

한국 법학교육의 새로운 미래 : 리걸 마인드를 구현한 GPT?

양정¹, 김요한², 김소연³, 지원석³, 경기룡³

¹서강대학교 ICT 법경제연구소, ²법무법인(유) 로고스, ³서강대학교 법학전문대학원

yj7do@snu.ac.kr, yohan.kim@lawlogos.com, sarahkim15@sogang.ac.kr,

jiwonseok@sogang.ac.kr, binplace@sogang.ac.kr

An Unexplored Future of Lawyers Training Systems in Korea : GPT Embodied the Legal-Mind?

요 약

미국 변호사 시험에서 GPT-4의 성능은 매우 탁월하여 쟁점 파악 능력 및 풍부한 분석 능력에서 두드러진 성과를 보였다. 실제 답안과 비교하여도 부족한 부분을 찾아볼 수 없을 정도로 우수하였다. GPT-4를 한국의 변호사 시험에 적용할 경우, 법규와 판례의 차이, 문제의 출제 방식에 따른 차이, 전문용어의 차이가 예상되는 문제점들이었다. 그럼에도 불구하고 GPT-4는 리걸 마인드의 측면에서 논리적으로 사안에 법리를 적용하여 답안을 작성하는 능력이 매우 우수하므로 한국의 고유한 법학지식을 충분히 학습시킬 경우, 한국 변호사 시험에서도 우수한 성과를 거둘 수 있을 것으로 기대할 수 있다.

1. 서 론

종래 인문학적 방법론을 채택한 학문에서의 인공지능 활용은 매우 제한적이었다. 이른 시기부터 소설이나 영화에는 인간처럼 생각하고 대화할 수 있는 소위 튜링테스트를 통과한 인공지능이 등장했으나, 최근까지도 현실에서 이런 기능을 갖춘 인공지능은 제한적으로도 구현되지 못했다. 그러나 대규모 언어 모델(LLM)이 등장하면서 훨씬 자연스럽게 언어를 구사하는 일이 가능해졌고, 인간의 배타적인 영역으로 보였던 작문이나 창작의 영역에도 서서히 인공지능이 사용되기 시작하였다.

하지만 여전히 법학분야에서의 인공지능은 그 활용이 제한적이었다. 그 이유로는 첫째, 인공지능이 답을 내놓는 방식을 완전히 해독하여 리걸 마인드를 갖춘 사람의 논리적인 문제해결 방식과 비교·검토하지 못했다는 점, 둘째, 법학에서만 사용하는 고유한 어휘와 전달방식은 일반적인 대규모 언어 모델이 학습한 일상적인 회화체계와 상이하여 답이 충분히 법률 전문가스럽지 않았다는 점 등을 꼽을 수 있을 것이다.

그러나 최근 새롭게 선보인 GPT-4는 기존의 GPT-3.5와 다르게 미국 변호사 시험에서 탁월한 성능을 보여주면서 법학분야에서의 인공지능 활용에 새로운 도약을 이뤄냈다는 평가를 받았다. GPT-4의 답안을 전체 응시자와 비교하면 상위 10%정도의 우수한 성적에 해당하는 수준

으로 미국 변호사 시험의 합격선을 훌쩍 넘겼다.¹⁾ 소위 변호사만큼 잘하는 인공지능이 현실에서 구현된 것이다.

Exam	GPT-4	GPT-4 (no vision)	GPT-3.5
Uniform Bar Exam (MBE+MEE+MPT)	298 / 400 (~90th)	298 / 400 (~90th)	213 / 400 (~10th)
LSAT	163 (~88th)	161 (~83rd)	149 (~40th)

그림 1 GPT-4와 GPT-3.5 성적 비교

이렇듯 개선된 GPT-4가 한국 변호사 시험에서도 성과를 거둘 수 있을지, 나아가 발전된 인공지능의 등장이 한국 법학교육과 법률분야에서 가지는 의미는 무엇인지 검토해보고자 한다.

2. 미국 변호사 시험의 체계와 사례 분석

2.1 리걸 마인드와 변호사 시험

리걸 마인드(Legal-Mind)²⁾는 잘 훈련된 법률전문가가 갖추어야 할 지식, 학습능력, 자세를 모두 포괄하여 통칭하는 것으로, 법적 문제를 해결하기 위해서 어떤 방식으로

1) Latest AI ChatGPT Successor GPT-4 Proffers Both Legal Promise and Legal Perils

<https://www.law.com/legaltechnews/2023/03/17/latest-ai-chatgpt-successor-gpt-4-proffers-both-legal-promise-and-legal-perils/?slreturn=20230509110025>

2) <https://www.lsd.law/define/legal-mind>

접근해야 하는지에서부터 어떤 답을 내놓는지까지의 과정을 통틀어서 작용하는 지적인 사고과정 전부를 의미한다고 볼 수 있다. 인공지능이 인간의 학습, 추론, 직각 능력을 모사(摹寫)하여 개발된다고 한다면, 인공지능이 법률전문가의 지적 능력을 모사하기 위해서는 이른바 이 ‘리걸 마인드’를 구현해내어야 한다.

변호사 시험은 미국에서든 한국에서든 이러한 리걸 마인드를 충분히 체화하였는지를 시험하는 것이 그 목적으로, 특정한 인공지능이 변호사 시험을 통과하였다면 법률전문가로 인정받을 수 있을 정도의 지적 능력을 갖추었다는 방증이 된다고 볼 수 있기 때문에 GPT-4의 미국 변호사 시험에서의 성취는 매우 의미 있는 발전이라 할 수 있다.

2.2 미국 변호사 시험의 체계와 논술형 답안 작성

미국에서 가장 널리 통용되는 통합 변호사 시험(Uniform Bar Examination ; UBE)은 National Conference of Bar Examiners(NCBE)가 주관하는 시험으로 Multistate Bar Examination (MBE, 50%), Multistate Essay Examination (MEE, 30%), Multistate Performance Test (MPT, 20%)의 세 영역으로 이루어져 있다. 이는 한국의 변호사 시험과도 유사한 구조로 각각 한국 시험의 선택형(25%), 사례형(50%), 기록형(25%)이라는 세 영역에 대응한다고 볼 수 있다.

변호사 시험에서 가장 중점적으로 평가하고자 하는 영역은 논술에 해당하는 [MEE-사례형]으로, 이는 특정한 법적 문제에 대한 논리적이고 체계적인 답변을 내놓을 수 있는지를 평가한다. 구체적으로 사례형 문제는 특정한 법적 쟁점을 포함하고 있는 사실관계를 제시하고 수험자는 제시된 사실관계에서 문제가 되는 법적 쟁점(Issue)을 추출한 뒤 자신의 법규(Rule)와 판례에 대한 법률지식을 적용(Application)하여 쟁점에 대해 논리적이고 체계적인 답안(Conclusion)을 서술하는 방식(IRAC)으로 풀어나가게 된다.

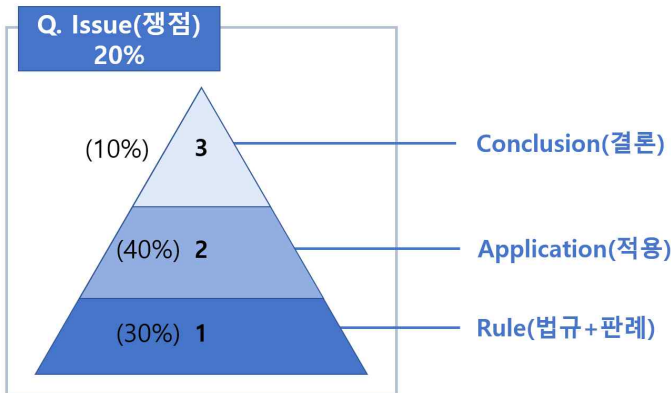


그림 2 논술형 답안의 구성

이러한 복잡하고 여러 단계의 사고과정을 거쳐서 풀어나가야 하는 사례형 문제의 해결에서 GPT-4는 기존의

chatGPT뿐 아니라 미국 변호사 시험의 합격선과 비교해서도 나은 성적을 기록하여 소위 변호사 시험에 합격할 수준이라는 평가를 받기에 이르렀다.³⁾ 이와 같은 성취는 객관식[MBE-선택형]이 아닌 장문의 텍스트를 작성하는 과제에서 일궈낸 것으로 대규모 언어 모델의 강점이 잘 발휘된 결과라 할 수 있다. 그렇다면 GPT-4가 GPT-3.5에 비해서 어떤 개선점이 있었는지, GPT-4에 실제 문제를 입력한 뒤 생성된 답안을 GPT-3.5와 비교해 보면서 확인해 보고자 한다.

MEE Question Subject	GPT-4	ChatGPT
MEE 1 - Evidence	5.0 / 6.0	3.7 / 6.0
MEE 2 - Contracts	4.2 / 6.0	3.1 / 6.0
MEE 3 - Corporations	4.4 / 6.0	3.0 / 6.0
MEE 4 - Trusts / Estates	3.9 / 6.0	2.5 / 6.0
MEE 5 - Civil Procedure	3.5 / 6.0	2.8 / 6.0
MEE 6 - Real Property	4.2 / 6.0	2.7 / 6.0
Overall Score	4.2	3.0

그림 3 MEE 문항별 점수 요약

2.2.1 민사법 문제와 답안 실험 사례

공개된 Feb 2021 NY Bar Exam의 MEE 1-1-(a) 문제의 사실관계 및 문제와 답안을 요약하면 다음과 같다.

- 한 여성은 푸드트럭 운영에 필요한 재료로 남성에게 매일 아침 시장에 가서 농산물을 구매해오게 부탁함
- 이 남성은 구매 자체에 대해서는 매일 \$20+푸드트럭 이익의 10%를 받으면서, 농산물 대금은 일단 본인의 신용카드로 구입한 뒤 페이백받기로 함
- 어느 날 남성이 실수로 어느 농부의 농산물 가판대를 들이박았으나, 남성은 보상할 금전적 돈이 없음
- 농부는 여성이 농부에게 보상금을 제공해야 한다고 주장하고 있음

1-(a) Are the woman and the man partners in the food-truck business? Explain.

- 공개답안 : 일반 파트너십은 두 당사자가 수행하기로 합의한 경우에 형성됨. 첫째로, 공동소유자로서 이익을 그들이 공유하는지 여부가 중요함. 다만, 용역의 대가로 이익을 공유하는 경우에는 이익 공유가 있어도 일반 파트너십의 추정은 발생하지 않을 수 있음. 둘째로, 비즈니스에 대한

3) Katz, Daniel Martin and Bommarito, Michael James and Gao, Shang and Arredondo, Pablo, GPT-4 Passes the Bar Exam (March 15, 2023).

통제력 행사도 요소임. 여기서 여자는 남자에게 서비스에 대한 대가로 하루에 20달러와 푸드트럭 수익의 10%를 지급함. 따라서 서비스에 대한 지불로 이익이 주어졌기 때문에 일반 파트너십의 추정이 발생하지 않음. 또한 여성은 자신의 푸드트럭 사업을 계속 관리하면서 남성이 농산물을 구매하고 배달하는 일을 제외한 사업과 관련된 모든 업무를 수행함. 따라서 남자의 통제력 부족은 여기에서 파트너십이 생성되지 않았다는 결론을 뒷받침함.

• ChatGPT(3.5) : 여자와 남자는 푸드트럭 사업에서 파트너가 아님. 파트너십은 이익을 위해 사업을 함께 운영하는 두 명 이상의 개인 간의 법적인 관계임. 여기서 여자는 푸드트럭 사업을 직접 소유하고 운영하며, 남자는 특정 작업인 농산물 구매와 배송을 위해 고용되었음. 남자는 합의된 대로 푸드트럭의 이익의 10%를 보상으로 받을 뿐임.

• GPT-4 : 여자는 어떤 농산물을 구매할지 결정하고 푸드트럭 사업의 모든 다른 측면을 관리함. 남자는 단지 서비스를 제공하는 것뿐임. 그는 이익의 일정 비율을 받지만, 이는 파트너십에서 발생하는 이익의 공유보다는 커미션과 더 유사함. 그는 사업의 관리에 대해 어떠한 의사결정권도 가지고 있지 않으며, 손실을 공유하지도 않음. 또한 남자가 여자를 대신하여 계약이나 다른 사업 결정을 할 수 있는 대리권한이 없음. 이는 파트너십의 또 다른 핵심 요소임. 그는 특정 작업(농산물 구매)을 위한 권한을 부여받았고, 여성을 대신하여 다른 결정을 내릴 권한은 없음. 따라서 여자와 남자 사이에는 분명히 사업적 관계가 있지만, 이것이 파트너십 수준에 이르지 않는 것으로 보임. 이것은 고용 관계로 더 보기 쉬움.

위 민사법 사례에서 GPT-4는 GPT-3.5에 비해서 사업의 의사결정 권한이라는 판단기준을 추가로 제시하고 있으며, 공개된 답안에 비해서 “고용관계”에 가깝다는 추가적인 판단을 적시하고 있다.

2.2.2 형사법 문제와 답안 사례

공개된 Feb 2022 NY Bar Exam의 MEE 2-1-(b) 문제의 사실관계 및 문제와 답안을 요약하면 다음과 같다.

- 한 여성이 A주에서 조경사업을 운영하는데, 근처에서 일 하던 도중 전지가위를 소지한 상태로 요금을 미납한 집주인을 찾아가
- 자신이 받은 조경서비스 대금 \$100를 3개월이나 연체한 집주인은 여성이 전지가위를 소지한 상태로 돈을 요구하자 도둑인줄 알고 해당 금액을 내어줌
- 3개월간 연체된 요금을 회수했으나 지연에 대한 대가를 받지 못한 여성은 집주인의 허락 없이 조경용 조각상(figurine)을 가지고 옴
- 여성은 자신의 조수에게 \$200 가치의 조각상을 \$10에 팔고자 시도하였음
- A주에서는 절도를 재산 소유자의 재산을 영구적으로 박탈

탈할 의도로 다른 사람의 재산이나 권리권으로부터 불법적으로 재산을 가져가는 행위로 정의함

1-(b) theft of the figurine? Explain.

• 공개답안 : 여자는 조각상을 훔쳤습니다. A주에서 절도는 재산 소유자로부터 영구적으로 박탈할 의도로 다른 사람의 재산이나 권리를 불법적으로 가져가는 것입니다. 여기에서 여자는 절도의 모든 요소를 만족시킵니다. 그 여성은 집주인의 허락 없이 불법적으로 조각상을 가져갑니다. 그녀는 집주인의 잔디밭에서 “조각상”을 옮겨왔습니다. 그 여자는 조각상을 자신의 트럭에 싣고 차를 몰고 가버렸습니다. 조각상은 집주인의 재산에 속해 있었기 때문에 조각상은 타인의 재산입니다. 마지막으로, 그 여성은 집주인에게서 자신의 소유물이 아닌 조각상을 영구적으로 박탈할 의도가 있었습니다. 사실관계에 따르면 그녀가 뭔가 더 받을 자격이 있다고 생각했음을 나타냅니다. 또한 그 여자가 입상을 잡았을 때 집주인이 보고 있지 않은지 확인하기 위해 어깨 너머로 훔쳐 보았다는 사실도 있습니다. 게다가 그 여자는 그녀의 조수에게 입상을 팔겠다고 제안했고 실제로 팔았습니다. 이러한 사실은 그 여성이 집주인에게서 자신의 재산을 영구적으로 박탈하려는 특정한 의도를 가지고 있었음을 보여줍니다.

• ChatGPT(3.5) : 조각상의 절도에 대해 답변드리겠습니다. 절도는 재산을 불법적으로 가져가고 옮기는 행위를 의미합니다. 그 여성은 주택 소유주의 정원에서 청동 정원 조각상을 훔친 후, 그녀의 트럭에 넣고 도주하였습니다. 여성이 조각상을 가져갈 때, 그녀는 주택 소유주의 허락 없이 물건을 빼앗아 갔으며, 이는 불법적인 행위입니다. 따라서, 여성은 조각상에 대한 절도를 저지른 것으로 볼 수 있습니다. 위의 상황에서 여성은 소유주의 허가나 권한 없이 조각상을 채취하였으므로, 절도로 간주될 수 있습니다. 그리고 조각상은 불법적으로 얻은 재산에 해당하므로, 이후에는 도난된 재산의 형법적 소지에 대한 판단도 필요할 수 있습니다. 이에 따라, 여성은 조각상에 대한 절도를 저지른 것으로 판단됩니다. 상황과 법적 규정의 세부사항을 고려하여 이러한 판단이 이루어져야 합니다.

• GPT-4 : 각 범죄의 모든 요소를 분석할 때, 이 여성이 조각상에 대한 도난을 저질렀는지에 대해 설명하겠습니다. “도난”은 타인의 소유물을 불법적으로 가져가고 그 소유물을 영구적으로 소유자로부터 박탈하려는 의도가 있는 경우를 의미합니다. 여성은 훔오너의 전원에서 청동 조각상을 가져갔고 이것은 훔오너의 허락 없이 이루어진 행위이므로 이것은 불법적인 행위입니다. 또한, 그녀가 조각상을 그녀의 보조에게 팔아서 그녀의 이익을 취하려고 했던 것으로 보아, 그녀는 훔오너를 영구적으로 그 소유물로부터 박탈하려는 의도가 있었습니다. 따라서, 이 여성은 “도난” 범죄의 모든 요소를 충족시키는 것으로 보입니다. 그녀는 불법적으로 소유물을 가져가고 그 소유물을 영구적으로 소유자로부터 박탈하려는 의도가 있었습니다. 그러므로, 그녀는 조각

상에 대한 도난을 저질렀다고 볼 수 있습니다.

위 형사법 사례에서 GPT-4는 GPT-3.5에 비해서 영구적인 박탈의도라는 주관적 구성요소에 해당하는 판단기준을 추가로 제시하고 있으며, 범죄의 성립여부를 해당 범죄의 “구성요소” 해당여부에 따라 판단한다고 적시하고 있다. 이외에 “상황과 법적 규정의 세부사항을 고려하여 이러한 판단이 이루어져야 합니다” 와 같은 추상적이고 무익한 기재도 찾아볼 수 없다. 공개된 답안과 비교하였을 때에도 빠진 논점 없이 모두 기술하고 있다.

2.2.3 총평

이러한 GPT-4의 우월한 성능은 작성된 답안의 길이와 유창성에 기인할 뿐 아니라 향상된 쟁점 파악 능력 및 풍부한 분석 능력에도 힘입은 결과라 할 수 있다. 실제 공개되었던 시험답안과 비교하여서도 부족한 부분을 찾아볼 수 없을 정도이다.

2.3 새로운 과제 : 한국의 변호사 시험

그러나 GPT-4가 내놓은 놀라운 결과에도 불구하고 여전히 한국 변호사 시험에서도 잘 작동할지는 의문이다. 이에 한국의 사례형 문제를 예시로 쟁점 추출 및 답안 작성의 각 단계에서 사람과 GPT-4가 작업을 수행한 경우를 간단하게 비교해 본다.

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 소유권이전등기를 경로하여 주기로 약정하였다. 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 경로되어 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).

- 사람이 작성한 쟁점
- 1. 이행지체를 이유로 해제할 수 있는지 여부

2. 이행불능을 이유로 해제할 수 있는지 여부
 3. 해제로 인한 A와 B 사이의 원상회복청구 거부
 4. 손해배상청구 거부
- GPT-4가 추출한 쟁점
 - 1. 매매계약 해제의 합법성
 - 2. 원상회복청구의 가능성
 - 3. 손해배상청구의 가능성

위 예시에서 1번 쟁점 중 이행지체를 이유로 해제할 수 있는지 여부의 결론에 대하여 사람의 답변과 GPT-4의 답변은 아래와 같다.

- 사람이 작성한 답변
- 2016.5.1. 이행기가 도래하였음에도 A는 소유권이전등기의무를 지체하였으며, 이에 특별한 사정이 없는 한 귀책사유 도 있다고 할 것이다. 다만 2016.5.1. 이행기가 도래하였으므로 A는 동시이행항변권을 갖고, B의 이행 또는 이행제공이 없으므로, A가 이행기에 이행을 하지 않더라도 위법하지 않으므로 이행지체 책임을 지지 않는다. B는 A의 이행지체를 이유로 매매계약을 해제할 수 없다.
- GPT-4가 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 경로해야 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.

검토해보면 GPT-4의 결론이 틀렸는데, 이는 “동시이행관계”에 대한 판례의 법리가 학습 데이터에 포함되어 있지 않기 때문으로 보인다. 그래서 GPT-4에게 동시이행관계에 대한 판례에 따른 법리를 알려주고 다시 결론을 내도록 지시하였다.

- GPT의 prompt에 입력한 정보
- “쌍무계약에서 쌍방의 채무가 동시이행관계에 있는 경우 일방의 채무의 이행기가 도래하더라도 상대방의 이행제공이 있을 때까지는 그 채무를 이행하지 않아도 이행지체의 책임을 지지 않는 것이다.”
- 정보를 입력한 뒤 GPT-4가 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 잔금지

급과 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.

정리해보면 GPT-4가 판단할 때에 판례에 의해 실시된 법리를 제공하고 다시 결론을 내도록 지시한 결과 법리에 따라 올바른 답변을 생성하였다. 이처럼 특정 정보를 입력해 주는 것이 GPT-4가 더 적합한 결론을 도출하는데 도움이 되는 것으로 보인다. 나아가 한국의 경우에도 GPT-4에게 적절한 법규 및 판례에 따른 법리가 올바르게 제공된다면 미국의 사례에서처럼 주어진 사실관계에 적합한 답안을 훌륭하게 작성할 수 있을 것으로 보인다.

그러나 이에 새로운 문제가 제기된다. 과연 판례 위주의 보통법(Common Law) 체계를 학습한 미국의 인공지능이 법조문을 우선하는 한국의 성문법 체계에서 잘 작동할 수 있을까? 한국은 영미의 보통법과는 법체계부터 차이가 나는데, GPT-4는 이러한 차이를 충분히 학습하지 못하였다. 때문에 GPT-4가 한국의 변호사 시험에 적용될 때에 예상되는 문제점을 추가로 검토해보고자 한다.

3. 한국의 변호사 시험에의 적용과 한계

3.1 적용시 예상되는 문제점

3.1.1 법규와 판례의 상이성

한국의 변호사 시험은 공법, 민사법, 형사법, 전문적 법률분야에 관한 과목으로 구성되어 있다. 이중 민사법과 형사법은 전세계 공통의 규칙으로 통용되는 부분이 많은 비중을 차지하고 있어, 법적 문제 해결을 위한 리걸 마인드의 측면에서 차이가 난다고 하기는 어렵다. 예를 들어 사람을 죽이면 안된다는 규정은 구체적인 살해 방식, 동기, 양태들에 따라서 다르게 분류되어 있을지언정 미국에서나 한국에서나 마찬가지로 금지되어 있다.

하지만 양국의 법조문과 판례는 서로 상이하다. IRAC에 따라서 서술할 때에 Rules에 해당하는 내용이 전혀 다른 것이다. 인공지능에게 어떤 법률을 전제로 분석할 것인지 지정하지 않는다면 답안은 전혀 엉뚱한 근거를 들어서 결론을 도출할 수 있다. 이러한 차이는 GPT-4에 한국 변호사 시험이 생각보다 어렵게 느껴지는 장벽이 될 수 있다.

3.1.2 쟁점 추출 및 결론 도출의 문제

미국 변호사 시험은 단문화되어 구체적인 쟁점을 상당 부분 문제에서 제시하고 있다. 종래 한국의 변호사 시험이 쟁점 추출까지 시험해왔던 것에 비해 GPT-4에게 유리한 방식이라 할 수 있다. 예를 들어 미국의 시험은 특정 사실관계에서 “여성과 남성은 파트너십인가”라고 질문하여 “파트너십”이 쟁점임을 적시하고 있다. 하지만 한국의 시험은 “법원은 어떠한 판단을 하여야 하는지 ①결론(소 각하/청구 기각/청구 인용/청구 일부 인용 - 일부 인용의 경우 인용 범위를 특정할 것)과 ②논거를 서술하십시오”와 같이 추상적인 질문을 던지면서 직접 쟁점-논거-결론의 논리적인 답을 수험자가 내놓기를 바라고 있다. 특히 구체적인 사실관계에서 추상적인 쟁점을 추출해야 하는 작업은 분명 GPT-4에게 부담이 될 것이다. 다만 최근 한국의 변호사 시험도 점차 사실관계나 문제에서 쟁점을 제시하는 형태로 점차 변화하고 있어 이러한 문제는 적용적으로 극복할 수 있을 것으로 기대할 수 있다.

3.1.3 전문적인 언어 능력의 한계

앞서 언급한 문제들이 해결되어 한국의 변호사 시험은 한국어로 답변하고, 미국의 변호사 시험은 영어로 답변하는 차이가 있을 뿐이라면 매우 뛰어난 번역능력을 자랑하는 대규모 언어 모델은 이러한 언어상의 차이를 쉽게 극복할 수 있을 것처럼 보인다. 그러나 리걸 마인드의 범용성에도 불구하고 전문적인 법률용어 구사로 인한 차이는 쉽게 극복되지 않을 것으로 예상된다.

한국의 법조문과 판례는 점차 순화되어가고 있기는 하나 여전히 ‘법률용어’로 지칭되는 특이한 용법에 따라 작성되어 있다. 예를 들어 ‘보상’과 ‘배상’은 일상적인 언어 체계에서는 유의어에 해당한다. 그러나 법적인 측면에서 보상은 적법한 행위에 의해서 발생한 피해를 갚아주는 것인 반면 배상은 타인의 권리를 불법적으로 침해하여 발생한 피해를 갚아주는 것을 뜻한다. 대규모 언어 모델이 학습한 일반적인 언어 대중의 체계와 전문가들이 사용하는 법률용어의 체계가 이처럼 다른 경우, 인공지능이 생성해내는 산출물은 미묘한 차이를 구분하여 작성되지 못할 수 있고 결과적으로 잘 훈련된 전문가의 수준에 미달하게 될 것이다.

3.2 결론

GPT-4를 한국의 변호사 시험에 적용할 경우 발생할 수 있는 문제점들을 검토해보았다. 주로 법규와 판례의 차이, 문제의 출제 방식에 따른 차이, 전문용어의 차이가 예상되는 문제점들이었다.

그럼에도 불구하고 GPT-4는 리걸 마인드의 측면에서 논리적으로 사안에 법리를 적용하여 답안을 작성하는 능력이 매우 우수한 것은 사실이고, 이미 알려진 바와 같

이 쟁점을 추출하는 능력도 부족하다고 볼 수는 없다. 따라서 한국의 고유한 법학지식을 충분히 학습시킬 경우, 한국 변호사 시험에서도 우수한 성과를 거둘 수 있을 것으로 기대할 수 있다.

4. 한국 법학교육과 인공지능 : 결론을 같음하여

앞서 검토한 한국의 변호사 시험에 발생할 수 있는 문제들 이외에 GPT-4에 제기되는 일반적인 문제들도 있다.

첫째, 대규모 언어 모델에 따라 생성되는 답변은 지어내기(hallucination)의 문제가 발생한다. 다시 말해 아무런 법조문이나 판례 없이 적당한 말로 이를 지어낼 수 있다는 뜻이다. 정확성이 매우 중요한 법적 문제를 해결할 때에 GPT를 활용하는 범위가 제한되는 가장 큰 원인이라 할 수 있다. 법률 문서를 작성하거나 답변을 할 때에 잘못된 근거에 따라 답변하거나 심지어 잘못된 답변을 한다면 신뢰성이 떨어지는 것은 물론이고 이에 따른 법적 책임을 져야 할 수도 있을 것이다.

둘째, 한국 고유의 법학 데이터베이스를 어떻게 구축하여 인공지능에게 학습시킬 것인지 문제가 된다. 미국은 상대적으로 소송과 관련된 서류를 폭넓게 공개되어 학습에 용이하나, 한국은 소송서류는 물론이고 하급심 판결문도 대중에게 잘 공개하지 않는다. 따라서 대규모 언어 모델이 학습할 수 있는 자료의 양이 아주 적을 뿐 아니라, 공개되어 있는 자료들도 고도로 정제되어 있다. 따라서 인공지능에 한국의 법리를 내재화하기 위해서 구체적인 사례로부터 추상적인 개념을 추출하거나 추상적인 개념을 구체적인 사례에 적용하는 등의 복잡한 사고 과정을 훈련하기에는 자료 부족의 문제가 심각하게 나타날 수 있을뿐더러, 미국의 법리와 차이가 나는 부분에 대해서는 잘못된 답을 내놓을 가능성이 있다.

잘 훈련된 법률전문가를 대체하는 것을 목표로 나날이 발전하고 있는 인공지능에 한국의 법학교육은 어떻게 대응할 수 있을까? 인공지능에 대체될 수 있는 암기와 작문보다는 법적 추론능력과 직관적인 문제해결능력 개발에 집중하고, 나아가 변호사 시험도 지식의 양을 측정하기보다 이러한 능력을 평가하는 방향으로 발전해야 한다고 볼 수 있을 것이다. 하지만 반대로 앞서 언급한 인공지능의 한계인 틀린 결론을 생성하는 점을 보완할 수 있도록 항상 정확한 답을 도출하도록 엄격하게 훈련하는 방향으로 발전시켜야 한다고 볼 수도 있을 것이다. 어느 쪽이든 발전해나가고 있는 인공지능을 염두에 두지 않고서는 한국 법학교육의 진로를 설정하기는 쉽지 않을 것이라는 점은 분명하다.

참 고 문 헌

[1] Katz, Daniel Martin and Bommarito, Michael James and Gao, Shang and Arredondo, Pablo, GPT-4 Passes the Bar Exam (March 15, 2023).