints", Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 648-664, Dublin, Ireland, Association for Computational Linguistics, 2022.

9) Fuller, L. L., "The Legal Mind", *The Atlantic*, 1933. 하버드 대학교 파웰(Powell) 교수는 1969년에 풀러를 인용하여 리걸 마인드를 "만약 당신이 그것과 관련된 것에 대해 생각하지 않고 어떤 것에 관련된 것에 대해 생각할 수 있다면, 당신은 리걸 마인드를 가진 것이다"라고 설명하였다. Brozek, Bartosz, *The Legal Mind: A New Introduction to Legal Epistemology*, Cambridge University Press, 2020; 브로첵 교수는 법적 사고를 철학적 개념인 법적 추론과 연결시키는 통념과 달리 인지과학의 발견도 반영하여 법적 사고를 직관, 상상 그리고 언어적 사고라는 3가지 인지적 메커니즘의 동시적 사용과 협력으로 설명한다. 이 글에서는 통념에 따라 법적 사고를 법적 추론을 하는 사고로 보되, 법적 사고를 하는 존재를 인지심리학적 개념인 마음을 사용하여 리걸 마인드로 표현하는 것은 이런 접근과도 연결될 수 있을 것으로 생각한다.

법조인이 갖추어야 할 기본적인 소양으로 많이 일컬어진다. 리걸 마인드와 구별될 수 있는 용어는 법적 추론(legal reasoning)이다. 법적 추론은 추론 중에서도 특별히 법적 논증 및 의사결정에서 집중적으로 두드러지게 나타나는 추론의 방법인데, 법적 추론이란 것이 존재한다는 것과 그 결과 우리가 ‘법률가처럼 사고하기’(thinking like a lawyer)라고 정확하게 특징지을 만한 무언가가 존재한다고 전제할 수 있다.¹⁰⁾ 리걸 마인드는 법적 사고를 인지심리학에서 신체(body)에 대응하여 정신활동이 이루어지는 영역 또는 틀을 가리키는 마음(mind)라는 용어와 연결시킨다. 법적 사고를 법적 추론을 하는 사고로 정의한다면, 리걸 마인드는 법적 추론을 하는 마음이라고 할 수 있다. 여기서 법적 추론이 기능 또는 작용이라면 마음은 그런 기능 또는 작용을 하는 존재이다. 인공지능 분야에서 생각하는 기계를 만든다는 것은 인간의 생각에 대한 흉내 게임(imitation game)을 잘해내는 존재인 에이전트(agent)를 만들어내는 것을 뜻하므로,¹¹⁾ 이를 리걸테크 분야에 적용할 때 법적 추론과 관련된 인공지능 기술을 리걸 마인드로 이름 붙일 수 있겠다.

리걸 마인드를 위와 같이 정의하고 앞서 본 변호사들의 업무의 4단계 분류에 따르면, 리걸 마인드는 [2]단계 사실관계를 입력하여 [3]단계 쟁점 추출과 [4]단계 법령·판례에 따른 결론을 도출하는 것을 목적으로 하여 이런 기능을 수행하는 에이전트를 개발하는 리걸테크 분야의 기술이다. 이러한 쟁점·결론 추출 모델은 일상어로 표현된 사실관계를 법률 용어로 재구성한 후 결론까지 도출해준다는 점에서 가장 업무상 효용성이 높을 것으로 예상된다. 다만, 이러한 특성은 동시에 개발의 어려움을 보여주기도 한다.

따라서 쟁점·결론 추출 모델의 개발은 다양한 법적 쟁점을 해결하는 변호사 실무에 직접 적용하기에 앞서, 가장 전형적인 법적 쟁점의 해결을 목적으로 해야 한다. 이에 따라 법률 실무에서 가장 전형적인 법적 쟁점을 포괄하고 있는 변호사시험을 해결할 수 있는 모델을 개발하는 것이 이 연구에서 시도하는 리걸 마인드 개발의 목적이다.

10) Schauer, Frederick, *Thinking Like a Lawyer: A New Introduction to Legal Reasoning*, Harvard University Press, 2009(이 책의 한국어 번역본은 김건우 역, 『법률가처럼 사고하는 법: 법적 추론 입문』, 도서출판 길, 2019), pp.1-12.

11) 김재인, 『인공지능의 시대, 인간을 다시 묻다』, 동아사이, 2018, 25-56면. 인공적 에이전트는 생각하고와 같이 객관화하기 어려운 마음의 영역을 갖는 존재를 인간이건 기계이건 상관없이 포괄하기 위하여 스튜어트 러셀과 피터 노빅이 사용한 개념이다(스튜어트 러셀, 피터 노빅 저, 류광 역, 『인공지능: 현대적 접근방식(제3판) 제1권』, 제이펍, 2016, 5-6면). 생각하는 기계에 대한 화두를 던져 인공지능 구축의 시대를 연 튜링은 에이전트나 마음이라는 용어를 사용하지는 않으면서 어떤 에이전트가 생각하는지를 판정하는 기준을 제시하였는데, 그 주장을 담은 튜링의 논문이 1950년에 실린 저널 이름이 “Mind”라는 점은 의미심장하다(Turing, A. M., “Computational Machinery and Intelligence”, *Mind*, Volume LIX, Issue 236, 1950, pp.433-460).

(3) 쟁점·결론 추출 모델의 개발 정도

후술하는 인공지능 리걸 마인드(AI Legal Mind) 형성에서 구체적인 프롬프팅(Prompting) 방법론(실험 설계)을 소개하고 그에 따른 실험 결과값을 통하여 리걸 마인드 개발의 정도를 상세히 설명하겠지만, 리걸테크 분야 중 쟁점·결론 추출 모델에 대한 실험 사례, 특히 GPT가 미국의 변호사시험(Bar exam)을 얼마나 잘 해결하는지에 대한 실험 사례를 통하여 리걸 마인드 분야의 전반적인 개발 정도를 이해할 수 있다.

현재까지 쟁점·결론 추출 모델에서 가장 명확한 결과값을 보여준 것은 GPT로 미국 변호사시험을 해결하려는 시도이다.¹²⁾ 이 논문에서는 사람이 리걸 마인드를 활용하여 변호사시험 문제를 해결하려는 것과 유사하게 GPT가 변호사시험 문제를 해결하도록 하였고, 특히 GPT-4는 합격 수준을 넘어 70% 이상의 높은 정답율을 보여주었다고 설명한다. 하지만 이러한 정답율은 다음과 같은 이유에서 그대로 신뢰하긴 어렵다. 먼저 주관식 시험이라는 특성상 평가자가 누구인지에 따라 그 정답율이 달라지는데 그 평가자가 일부 다른 연구자의 블라인드 평가로 이루어졌으며 지표가 인간 선호도(Human preference)인지 정확성(precision)인지도 명확하지 않아 실제로 70%의 정답율을 보였다고 확신하기 어렵다. 또한 실험은 반복성이 매우 중요한데, 위 논문에서는 2022년 7월에 실시된 단 하나의 시험 문제를 풀게 하였다. 이에 따라 매년 70% 이상의 정답율을 보일 것이라고도 확신하기 어렵다. 이에 더하여 GPT로 다른 변호사 시험을 실험하고 채점해본 결과 합격 수준에 한참 못 미치는 정답율을 자주 보이기도 하였다.

리걸테크 분야 중 쟁점·결론 추출 모델은 결국 변호사 실무에 적용될 정도로 개발된 적이 아직 없으며, 그 전 단계로 전형적인 법적 쟁점을 포괄하는 변호사 시험 문제를 해결하는 것에 대하여도 아직 합격 수준으로 개발된 정도가 없다고 할 수 있다.

4. 리걸 마인드를 탑재한 대규모 언어모델이 법률 시장에 미칠 수 있는 영향

자연어 인공지능의 발전이 법률 시장 및 법조인의 업무에 큰 변화를 가져올 것이라는 점은 자명하다. 예컨대, 법률 인공지능은 영미법계 국가의 민사소송절차 중 하나인 전자증거개시(eDiscovery), 유사 판례 및 법리 검색(legal search), 판결문이나 계약서 등의 법률

12) Katz, Daniel Martin and Bommarito, Michael James and Gao, Shang and Arredondo, Pablo, "GPT-4 Passes the Bar Exam", 2023. Available at SSRN: <https://ssrn.com/abstract=4389233>

문서의 작성 등에서 가장 활발하고 효과적으로 이용될 것이라고 기대되고 있다. 그러나 법률 인공지능이 가지게 될 정확성, 범용성, 자연어 처리 능력을 고려하면 앞서 말한 방안들뿐만 아니라 현재 미처 예상하지도 못하고 있는 법조 업무에 활용될 가능성 또한 충분히 존재한다.

결과적으로 리걸 마인드를 갖춘 법률 인공지능은 수많은 법률 문서를 빠르게 요약하고 유사한 사실관계를 찾아내는 데에 뛰어난 능력을 가지게 될 것으로 기대되는데, 이를 활용하는 다양한 방안들의 핵심은 결국 반복적이고 시간 소모적인 작업들을 처리함으로써 변호사들이 보다 가치 있는 작업에 집중할 수 있게 보조하는 수단으로서 기능하는 데에 있다.

‘법조인의 지능형 비서’(intelligent assistants)로서 법조업무의 효율성을 증진시키는 법률 인공지능의 발전 양상의 전망과 관련하여 변호사와 사법부 각각에 미칠 영향과 ODR의 발전에 기여함으로써 국민의 사법 접근권을 제고하는 효과로 나누어 IV.에서 후술한다.

II. 자연어 인공지능 모델의 발전 과정 및 현황

1. 자연어 인공지능 발전사 소개 이유

인간이 사용하는 자연어를 컴퓨터에게 이해시키고, 컴퓨터가 자연어를 출력하게 하려는 시도는 과거부터 있어왔으나, 해당 기술을 법률 서비스에 적용하고자 하는 시도는 드물었다. 이하에서는 자연어 인공지능 모델이 어떻게 발전해 왔는지, 그리고 왜 최근에는 법률 분야에도 자연어 인공지능 기술이 적용되고 있는지 대해서 설명한다. 이러한 자연어 인공지능의 발전사를 알아봄으로써, 최근 다양하게 출시되는 법률 인공지능 서비스의 기능과 한계를 이해하는 데 도움이 될 수 있으리라 기대한다.

2. 자연어를 처리하기 위한 초창기 모델, 순환인공신경망의 출현

인간의 두뇌를 논리적으로 표현하려 한 최초의 시도는 1943년, 논리학자인 피츠(Pitts)와 신경외과의인 맥컬로치(McCulloch)의 논문에서 이루어졌다.¹³⁾ 그 뒤 현재 인공지능의 기

13) WARREN S. MCCULLOCH AND WALTER PITTS, "A Logical Calculus of Ideas Immanent in Nervous Activity", *Synthese* 162(2):235-250, 1943.

반인 인공신경(Perceptron) 개념을 1958년 로젠블라트(Rosenblatt)가 고안해냈고¹⁴⁾ 1982년 위 인공신경을 기초로 자연어 인공지능의 초기 모델이라 할 수 있는 순환신경망(Recurrent Neural Network, RNN)이 탄생하였다.¹⁵⁾

순환신경망은 여러 인공신경이 쌓인 단순한 인공신경망과 유사한 구조를 가졌지만, 입력 값과 출력 값이 각각 독립적인 일반 인공신경망과는 달리 순환신경망은 각 입력 값에 대한 출력 값이 다음 입력 값에도 영향을 미치면서 학습이 진행된다. 즉 서로 연관성을 가지는 연속적인 입력 값과 출력 값을 다루는 데 적합한 모델이므로 시계열 데이터인 음성, 기온, 주가 등을 비롯하여 자연어를 처리하는 데에도 적합한 모델이다.

그러나 초기 순환신경망의 단점은 입력하는 데이터의 길이가 길어질 경우, 그 성능이 매우 떨어진다는 점에 있었다.¹⁶⁾ 간단한 몇 문장 정도를 입력했을 경우에는 그 다음에 올 단어를 예측할 수 있었지만, 수십 문장이나 문서 형태로 된 데이터를 입력하게 되면 무작위의 단어를 출력하는 수준으로 모델의 성능이 떨어졌다. 이를 장기 의존성 문제(the problem of Long-Term Dependencies)라고 한다.

3. 순환신경망의 한계를 극복하기 위한 LSTM 모델

장기 의존성 문제를 해결한 것은 1997년에 나온 논문이었다.¹⁷⁾ 장단기 기억(Long Short-Term Memory, LSTM)이라는 뜻의 직관적인 제목을 가진 이 논문은 기존 순환신경망이 가진 치명적인 한계를 극복하고자 하였다. LSTM 모델이라 불리는 이 모델은 입력 게이트(Input gate)와 삭제 게이트(Forget gate) 구조를 이용하여 순환신경망의 ‘(1) 길이가 긴 데이터를 입력 받거나, (2) 중요한 데이터가 앞 단에 있는 경우에 성능이 극적으로 떨어지는 한계’를 해결하였다. LSTM 모델은 두 개의 게이트를 이용하여 입력한 데이터 중에서 ‘결과에 큰 영향을 미쳐서 마지막까지 기억해야 할 데이터’와 ‘결과에 별 영향을 미치지 않아 잊어버려도 되는 데이터’를 구분하여 학습하는 방식이다. 전자인 중요한 데이터는 입력 게이트로, 후자인 잊어버려도 되는 데이터는 삭제 게이트를 통과시킴으로써 길이가

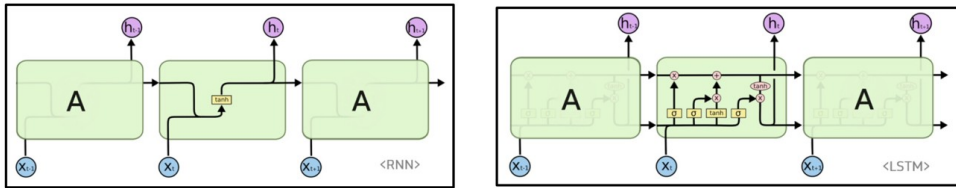
14) Rosenblatt, F., “The perceptron: A probabilistic model for information storage and organization in the brain”, *Psychological Review*, 65(6), 386-408, 1958.

15) J.J.HOPFIELD, “Neural networks and physical systems with emergent collective computational abilities”, *Proc.NatAcad.Sci. USA*, Vol.79,pp.2554-2558, 1982.

16) Yoshua Bengio et. al., “Learning Long-Term Dependencies with Gradient Descent is Difficult”, *IEEE Transactions on Neural Networks*, Volume: 5, Issue: 2, 1994.

17) Sepp Hochreiter et. al., “long short-term memory”, *Neural Computation* 9(8):1735-1780, 1997.

긴 데이터나 중요한 부분이 앞 단에 있는 데이터를 입력 받은 경우에도 만족할만한 결과 값을 가지게 되었다.



(그림 1)¹⁸⁾

4. 문서 내의 여러 단어를 제한된 공간에서 표현해 낸 단어의 벡터 전환(Word to Vector) 모델

(1) 단어 임베딩(Word embedding) 방식

사람의 언어인 자연어를 컴퓨터에게 곧바로 입력할 수는 없으므로, 자연어 인공지능을 연구하는 작업에서 자연어를 컴퓨터의 언어로 변환하는 과정은 필수적이다. 가장 일반적이고 현재 널리 사용되고 있는 방식은 자연어인 단어를 컴퓨터가 이해할 수 있는 벡터 공간(Vector space) 상에 일대일 대응을 시키는 방식이다. 각 단어를 벡터 공간에 박아 넣는 작업이라 하여 단어 임베딩이라고도 부른다.

(2) 직관적이지만 비효율적인 원핫 인코딩(one-hot encoding) 방식

초기 자연어 인공지능 모델은 데이터에 존재하는 모든 단어의 개수만큼의 벡터 차원을 생성한 뒤에 해당 단어 위치에는 1, 나머지는 0을 부여하는 원핫 인코딩 방식을 채택하였다. 예를 들어 모델에 입력하는 데이터가 “서강대학교 법학전문대학원” 이라면 “서강”, “대학교”, “법학”, “전문”, “대학원”의 다섯 개의 단어가 존재하므로, “서강”은 [1, 0, 0, 0, 0]로 표현하고, “대학교”는 [0, 1, 0, 0, 0]로 표현하는 방식이다. 이러한 방식은 단순하고 직관적이지만, 매우 비효율적이다. 단어가 1,000개 정도 되는 아주 짧은 문서를 분석하려고 해도

18) <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

1,000 차원의 벡터가 필요해지기 때문이다. 문서가 길어질수록 계산에 들어가는 비용이 급격히 증가하게 되는데 이를 차원의 저주(The curse of dimensionality)라고 부른다.

(3) 단어의 의미를 벡터 공간 상에 표현하는 단어의 벡터 전환 모델

이후 원형 인코딩의 단점을 보완하는 변환 방식들이 많이 개발되었는데, 변화의 시작점이 된 모델이 2013년에 공개된 단어의 벡터 전환(Word2Vec) 모델이다.¹⁹⁾ 단어의 벡터 전환 모델도 마찬가지로 자연어를 벡터 공간 상에 일대일 대응시키는 워드 임베딩 방식이지만, 벡터의 차원을 미리 정해 놓은 뒤에 단어의 의미를 벡터 공간 상에 투영시키는 학습과정을 거친다. 예를 들어서 “서강대학교 법학전문대학원”이라는 데이터를 3차원 벡터로 표현하기로 정해 놓은 뒤에 “서강”은 [0.2, 0.08, -0.1]로 표현하고 “대학교”는 [0.8, -0.9, 0.1]로 표현하고 “대학원”은 [0.79, -0.91, 0.1]로 표현하는 방식이다. 예시이긴 하지만, “대학교”와 “대학원”은 연관성이 높으므로 벡터 공간 상에서도 유사한 곳에 위치하도록 표현된다.

단어의 벡터 전환 모델은 사람이 직접 유사한 단어를 비슷하게 표현하지 않는다. 다만 대량의 말뭉치 데이터(corpus)가 주어졌을 때, 인공지능 모델이 직접 유사한 단어를 벡터 공간 상에 근접하게 표현하도록 학습한다. 모델 스스로 특정 단어를 지우고, 그 지운 단어에 들어갈 법한 데이터를 예측하는 방식으로 학습이 진행된다. 예를 들어 “부동산 매매계약에서 ()만 지급된 단계에서는 어느 당사자나 이를 포기하거나 그 배액을 상환함으로써 자유롭게 계약의 구속력에서 벗어날 수 있다.”라는 문장이 주어지면 모델은 빈 칸에 들어갈 단어가 “계약금”이라는 것을 학습하는 방식이다.

단어의 벡터 전환 모델이 개발된 뒤로, 입력되는 문장이 길거나 문서 자체가 입력되는 경우에도 적은 크기의 벡터를 이용해 자연어를 컴퓨터의 언어로 변환시킬 수 있게 되었다.

5. 현재 자연어 모델의 기본 구조, 트랜스포머(Transformer)

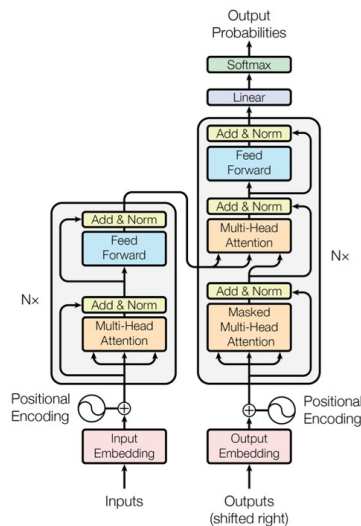
(1) 어텐션(Attention) 기법으로부터 비롯된 트랜스포머

현재 가장 성능이 좋은 자연어 인공지능 모델들, GPT나 BERT 등의 모델들은 거의 다수가 트랜스포머 구조로 되어 있다. 위 트랜스포머 구조는 2017년에 공개된 “Attention

19) Tomas Mikolov et. al., “Efficient Estimation of Word Representations in Vector Space”, arXiv: 1301.3781 [cs.CL], 2013.

Is All You Need”라는 혁신적인 논문²⁰⁾에서 비롯되었다. 논문의 제목에서부터 알 수 있듯이 트랜스포머 구조는 어텐션이라는 기법을 바탕으로 만들어졌는데, 이 어텐션 기법은 자연어 모델에서 입력 값과 출력 값 사이의 관련성을 알기 위해서 만들어진 기법이다. 2014년에 제안된 어텐션이라는 기법은, 모델이 특정 단어를 출력하였을 때 그 출력 값에 각 입력 단어들이 얼마나 영향(attention)을 주었는지 알기 위해 제안된 방법이다.²¹⁾ 출력 값에 영향을 준 정도를 입력한 단어의 개수 길이만큼의 어텐션 벡터(Attention Vector)로 표현하였다. 트랜스포머 구조는 이 어텐션 벡터만을 사용하여 만든 구조체이다. (〈그림 2〉)

(2) 트랜스포머의 구조 분석



〈그림 2〉²²⁾

복잡해 보이는 〈그림 2〉의 트랜스포머 구조는, 크게 두 부분으로 나눌 수 있다. 입력 값이 들어가는 인코더(Encoder) 부분과, 출력 값이 나오는 디코더(Decoder) 부분이다. 쉽게 말해서 인코더 부분은 자연어를 벡터 공간 상에 표현하는, 즉 단어 임베딩 작업을 진행하는 부분이다. 반대로 디코더는 벡터 공간 상에 존재하는 벡터 값을 다시 자연어로 출력

20) Ashish Vaswani et. al., “Attention Is All You Need”, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017.

21) Dzmitry Bahdanau et. al., “Neural Machine Translation by Jointly Learning to Align and Translate”, arXiv: 1409.0473 [cs.CL], 2014.

22) “Attention Is All You Need” (2017)의 Figure 1.

하는 부분이다. 다시 말해서 잘 학습된 트랜스포머 구조체는 입력된 자연어를 컴퓨터의 언어로 바꾸는 부분과, 컴퓨터의 언어를 다시 자연어로 바꾸는 부분으로 구성되어 있다고 볼 수 있다.

(3) 트랜스포머 구조를 사용한 대표적인 자연어 모델, BERT와 GPT

최신 자연어 모델 두 가지는 모두 트랜스포머 구조를 기반으로 한다. 2019년에 나온 BERT²³⁾와 2018년에 나온 GPT²⁴⁾가 그것이다. BERT는 트랜스포머의 인코더 부분만을 사용하였고, GPT는 디코더 부분만을 사용하였다. 즉 BERT는 자연어를 컴퓨터 언어로 바꾸는 단어 임베딩에 특화되어 있고, GPT는 컴퓨터의 언어를 자연어로 바꾸는 작업에 특화되어 있다.

BERT는 자연어를 컴퓨터의 언어로 바꾸어 유사한 문서들이 무엇인지 잘 판별해 낼 수 있다. 이는 비슷한 뉴스나 논문을 추천해 주는 서비스에 이용되고 있는데, 요즘 법률 시장에서는 사실관계가 비슷한 판례를 찾아주는 서비스에도 이용되고 있다.²⁵⁾

GPT는 컴퓨터의 언어를 자연어로 바꾸어주는 기능, 즉 챗봇에 특화되어 있다. 최근 가장 큰 영향력을 보이고 있는 ChatGPT가 그 대표적인 예이다.

6. 자연어 모델 발전사를 통해 본 법률 인공지능 서비스의 현재와 미래

앞서 살펴봤듯이, 초기 자연어 인공지능 모델은 두 가지의 뚜렷한 한계점을 가지고 있었다. (1) 입력하는 문장의 길이가 길어지면 성능이 좋지 않아진다는 것과, (2) 낮은 수준의 단어 임베딩 방식이 그것이다. 법률 문서의 경우에는 그 길이가 길고 사용되는 단어가 일반 문서와 다른 경우가 많아 초기의 자연어 인공지능 모델로는 접근이 불가능하였다. 그러나 LSTM 모델처럼 긴 문장을 입력하기 위한 방법론들이 개발되었고, 트랜스포머 구조를 이용한 효율적인 단어 임베딩 방식이 개발됨에 따라 자연어 인공지능의 성능이 비약적으로 증가하였다. 판례나 법 관련 문서들을 입력 데이터로 받아들이 수 있게 되었고, 그렇게

23) Jacob Devlin et. al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", arXiv: 1810.04805 [cs.CL], 2019.

24) Alec Radford et. al., "Improving Language Understanding by Generative Pre-Training", 2018. Available at: <https://www.mikecaptain.com/resources/pdf/GPT-1.pdf>

25) LBOX의 "유사판례 검색" 서비스.

받아들이는 데이터를 기초로 비슷한 판례를 묶어주거나 사람과 유사하게 자연어로 문답을 할 수 있게 되었다.

무엇보다 기존 대화형 인공지능 모델은 바로 직전의 입력 값이나 출력 값에 대한 질문에 제대로 답변하지 못하는 한계점을 가지고 있었기에, 왜 인공지능 모델이 해당 답변을 출력하였는지에 대한 이유를 알 수 없었다. 그러나 OpenAI의 ChatGPT는 사람에 가까운 문답능력을 보여주면서, 이전에 했던 질문과 답변들을 기억하고 그것을 토대로 새로운 답변을 출력해 낼 수 있다. 주어진 사실관계에서 특정 쟁점을 도출한 이유나, 그 쟁점에 대해 어떤 판례나 규정을 근거로 해결책을 제시하였는지 등 이전 인공지능 모델들은 가지지 못한 “설명성”을 확보하게 되었다.

즉 이제는 자연어 인공지능의 추론 능력이 리걸 마인드를 가질 수 있는 정도가 되었는지 실험해 볼 수 있는 단계가 되었다. 본 논문에서는 아래 III.에서 ChatGPT를 기반으로 자연어 인공지능에게 리걸 마인드를 형성시키는 실험을 진행한다.

III. ChatGPT를 기반으로 한 리걸 마인드 형성 실험²⁶⁾²⁷⁾

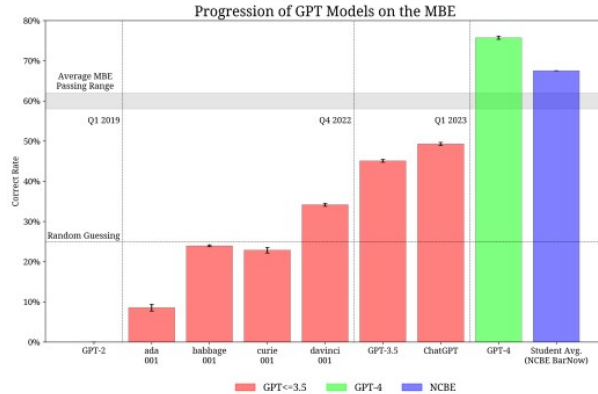
1. 연구 배경

대규모 언어모델은 다양한 자연어 처리 문제들에 적용되고 있으며 뛰어난 성능을 보이고 있다. 법률분야에 대해서도 언어모델의 활용을 위한 실험이 꾸준히 진행되고 있으며, 최근에는 GPT-4가 미국 변호사시험에서 상위 10% 정도의 우수한 성적을 기록하면서 뛰어난 성능을 인정받기도 했다.²⁸⁾

26) 본 실험은 서강대학교 대학원 컴퓨터공학과와 서강대학교 ICT 법경제연구소(서강대학교 법학전문대학원)에서 공동으로 수행한 연구의 일부이다.

27) 본 실험 중 실험 1에 대한 설명은 컴퓨터공학적 관점을 중심으로 국내 학회에 발표된 내용(연희연·김민주·정해인·양정·구명완, “생성형 대규모 언어 모델을 활용한 법률분야 리걸 마인드 구현 방법론”, 『한국정보과학회 학술발표논문집』, 2023(6):316-318, 2023)을 토대로 법학적 관점에서 의미 있는 내용을 보완하여 재구성한 것이다. 본 실험은 실험 1의 결과를 재해석하고 OpenAI의 GPT-4를 이용하여 실험 1과 대비되는 실험 2를 추가로 실시한 후 그 결과를 비교하는 것을 내용으로 한다.

28) Daniel Martin Katz et. al., “GPT-4 Passes the Bar Exam”, 2023.



〈그림 3〉²⁹⁾

변호사시험 문제를 활용한 평가는 인공지능의 법률분야 활용에 앞서 언어모델의 법률 지식과 추론 능력을 기초적이나마 확인할 수 있다는 점, 정량적인 점수를 통해 실제 시험 응시자들과의 비교가 가능하다는 점에서 의미가 있다. 때문에 본 논문은 언어모델(본 논문에서는 GPT-3.5와 GPT-4)이 한국 변호사시험, 특히 사례형 문제에 대해서 어느 정도의 성능을 보일 수 있는지, 프롬프팅(Prompting) 방법론을 적용하여 실험하고자 하였다.³⁰⁾

2. 아이디어

언어모델의 성능을 향상시키는 가장 근본적인 방법은 모델의 규모를 키우고, 더 많은 데이터를 학습시키는 것이다. 범용성을 가지는 대규모 언어모델(LLM)을 특정 분야에 적합한 언어모델로 만들기 위해, 해당 도메인에 특화된 추가 학습 데이터를 사용하여 사전학습된 언어모델의 매개 변수를 업데이트하는 미세조정(Fine-tuning) 역시 널리 활용되고 있다. 현재 OpenAI에서는 GPT-4에 대한 전면적 미세조정은 지원하지 않고 있으나, 미세조정을 통해 학습된 GPT-3.5가 특정 도메인에 대해서는 GPT-4보다 더 뛰어난 성능을 보이기도

29) Daniel Martin Katz et. al., “GPT-4 Passes the Bar Exam”, 2023.

30) 유사한 실험을 한 국내 선형연구로는 한국의 법학석사시험 LEET를 사용하여 GPT-4의 추론 및 논증 능력을 평가하기 위한 실험을 한 박성미·박지원·안정민, “법률영역에서 GPT-4 활용 가능성과 시사점-법학석사시험(LEET) 중심으로”, 『경제규제와 법』 제16권 제1호, 2003이 있다. 다만 LEET는 리걸 마인드를 본격적으로 함양하기 위한 법학전문대학원 입학에 필요한 기본적인 추론능력을 측정하기 위한 시험이라는 점에서 이 연구에서 실험 대상으로 한 한국 변호사시험과는 차이가 있다. 이 점에서 한국 변호사시험을 실험 대상으로 이 연구의 차별점이 있다.

한다. 그러나 위와 같은 방법들은 대량의 학습 데이터와 강력한 컴퓨팅 자원을 필요로 하며, 막대한 비용과 시간이 든다는 문제점이 있다.



〈그림 4〉

한편, 추가적인 학습 없이도 언어모델이 최적의 결과물을 생성할 수 있도록 입력(요청 값)을 설계하는 프롬프트 엔지니어링(Prompt Engineering), 즉 프롬프팅의 중요성이 커지고 있다. 프롬프트는 다양한 형태로 주어질 수 있다. 퓨샷(Few-shot) 프롬프팅은 입력 단계에서 몇 가지 예시를 제시하는 방법이며, 연쇄적 사고(Chain-of-thought, CoT) 프롬프팅은 언어모델이 중간추론 단계(intermediate reasoning step)를 거쳐 최종적인 답을 구할 수 있도록 지시하는 프롬프팅 방식이다.³¹⁾ 구체적으로는 문제 풀이에 필요한 논리 절차를 세부적으로 명시하여 순서에 따라 문제를 풀도록 유도하는 방식으로 이루어진다. 규모가 큰 모델일수록, 난이도가 높은 복잡한 문제일수록 CoT 방식의 프롬프팅이 성능 향상에 효과적인 것으로 나타난다.

본 연구에서는 프롬프트 엔지니어링을 통해, 이미 존재하는 언어모델(본 연구에서는 GPT-3.5와 GPT-4)에 최소한의 리소스를 투입하면서도 법률 도메인에 대한 성능이 향상될 수 있는지를 실험하였다. (〈그림 4〉)

3. 법률 지식의 투입 없는 프롬프팅 실험(실험1)

(1) ‘리절 마인드 프롬프트’ 설계

사람이 생각하는 방식과 유사하게 논리의 흐름대로 프롬프팅하는 방법(CoT)에는 여러가지가 있다. 질문과 함께 “Let’s think step by step”과 같은 명령어를 제시하는 방법이나,

31) Jason Wei et. al., “Chain of Thought Prompting Elicits Reasoning in Large Language Models”, arXiv:2201.11903 [cs.CL], 36th Conference on Neural Information Processing Systems, NeurIPS, 2022.

직접 ‘입력-과정-정답’의 쌍을 CoT의 예시로서 제시하는 방법(prompting)은 수학 문제나 일반적인 상식 추론 문제에서 유효한 성능 향상을 보인다. 본 실험에서는 한발 더 나아가 전문 지식을 기반으로 법률 전문가만의 문제 해결 방식(앞서 정의한 ‘리걸 마인드’)을 따라 최종적인 답을 얻어야 하는 법률 분야에 적합한 프롬프팅 방식을 설계하고자 하였다.



〈그림 5〉³²⁾

우선 법률 전문가의 문제해결 단계(‘리걸 마인드’)를 정의하였다. 변호사시험의 사례형 문제와 같이 ‘법률 사례’를 풀이하기 위해 ‘사례로부터 사실관계 추출’, ‘추출한 사실관계로부터 쟁점 도출’, ‘쟁점에 대한 결론 도출’이라는 3단계의 해결방식을 따른다고 가정하였다.

〈표 1〉

법률추론 단계	프롬프트 방식	
	프롬프트	Output
사실관계추출	사례	사실관계
쟁점도출	사례+사실관계 질문	쟁점
결론	사례+사실관계+쟁점 질문	결론

그리고 각 단계에 따라 결과를 누적하는 방식으로 프롬프트를 구성하여, 전문가의 사고 과정을 따라가면서도 각 단계별 성능을 확인할 수 있도록 하였다. 사실관계추출 단계에서는 사례와 질문이 주어지고, 주요 사실관계를 출력한다. 쟁점도출 단계에서는 사례와 주요

32) 연희연 외, 앞의 논문.

사실관계 및 질문이 주어지고, 관련 쟁점을 출력한다. 결론도출 단계에서는 사례, 주요 사실관계, 쟁점이 모두 주어지고 질문에 따라 결론을 출력한다(〈표 1〉). 위와 같은 프롬프팅 방식은 ‘Multi-stage 방식의 CoT’ 또는 ‘Chain of Subtasks’에 해당한다고 할 수 있다.³³⁾

(2) 실험 데이터 구축

실험을 위해 변호사시험 데이터를 정리하였다. 대한민국 법무부에서 공개한 변호사시험 문제 중 민사법 사례형 문제 94개를 수집한 후, 모범 답변을 작성하였다. 모범 답변은 서강대학교 법학전문대학원 석사과정 학생들이 교과서와 시중에 나와 있는 변호사시험 사례형 기출 해설문제집을 참고하여, 위에서 정의한 법률사례 해결의 3단계(사실관계 추출, 쟁점 도출, 결론)에 맞게 작성하였다(〈표 2〉).

〈표 2〉

변호사시험 사례형 문제
<p>A는 국가 소유인 312번지 토지를 점유하고 있었고, B도 역시 국가 소유인 313번지 토지를 점유하면서 그 지상에 건물을 건축하여 이를 사용하여 왔는데, 312번지 토지와 313번지 토지는 서로 인접하여 있다. 그런데 A는 국가와 312번지 토지를 매수하기로 협의하였는데, 매매계약을 체결하는 과정에서 매매목적물을 313 번지 토지의 지번으로 기재하였고, 이에 기하여 313번지 토지에 관하여 A명의로 소유권이전등기가 경료되었다. 그 후에도 A는 312번지 토지를 계속 점유, 사용하여 왔는데, 313번지 토지를 점유하던 B가 313번지 토지를 국가로부터 매수하는 과정에서 위 사실이 밝혀졌고, A도 역시 이때에 그 사실을 알게 되었다.</p> <p>위 경우에 312번지 토지에 대해서 매매계약이 성립하는가?</p>
주요 사실 관계
<ol style="list-style-type: none"> 1. 312번지, 313번지 토지는 국가소유이다. 2. A는 312번지, B는 313 번지 토지를 점유하고 있다. 3. A는 312번지 토지를 국가로부터 매수하였다. 4. A와 국가는 착오로 313번지 토지를 계약서 상의 매매목적물로 기재하였다. 5. 313번지 토지에 관하여 A 명의로 소유권이전등기가 경료되었다.
쟁점
<p>법률행위의 해석 방법. 구체적으로는 매매계약서 내용 작성 과정에서 당사자의 착오가 있을 때, 매매목적물의 특정방법</p>
결론

33) Takeshi Kojima et. al., “Large Language Models are Zero-Shot Reasoners”, arXiv: 2205.11916 [cs.CL], 2022.

판례는 부동산의 매매계약에 있어 쌍방당사자가 모두 312번지 토지를 계약의 목적물로 삼았으나 그 목적물의 지번 등에 관하여 착오를 일으켜, 계약서상 그 목적물을 312번지 토지와는 별개인 313번지 토지로 표시하였다 하여도 312번지 토지에 관하여 이를 매매목적물로 한다는 쌍방당사자의 의사합치가 있는 이상, 위 매매계약은 312번지 토지에 관하여 성립한 것으로 보아야 할 것이라고 판시하여 자연적 해석방법을 인정하고 있다.

판례의 자연적 해석방법을 기초로 위 사실관계를 검토하면, 매매계약서상 기재와 등기 경로(313번 토지)에도 불구하고 매매계약의 쌍방당사자인 A와 국가간 의사의 합치가 이루어진 312번 토지를 매매 목적물로 삼아 매매계약을 체결한 것이 분명하다.

따라서, A와 국가 간의 매매계약은 312번 토지에 대해 성립한다.

(3) 사용 모델

Open AI의 GPT-3.5-turbo 모델에 ‘리걸 마인드 프롬프트 방식’을 사용하여 답변을 생성하게 하였다. GPT-3.5-turbo는 트랜스포머 디코더 기반의 GPT 모델로 입력 가능한 토큰 수는 4,096개이다. 파라미터 개수는 대략 20B로 알려져 있다. 성능 비교를 위한 지도학습 모델로는 파라미터 수 기준 GPT의 1/100 정도 크기(220M)의 한국어 T5³⁴⁾-base를 사용하였고, 앞서 수집한 변호사시험 기출 문제 중 테스트에 사용할 19개를 제외한 나머지 문제와 모범답안으로 추가학습(fine-tuning)시켰다.

(4) 평가 방식

자동평가와 전문가 평가를 합하여 3가지 방법으로 평가를 진행하였다. 자동평가는 모범답안과 출력값을 비교하여 토큰을 기준으로 유사도 점수를 산출하였다($F1\ score = \frac{2 \times precision \times recall}{precision + recall}$).³⁵⁾ 전문가 평가에는 법무법인 디라이트의 변호사, 서강대학교 법학전문대학원 석사과정 2학년 재학생들이 참여하였다. 인간 정확성(Human precision, 모델이 생성한 문장에서 올바른 요소의 비율을 나타낸 점수), 인간 선호도(Human preference, 모델의 답변이 형식, 내용, 유창성의 측면에서 전문가다운지를 나타낸 점수)를 각각 5점 척도로 평가하였고,³⁶⁾ 0에서 1 사이인 F1 score와 스케일을 통일하기 위해 1/5배로 변환하였다.

34) Colin Raffel et al., “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer”, arXiv:1910.10683 [cs.LG], 2019.

35) Precision : 결과값 중 모범답안에 포함된 비율, Recall : 모범답안 중 결과값에 포함된 비율

36) 1점(전체적으로 보아 전혀 전문가답지 않은 답변으로 느껴지는 경우), 2점~4점(형식, 내용, 유창성 측면에서 전문가답게 느껴지는 요소의 개수에 따라 점수 부여), 5점(모든 요소가 갖춰지고, 전체적으로 보아 실제 전문가의 답변과 구분할 수 없을 정도로 느껴지는 경우)

(5) 실험 결과

우선 단계적 프롬프팅 없이 [사례]를 입력하고 곧바로 [결론]을 출력하도록 한 실험에서 0점(사실상 평가 불가능)에 수렴하는 평가 점수를 기록한 데 반해, 프롬프팅을 통해 단계를 나눈 결과 어느 정도 유의미한 답변을 얻어낼 수 있었다.

〈표 3〉

법률추론 단계	프롬프트 방식		F1 score	Human Precision	Human Preference
	프롬프트	Output			
사실관계추출	사례	사실관계	0.575	0.893	0.770
쟁점도출	사례+사실관계 질문	쟁점	0.288	0.547	0.510
결론	사례+사실관계+쟁점 질문	결론	0.261	0.251	0.345

‘리결 마인드 프롬프트’ 실험에 대한 결과를 3가지 방식으로 평가하였다(〈표 3〉). 각 행은 앞서 정의한 3단계의 법률추론 단계를 나타낸다. 1행을 참고하면 법률 사례가 주어졌을 때, 주요 사실관계를 추출하는 과제에서 가장 좋은 성능을 보인 것을 알 수 있다. 이는 GPT가 일반적으로 강점을 보이는 자연어 요약과 유사한 과제이기 때문으로 분석된다. 그에 반해 2행의 [쟁점도출 단계]에서는 전문가 평가에서 0(매우 나쁨)과 1(매우 좋음)의 중간 정도의 점수를 기록하였고, 3행의 [결론도출 단계]에서는 가장 낮은 점수를 얻은 것을 확인할 수 있었다. [쟁점도출 단계]와 [결론도출 단계]는 학습된 법률 지식을 기반으로 사례에 적절한 조문과 판례를 적용하여 풀어내야 하는 과제이기 때문인 것으로 분석된다.

〈표 4〉³⁷⁾

쟁점도출 프롬프트 예시
<p>[사례]</p> <p>A는 국가 소유인 312번지 토지를 점유하고 있었고, B도 역시 국가 소유인 313번지 토지를 점유하면서</p>

37) 〈표 4〉는 〈표 3〉에서 쟁점도출 단계 실험에 사용된 프롬프트와 GPT의 답변 예시이다. [사례]와 함께 모범답안의 [주요사실관계]를 제시하고, 질문에 대한 답변에 앞서 어떠한 법률적 쟁점이 검토되어야 하는지 작성할 것을 주문하였다.

그 지상에 건물을 건축하여 이를 사용하여 왔는데, 312번지 토지와 313번지 토지는 서로 인접하여 있다. 그런데 A는 국가와 312번지 토지를 매수하기로 협의하였는데, 매매계약을 체결하는 과정에서 매매 목적물을 313번지 토지의 지번으로 기재하였고, 이에 기하여 313번지 토지에 관하여 A명의로 소유권이 전등기가 경료되었다. 그 후에도 A는 312번지 토지를 계속 점유, 사용하여 왔는데, 313번지 토지를 점유하던 B가 313번지 토지를 국가로부터 매수하는 과정에서 위 사실이 밝혀졌고, A도 역시 이때에 그 사실을 알게 되었다.

[주요 사실 관계]

1. 312번지, 313번지 토지는 국가소유이다.
2. A는 312번지, B는 313 번지 토지를 점유하고 있다.
3. A는 312번지 토지를 국가로부터 매수하였다.
4. A와 국가는 착오로 313번지 토지를 계약서 상의 매매목적물로 기재하였다.
5. 313번지 토지에 관하여 A 명의로 소유권이전등기가 경료되었다.

[질문] 위 경우에 312번지 토지에 대해서 매매계약이 성립하는가?

이에 대한 답변을 하기위해 검토해야할 법률적 쟁점이 무엇인지 간단히 작성하시오.

GPT의 쟁점도출 답변 예시

[답변] 매매계약이 성립하지 않는다

[쟁점]

매매목적물 기재 착오에 따른 효력 문제 : ... 내용생략 ...

A가 312번지 토지의 적법한 소유자인지 여부 : ... 내용생략 ...

B의 보호를 위한 제3자 보호의 원칙 : ... 내용생략 ...

〈표 5〉

사용모델 (파라미터 개수)	F1 score	Human Precision	Human Preference
GPT (20B)	0.575	0.893	0.770
T5-base (0.22B)	0.558	0.552	0.295

추가적으로 T5-base 지도학습 모델과의 [사실관계추출] 단계 성능을 비교하였다(〈표 5〉).³⁸⁾ 토큰의 유사도를 기준으로 한 2열의 F1 score는 ‘리걸 마인드 프롬프트’를 적용한 GPT3.5-turbo 모델과 ‘변호사시험 기출문제로 지도학습’을 진행한 T5-base 모델에서 점수 차이가 크지 않았다. 그러나 3, 4열의 전문가 평가를 참고하면 그 차이가 두드러지는 것을 확인할 수 있었다. 이는 T5-base 모델이 입력된 법률 사례의 문장들을 반복적으로 작성하

38) T5-base 모델의 쟁점 도출, 결론 도출 실험의 결과가 0점(사실상 평가 불가능)에 수렴하여, 사실관계추출 단계의 성능만을 비교하였다.

였을 뿐 주요 사실관계추출, 답변의 형식과 유창성 측면에서 법률 전문가의 답변이라고 보기에는 부족하였기 때문이다.

4. 법률 지식을 투입한 프롬프팅 실험(실험2)

(1) 사용 모델

OpenAI의 GPT-4 모델을 사용하였다. GPT-4는 32,768토큰을 입력받을 수 있고, 이는 GPT-3.5에 비해 8배 증가된 것이다. GPT-4는 1T 이상의 파라미터를 갖는 것으로 알려졌으나, 정확한 개수는 공개되지 않았다. GPT-3.5-turbo는 법률 지식을 투입한 프롬프팅을 적용하더라도 유의미한 결과를 출력해내지 못했다.

(2) 실험 방법

한국 변호사 시험의 사례 문제를 제시할 때에, 관련된 법조문 또는 관련 판례를 함께 프롬프팅한 경우와 프롬프팅하지 않은 경우의 답변을 모범답안과 비교하였다.

(3) 실험 결과

〈표 6〉

변호사시험 사례형 문제
<p>A는 2016. 1. 1. 시가 약 5억 원 상당의 그 소유의 X토지를 B에게 매도하기로 하면서 계약금 5천만 원은 계약 당일, 중도금 2억 원은 2006. 3. 1. 잔금 2억 5천만 원은 2006. 5. 1.에 각 지급받되, 중도금을 지급받음과 동시에 위 부동산을 인도하여 주고, 잔금을 지급받음과 동시에 위 부동산에 설정된 근저당권설정등기 및 압류등기 등을 모두 말소한 아무런 부담이 없는 완전한 상태에서의 소유권이전 등기를 경료하여 주기로 약정하였다.</p> <p>이에 B는 A에게 계약금 5천만 원과 중도금 2억 원을 각 지급기일에 모두 지급하였고 A는 중도금을 지급받음과 동시에 위 부동산을 B에게 인도하였다. 그런데 위 부동산에 설정된 근저당권설정등기는 채권최고액이 6억 원으로 이미 부동산의 가액을 넘어선 상태였고, 청구금액을 3억 원으로 하는 압류등기 또 한 경료되어 있었으며 A는 위 등기들을 말소시킬 만한 자력이 없는 상황이었다. A가 위 근저당권설정등기 및 압류등기를 말소시키지 못한 상황에서 잔금지급기일이 도래하자 B는 잔금을 지급하거나 그 이행을 제공하지 않은 채 A에게 위 등기들의 말소를 독촉하였고, A가 자력이 없어 위 등기들을 말소시키는 것이 어렵다고 판단되자 2016. 6. 10. 위 매매계약을 해제하였다. (여전히 B의 잔금 이행 및 이행의 제공은 없었다)</p> <p>질문 : B는 A의 이행지체를 이유로 계약을 해제할 수 있는가?</p>

프롬프트

“쌍무계약에서 쌍방의 채무가 동시이행관계에 있는 경우 일방의 채무의 이행기가 도래하더라도 상대방의 이행제공이 있을 때까지는 그 채무를 이행하지 않아도 이행지체의 책임을 지지 않는 것이다.”

변호사시험 민사법 사례형 문제(사례+질문)를 입력할 때, 프롬프트에 관련된 법조문 또는 판례를 입력해주었다. (〈표 6〉)을 참고하면 사례 해결을 위한 법적 지식으로 ‘쌍무계약에서의 동시이행관계’라는 개념을 참고할 수 있도록, 프롬프트로 관련 판례의 일부를 입력하였고, 그에 따라 GPT-4의 답변이 달라진 것을 확인할 수 있다.

〈표 7〉

모범답안

2016. 5. 1. 이행기가 도래하였음에도 A는 소유권이전등기의무를 지체하였으며, 이에 특별한 사정이 없는 한 귀책사유도 있다고 할 것이다. 다만 A는 동시이행항변권을 갖고, B의 이행 또는 이행제공이 없는 한, A가 이행기에 이행을 하지 않더라도 위법하지 않다. 따라서 A는 이행지체 책임을 지지 않고, B는 A의 이행지체를 이유로 매매계약을 해제할 수 없다.

프롬프트 입력 없이 GPT4가 생성한 답변

A와 B 사이에 매매계약에는 몇 가지 중요한 항목이 있습니다. 그 중 하나는 A가 잔금을 지급받는 동시에 부동산에 설정된 근저당권설정등기 및 압류등기 등을 모두 말소하고, 완전한 소유권이전등기를 경료해야 한다는 조건입니다.

그러나 A는 이러한 의무를 이행하지 못하였습니다. 따라서 이 경우, A는 이행지체 상태에 놓여있다고 볼 수 있습니다. 즉 A는 계약의 의무를 이행하지 않았으므로, 이행지체에 따른 책임을 지게 될 수 있습니다.

프롬프트 입력 받은 GPT4가 생성한 답변

당신이 제공한 대법원 판례에 따르면, 쌍무계약에서 양측의 의무가 동시에 이행되어야 하는 관계에 있을 때, 한 쪽의 의무이행 기한이 도래하더라도 상대방의 의무 이행 제공이 있을 때까지 그 의무를 이행하지 않아도 이행지체의 책임을 지지 않는다고 합니다.

이 사안에서 A와 B 사이의 계약은 쌍무계약으로 볼 수 있습니다. 계약에 따르면 B는 잔금을 지불하고 A는 잔금지급과 동시에 부동산에 설정된 근저당권설정등기 및 압류등기 등을 모두 말소하고 완전한 소유권이전등기를 경료해야 합니다. 이는 둘 다 상대방의 이행제공이 있을 때까지 자신의 의무를 이행하지 않을 수 있음을 의미합니다.

따라서 이 판례를 바탕으로 볼 때, A는 잔금의 지급이 이루어지지 않았기 때문에, A는 등기의 말소를 이행하지 않아도 이행지체 책임을 지지 않을 수 있습니다. A가 등기를 말소하지 못하였지만, 이는 B가 잔금을 지급하지 않은 상황에서 발생한 것이므로, 이 경우 A는 이행지체 책임을 지지 않는 것으로 보입니다.

[모범답안]에 따르면 A와 B의 채무는 동시이행관계에 있고, 따라서 B의 이행제공이 없는 한, A 역시 이행지체책임을 지지 않는다는 결론을 내려야 한다. [프롬프트 입력 없이 GPT4가 생성한 답변]에 따르면 동시이행관계를 고려하지 않은 채 이행지체책임을 지는 것으로 결론을 내리고 있다. 이는 법률 용어로서의 '이행지체'를 단순한 일반 상식상 '이행이 늦어진 것'으로 파악하고, 의무를 이행하지 않았다는 점에 중점을 둔 답변으로 분석된다. 그러나 [프롬프트를 입력받은 GPT4가 생성한 답변]에 따르면, 주어진 판례를 기반으로 사례에 주어진 계약이 쌍무계약임을 도출하고, B의 잔금지급이 이루어지지 않은 점을 정확히 지적하며 동시이행관계에 따라 A는 이행지체 책임을 지지 않는다는 결론을 내리고 있다.(<표 7>)

5. 실험 결론

본 실험은 3년 이상 훈련받은 법학전문대학원 졸업생이 응시하는 '변호사 시험'을 기준으로 진행되었으며, 그 평가 역시 숙련된 법률전문가를 통해 이루어져 절대적인 평가점수는 부족한 측면이 있다. 그러나 대규모 언어모델인 GPT-3.5와 GPT-4의 사실관계 요약능력은 법률 분야에서도 뛰어난 성능을 보였다. 그리고 (실험1) T5-base 모델과의 비교를 통해서 '대규모 언어모델에 적절한 프롬프팅을 적용하는 것'이 '작은 규모의 언어모델을 미세조정하는 것'보다 오히려 나은 성능을 보일 수도 있음을 확인하였다.³⁹⁾ 다만, 사실관계추출 이후의 쟁점 및 결론 도출 단계로 갈수록 정확도가 낮아지며, 법률적 전문성을 필요로 하는 추론 단계에서는 여전히 부족함을 보이고 있다. 사실관계 속에 쟁점이 명시적으로 드러나지 않은 경우에는 논점을 일탈하기 쉬웠고, 논점을 일탈하지 않았더라도 명확한 결론을 내놓지 못하는 경우가 많았다. 영어 기반의 GPT에 한국어 기반의 법률, 판례정보가 부족하기 때문으로 분석된다. 이는 향후 현실에서의 활용을 위해 질문의 명확성을 높이는 것이 중요하다는 방증이며, 일반인에 비해 훈련된 법률 전문가가 법률 인공지능을 더욱 잘 활용할 수 있다는 의미이다.

추가적으로 법률분야는 일반적인 언어 체계와 달리 '법률용어'의 엄밀한 활용이 요구되므로, 전문용어와 일반용어의 혼용 문제, 존재하지 않는 법률이나 판례를 인용하는 환각

39) 한국어 T5-base 모델에 추가로 학습(fine-tuning)시킨 데이터가 변호사시험 문제와 모범답안 70여개에 불과해 매우 적었던 것 역시 T5-base 모델의 성능 저하의 한가지 요인이다. 이는 특정 도메인에 대한 데이터를 학습시키는 과정에서 마주치는 일반적인 문제이며, 특히 법률분야는 개인정보 문제와 결부되어 있어, 양질의 학습데이터를 확보하는 데 어려움이 있다. 법률데이터 확보 문제는 향후 리걸테크의 개발 과정에서 기술 외적으로 해결해야 할 가장 중요한 부분으로 생각된다.

문제(Hallucination) 등이 타 분야에 비해 답변의 신뢰수준을 심각하게 저해할 수 있다. 예를 들면, 법률용어로서 ‘소유’란 일상생활에서의 포괄적인 개념(점유나 보유 등)과는 법적 의미가 다르기 때문에, 그 사용에 신중을 기하여야 한다. 그러나 생성된 답변은 이를 혼용하여 사용하고 있고, 법률전문가 입장에서는 ‘답변이 전문가다운지’를 평가할 때 큰 감점 요소로 삼을 수밖에 없었다. 이는 전문가 평가 방식의 Human preference 점수를 통해 드러났다. 정답을 도출하는 것만큼이나, 답변에서 확실한 오류를 제거하는 것 역시 중요한 과제임을 알 수 있다.

반면 (실험2)에서는 적절한 법률지식(법조문 또는 판례)만 제시된다면, 이를 사례에 적용하고 결론을 도출하는 GPT의 법적추론능력 및 사안 포섭능력이 충분히 발휘되는 것을 확인할 수 있었다. 장차 법률분야에 대한 언어모델의 활용성을 기대해볼 수 있는 결과이다.

대규모 언어모델을 특정 분야에 적용하기 위해 직접 사전훈련(Pre-train)하거나, 추가학습(fine-tuning)하는 데에는 충분한 학습데이터 수집, 강력한 컴퓨팅 리소스 확보, 비용 등 현실적인 문제가 존재한다. 특히 법률분야 데이터는 개인정보 문제와 연결되어 수집에 많은 어려움이 있다. 그러나 프롬프팅 방법론을 통해서 기술적으로도 대규모 언어모델의 법률 답변 성능을 어느 정도 향상시킬 수 있음을 확인하였다. (실험1)과 같이 ‘리걸 마인드 프롬프팅’을 통해 단계적 추론 방식으로 정확도를 높이고, 동시에 (실험2)와 같이 관련된 한국의 민법 규정과 대법원 판례를 프롬프팅함으로써 새로운 학습과정 없이도 점진적으로 리걸 마인드를 형성시킬 수 있다.

IV. 법률 인공지능의 발전 전망

전세계의 법률 전문가 및 리걸테크 관련 기업들은 법률 업무에 생성형 인공지능(Generative AI)을 도입하는 것은 불가피한 추세라고 말하고 있다. 미국의 한 시장조사업체 (MarketResearch.Biz)는 2032년까지 생성형 인공지능과 관련된 법률 시장 규모가 약 675.1억 달러를 넘어설 것으로 예상하였고, 2023년부터 2032년까지의 연평균 시장성장률이 30.7%에 달할 것으로 전망하였다.

사회적 사실관계에서 법적 쟁점과 결론을 도출할 수 있는 자연어 인공지능 모델인 리걸 마인드를 비롯한 법률 인공지능은 지능형 비서로서 법조업무 효율화에 기여하고 이를 통해 법률 시장의 사회적 비용을 감축하며 일반 국민의 사법접근성을 제고할 수 있을 것이다.

1. 변호사의 업무효율성 증대

(1) 법령 및 판례 검색

고도화된 법률 인공지능은 자연어 처리 기술을 사용하여 기존의 검색 시스템에 비하여 복잡한 법률 용어와 개념을 더 잘 이해하고, 관련 정보를 빠르게 탐색할 수 있다. 예컨대, 변호사가 “계약의 해제”에 대한 법령이나 판례를 검색하면, 법률 인공지능은 “계약의 해제”라는 문구뿐만 아니라 “계약의 해지”, “계약의 파기”와 같이 유사한 문구를 포함하는 법문서까지도 함께 검색하고 그 결과를 검색어와의 관련성 순으로 정렬할 수 있으므로 변호사는 보다 빠르고 정확하게 원하는 정보를 찾아 그 업무를 수행할 수 있게 된다.

실제로 톰슨 로이터(Thomson Reuters)에서 제공하는 법률검색시스템 중 Westlaw Edge, Westlaw Precision은 인공지능을 기반으로 하여 기존의 검색시스템에서의 검색으로 놓친 문헌이나 법령을 알려주는 서비스를 제공하고 있다.

또한, 최근에는 텍스트 생성을 넘어 이미지의 입력·처리·추론이 가능한 대규모 언어 모델이 출시되면서 멀티모달(Multi-Modal) 인공지능이 점차 가시화되고 있다. 법률 인공지능이 이러한 방향으로 발전된다면, 변호사와 고객 사이 음성 데이터를 기반으로 유관 법률 정보들을 검색해주는 서비스나 사진 및 동영상 등 멀티미디어 형식의 증거물 등을 업로드 하면 텍스트 환경에서와 같이 법적 쟁점을 도출하고 관련 정보들을 도출하여 제공하는 서비스 또한 가능할 것이다.

인공지능을 이용한 대량의 계약서 검토가 필요한 법률실사, 반복적이고 정형적인 계약서 검토 등의 업무는 이미 상용화되고 있으나, 변호사의 계약서 검토 작업을 직접적으로 보조하여 계약서에 필수적으로 기재되어야 하는 조항이 포함되어 있는지 여부를 판단하거나 기존 계약서와 내용 측면의 비교를 하는 등의 기능은 자연어 처리 모델을 탑재한 인공지능의 개발에 따라 가까운 미래에 적극 활용될 것으로 예상된다.⁴⁰⁾

나아가, 본 연구가 목표로 하는 ‘리걸 마인드를 탑재한 법률 인공지능’은 주어진 상황에서 법적 문제를 포착하고 이를 기존 법제에서 해결하는 방안을 도출할 능력을 가지므로, 특정한 계약서 조항의 적법성 여부나 당사자애의 유불리 여부 등을 판단하거나 질문자가 입력하는 정보를 바탕으로 그가 목표하는 것을 달성하기 위하여 필요한 법문서의 초안을

40) 강석구 외 8, 『리걸테크 도입 및 대응을 위한 법무정책 연구』, 22-B-16, 한국형사법무정책연구원, 2022, 52-53면.

작성하는 것 또한 가능할 것이다. 이에 관하여 변호사법 위반의 여지가 있다는 우려 또한 존재하는데, 이러한 법률 인공지능의 활용 이전에 이해관계자들 간 담론을 형성하고 법령을 정비하는 것이 필요하다.

2. 사법 지원(Judicial Support)

리걸 마인드를 탑재한 법률 인공지능은 법관의 지능형 비서로서 기능하며 사법적 의사결정을 보조하고 지원할 수 있을 것으로 기대된다. 그 양상은 법률 인공지능이 자연어 처리 및 빅데이터 처리 능력을 기반으로 법관이 필요로 하는 법령 및 유사 판례 자료를 수집, 정리한 후 이를 분석한 결과를 제공하는 것, 사실관계를 확정하고 이에 대해 판단하는 것, 소송당사자들이 제출한 문서들 중 법적으로 의미 있는 부분들을 일차적으로 식별하여 표시하고 이를 요약한 자료를 제공하는 것 등으로 나타날 수 있다. 뿐만 아니라, 대규모 언어모델을 기반으로 하는 만큼 법률 인공지능이 법관이 작성할 문서의 초안을 작성하는 것 또한 가능할 것이다.

이렇듯 사법적 의사결정의 과정에서 법률 인공지능을 지능형 비서로 적극적으로 활용한다면, 시간소모가 크거나 반복적인 업무는 인공지능이 수행하게 되고 법관은 보다 복잡하고 긴급한 사건들, 인간인 법관의 논리적인 검토나 판단이 필요한 영역들에 더욱 집중할 수 있게 되고, 이는 사법부의 재판 지연 문제의 해결 및 사법부에 대한 신뢰 제고에 기여할 것이다.

리걸 마인드를 탑재한 법률 인공지능은 전형적인 법적 쟁점에 대한 문제를 해결할 수 있는 능력까지도 갖출 것으로 예상되는데, 단순한 사법 지원을 통한 업무효율화를 넘어 법률 인공지능의 판단에 의한 판결이나 결정이 일부 가능할 것이라는 전망 또한 존재한다. 이는 특히 양형 계산이나 보석 심사 등 객관적, 정량적 분석을 주요한 근거로 삼는 영역들에 적용될 수 있다고 여겨진다. 하지만 법률 인공지능을 이에 활용할 것인지는 인공지능과 차별의 문제, 헌법 제27조에 기한 법관에 의한 재판을 받을 권리 등 윤리적, 법정정책적 문제가 함께 논의되어야 한다.

3. 온라인 분쟁해결(ODR)의 발전을 통한 사법접근권 제고

재판 외 분쟁해결(Alternative Dispute Resolution, “ADR”)이란 소송을 갈음하는 분쟁해

결제도로서 화해, 조정, 중재 등이 있다. 소송에 의한 분쟁해결은 상당한 시간과 비용이 소요되고, 소송 당사자 간의 분쟁을 법적으로 구조화된 틀에서 해결하기 때문에 당사자들의 구체적인 이해관계와 의사를 반영하기 어렵다는 한계를 가지므로, ADR이 대안으로 주목받고 있다. 또한, 코로나 바이러스로 인해 법원이 폐쇄되고 재판이 연기되는 상황에 직면하면서 각국은 ADR의 과정 일부 또는 전부를 온라인상에서 수행하는 온라인 분쟁해결(Online Dispute Resolution, “ODR”)⁴¹⁾을 확대하려고 서두르고 있다.⁴²⁾

제1세대 ODR 시스템에서는 인간이 계획과 결정 과정에서 주도적인 역할을 하고 컴퓨터는 도구에 불과했지만, 제2세대 ODR 시스템에서 컴퓨터는 정보 접근을 용이하게 하는 도구에서 나아가 계획·전략 수립·결정이 가능한 단계에 이르렀고⁴³⁾, 인공지능의 기술은 ODR의 발전을 이끄는 중요한 요소라고 할 수 있다.⁴⁴⁾ 분쟁 당사자들이 인공지능을 활용함으로써 계약 검토 프로세스 가속화, 중재판정의 결과 예측, 합의를 위한 제언 생성 기능 등을 통해 ODR을 보다 신속하게 진행할 수 있기 때문이다. 이는 소송경제에 기여한다는 ADR 또는 ODR의 취지를 달성하고 국민의 사법접근권의 실질화에 이바지할 것이다. 이렇듯 인공지능이 ODR의 발전과 활용성에 미치는 영향은 특히 소액사건에서 크게 나타날 것으로 기대된다.

예컨대, 캐나다 브리티시 컬럼비아(British Columbia) 주의 민사해결법정(Civil Resolution Tribunal, “CRT”)은 ODR 메커니즘을 도입한 최초의 캐나다 법정으로, 소송목적의 값이 5,000캐나다 달러(약 490만 원) 미만인 소액사건 및 공동주택 관련 분쟁 등을 관할로 한다.⁴⁵⁾ CRT가 활용하는 인공지능 도구 중 하나는 조정의 초반부에서 이용되는 “the Solution Explorer”로 인공지능 기반의 전문가 시스템이 법률 상담을 진행하여 분쟁 당사자들이 CRT에 조정을 신청하는 대신 취할 수 있는 대체적인 해결 방안들을 안내한다. 또한, 분쟁 당사자의 입찰 전략과 선호의 우선 순위를 학습하여 금전적 협상안을 도출하는 것을 지원하는 알고리즘을 사용한 “Smartsettle ONE”이라는 ODR 도구가 활용된다. CRT가 공개한 자료에 따르면, 2023년 11월 한 달간 CRT는 55건의 주택 소유권 분쟁, 327건의 소액사건, 49건의 자동차 상해/사고보상/사고 책임 분쟁, 5건의 기타 사건을 종결하였다.⁴⁶⁾

41) 손승우, 『온라인 분쟁해결(ODR)에 관한 국제 규범 모델 연구(1) -ODR에 관한 국제 규범 동향 분석』, 비교법제연구, 한국법제연구원, 2011, 23면.

42) 강영주, “인공지능을 활용한 민사분쟁 해결에 관한 고찰”, 『안암법학』, 제64권, 2022.5., 401면.

43) Davide Cameiroa et. al., “Online Dispute Resolution : an artificial intelligence perspective”, *Artificial Intelligence Review*, vol.41, 2014, pp.214-215.

44) 강영주, 앞의 논문, 404면.

45) Civil Resolution Tribunal. Available at: <https://civilresolutionbc.ca>

또한, 2023년 11월 집계된, CRT에 참가한 분쟁 당사자들의 만족도 조사 결과에 따르면, 84%의 참여자가 다른 사람에게 CRT를 통한 분쟁 해결을 추천한다고 응답했다.⁴⁷⁾

에스토니아 또한 소액사건에서 인공지능을 이미 활용하고 있는데, 7,000유로(약 950만원) 미만의 소액사건은 인간으로부터 완전히 독립적인 인공지능 관련 시스템이 결정을 내릴 수 있도록 하고 있다. 소송 당사자들이 업로드한 각종 정보를 ODR이 분석해 판단에 이르는데, 인공지능 시스템에 의한 판단은 법적 구속력을 가지게 되고, 이에 관련된 항소만을 인간 법관이 담당한다.⁴⁸⁾ 2023년 9월에는 일본 정부 또한 ADR을 온라인화하기 위한 실증사업을 시작하였다.

V. 결론

기술의 진보로 인해 자연어 인공지능은 이전보다 더 넓은 분야에서, 더 다양한 문서들을, 더 낮은 비용으로 학습할 수 있게 되었기에, 자연어 인공지능 모델은 법률 시장에 커다란 변화를 가져올 것으로 예상된다. 자연어 인공지능의 발전사를 살펴보면, 초기 순환신경망 모델과 최근 OpenAI의 ChatGPT의 성능 정도를 체감할 수 있다.

인공지능은 막대한 양의 법률 문서를 신속하게 요약하고 분석할 수 있을 뿐만 아니라, 유사한 판례들을 묶어서 정리하는 데에도 도움을 줄 수 있다. 이러한 인공지능의 능력은 변호사들이 반복적이고 시간 소모적인 작업을 처리하는 데 도움을 줄 수 있으므로, 법조인들이 보다 가치 있는 작업에 집중할 수 있도록 도와줄 것으로 기대된다.

특히 본 논문에서 실험한 것처럼 법적 추론능력을 가진, 즉 리걸 마인드가 탑재된 인공지능 모델이 상용화된다면, 법률 인공지능의 사용 범위와 활용도가 크게 증가할 것으로 예상된다.

▶ 논문투고일 : 2023.11.29, 논문심사일 : 2023.12.11, 게재확정일 : 2023.12.27.

46) CRT Key Statistics, November, 2023. Available at:

<https://civilresolutionbc.ca/blog/crt-key-statistics-november-2023/>

47) Participant Satisfaction Survey, November, 2023. Available at:

<https://civilresolutionbc.ca/blog/participant-satisfaction-survey-november-2023/>

48) 정채연, 『사법절차 및 사법서비스에서 인공지능 기술의 도입 및 수용을 위한 정책 연구』, 사법정책연구원 연구총서 2021-10, 2021, 71면.

■ 참고문헌

〈국내문헌〉

- 강석구 외 8, 『리걸테크 도입 및 대응을 위한 법무정책 연구』, 22-B-16, 한국형사법무정책연구회, 2022.
- 김재인, 『인공지능의 시대, 인간을 다시 묻다』, 동아사이, 2018.
- 손승우, “온라인 분쟁해결[ODR]에 관한 국제 규범 모델 연구[1] -ODR에 관한 국제 규범 동향 분석”, 『비교법제연구』 한국법제연구원11-16-1-1, 2011
- 스튜어트 러셀, 피터 노빅 지, 류광 역, 『인공지능: 현대적 접근방식(제3판) 제1권』, 제이펍, 2016.
- 정채연, 『사법절차 및 사법서비스에서 인공지능 기술의 도입 및 수용을 위한 정책 연구』, 사법정책연구원 연구총서 2021-10, 2021.
- Ashley, Kevin D., 오태운·정영수·조동관 역, 『법을 분석하는 인공지능』, 박영사, 2020.
- Schauer, Frederick, 김건우 역, 『법률가처럼 사고하는 법: 법적 추론 입문』, 도서출판 길, 2019.
- 강영주, “인공지능을 활용한 민사분쟁 해결에 관한 고찰”, 『안암법학』, 제64권, 2022.5.
- 박성미·박지원·안정민, “법률영역에서 GPT-4 활용 가능성과 시사점-법학적성시험(LEET) 중심으로”, 『경제규제와 법』 제16권 제1호, 2003.
- 연희연·김민주·정해인·양정·구명완, “생성형 대규모 언어 모델을 활용한 법률분야 리걸 마인드 구현 방법론”, 『한국정보과학회 학술발표논문집』, 2023(6):316-318, 2023.

〈해외문헌〉

- Ashley, Kevin D., Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age, Cambridge, 2017.
- Brozek, Bartosz, The Legal Mind: A New Introduction to Legal Epistemology, Cambridge University Press, 2020.
- Schauer, Frederick, Thinking Like a Lawyer: A New Introduction to Legal Reasoning, Harvard University Press, 2009.
- Althammer, Sophia, Sebastian Hofstätter, Mete Sertkan, Suzan Verberne, Allan Hanbury, “PARM:- A Paragraph Aggregation Retrieval Model for Dense Document-to-Document Retrieval”, arXiv: 2201.01614 [cs.IR], 2022.
- Bahdanau, Dzmitry et. al., “Neural Machine Translation by Jointly Learning to Align and Translate”, arXiv: 1409.0473 [cs.CL], 2014.
- Bengio, Yoshua et. al., “Learning Long-Term Dependencies with Gradient Descent is Difficult”, IEEE Transactions on Neural Networks, Volume: 5, Issue: 2, 1994.

- Carneiroa, Davide et. al., “Online Dispute Resolution: an artificial intelligence perspective”, *Artificial Intelligence Review*, vol.41, 2014
- Devlin, Jacob et. al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, arXiv: 1810.04805 [cs.CL], 2019.
- Feng, Yi, Chuanyi Li, Vincent Ng, “Legal Judgment Prediction via Event Extraction with Constraints”, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 648-664, Dublin, Ireland, Association for Computational Linguistics, 2022.
- Fuller, L. L., “The Legal Mind”, The Atlantic, 1933.
- Hochreiter, Sepp et. al., “long short-term memory”, *Neural Computation* 9(8):1735-1780, 1997.
- HOPFIELD, J. J., “Neural networks and physical systems with emergent collective computational abilities”, *Proc.NatLAcad.Sci. USA*, Vol.79,pp.2554-2558, 1982.
- Katz, Daniel Martin and Bommarito, Michael James and Gao, Shang and Arredondo, Pablo, “GPT-4 Passes the Bar Exam”, 2023.
Available at SSRN: <https://ssrn.com/abstract=4389233>
- Kojima, Takeshi et. al., “Large Language Models are Zero-Shot Reasoners”, arXiv: 2205.11916 [cs.CL], 2022.
- Louis, Antoine and Gerasimos Spanakis, “A Statutory Article Retrieval Dataset in French”, arXiv: 2108.11792 [cs.CL], 2021.
- Ma, Yixiao Ai, Qingyao Wu, Yueyue Shao, Yunqiu Liu, Yiqun Zhang, Min Ma, Shaoping, “Incorporating Retrieval Information into the Truncation of Ranking Lists for Better Legal Search”, *SIGIR '22: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2022.
- MCCULLOCH, WARREN S., AND WALTER PITTS, “A Logical Calculus of Ideas Immanent in Nervous Activity”, *Synthese* 162(2):235-250, 1943.
- Mikolov, Tomas et. al., “Efficient Estimation of Word Representations in Vector Space”, arXiv: 1301.3781 [cs.CL], 2013.
- Raffel, Colin et. Al, “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer”, arXiv:1910.10683 [cs.LG], 2019.
- Radford, Alec et. al., “Improving Language Understanding by Generative Pre-Training”, 2018. Available at: <https://www.mikecaptain.com/resources/pdf/GPT-1.pdf>
- Rosenblatt, F., “The perceptron: A probabilistic model for information storage and organization in the brain”, *Psychological Review*, 65(6), 386-408, 1958.
- Turing, A. M., “Computational Machinery and Intelligence”, *Mind*, Volume LIX, Issue 236,

1950.

Vaswani, Ashish et. al., “Attention Is All You Need”, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017.

Wei, Jason et. al., “Chain of Thought Prompting Elicits Reasoning in Large Language Models”, Advances in Neural Information Processing Systems 35, 2022, arXiv:2201.11903 [cs.CL], 36th Conference on Neural Information Processing Systems, NeurIPS, 2022.

Zhong, Haoxi, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, Maosong Sun, “Legal Judgment Prediction via Topological Learning”, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3540-3549, Brussels, Belgium, Association for Computational Linguistics, 2018.

■ Abstract

Current Status and Prospects of Legal Mind Development using Large Language Models

Kyeong, Kiryong · Lee, Sang Yeob · Kim, Soyeon ·
In, Kyongseok · Ji, WonSeok* · Hong, Dae Sik**

This paper presents an assessment and outlook on whether large-scale language models can possess Legal Mind, the capability to solve attorney exam questions. Initially, it introduces the developmental history of natural language artificial intelligence models, traversing from Recurrent Neural Networks (RNNs) to Word to Vector techniques for Embedding, and the application of the Transformer architecture in BERT (Bidirectional Encoder Representations from Transformers). It delineates the developmental trajectory of large-scale language models like ChatGPT with 175 billion parameters and elucidates the recent focus on the utilization of natural language AI models in the legal domain.

Subsequently, it demonstrates the incremental formation of Legal Mind within ChatGPT without the need for new learning processes by inputting South Korean civil law provisions and Supreme Court precedents into the ChatGPT Prompter.

Finally, it outlines the developmental prospects of legal AI services and discusses the impact of integrating Legal Mind into large-scale language models on the legal market. Research aimed at instilling Legal Mind within large-scale language models is expected to bring innovative changes to the legal market, and this paper proposes future changes in the legal market in line with technological advancements.

• Key Words Large Language Model, Legal Mind, Legal AI Service, Legal Tech

* Student, Sogang University Law School

** Professor, Sogang University Law School