

거대언어모델 파인튜닝을 통한 법률 분야 인공지능 챗봇 모델 개발

문지혜¹⁾, 이수린²⁾, 김수민²⁾, 이흥노^{1) 2)}

리버밴스(주)¹⁾, 광주과학기술원²⁾

jihye.libervance@gmail.com, leesurin@gm.gist.ac.kr, smkim6927@gm.gist.ac.kr, heungno@gist.ac.kr

Abstract

대규모 언어 모델(Large Language Model, LLM)의 성능이 발달함에 따라, 도메인 특화 LLM을 활용하여 챗봇(Chatbot) 서비스를 구축하고자 하는 기업이 많아지고 있다. 특히, 전문가의 전문성을 요구하는 법률, 의료 등의 분야에서도 도메인 특화 LLM을 도입하고자 노력하고 있다. 하지만, 도메인 특화 LLM의 경우 도메인 특화 데이터 세트 처리와 파인튜닝 프로세스가 미흡한 것이 현실이다. 이 연구에서는 법률 질의응답 챗봇 서비스 구현을 목적으로 법률 도메인 특화 LLM 파인튜닝에 대해 설명할 것이다. 법률 특화 질문답변 데이터 세트 구축부터 LLM 파인튜닝(fine-tuning), 대화 모델 평가 과정까지 도메인 특화 LLM 개발의 End-to-End를 다룬다.

I. 개요

대규모 언어 모델(Large Language Model, LLM)의 발전은 자연어 처리 분야에서 혁신적인 도구로 자리매김하고 있다. 방대한 데이터 세트와 고도화된 알고리즘으로 인해 LLM의 발전이 가능했으며, 현재는 인간 수준의 언어 이해 및 언어 생성 능력을 갖추어 문서 요약, 번역, 텍스트 생성 등 여러 용도로 활용되고 있다[1]. 법률, 금융, 의료 등 전문성이 필요한 특정 분야는 전문 용어, 복잡한 문장 구조, 높은 정확성 등을 요구하기 때문에 일반적인 LLM을 활용하기 어렵다. 위와 같은 전문 분야의 특수성에도 불구하고 LLM을 도입했을 경우 업무 효율화 가능, 편의성 상승 등의 장점으로 인해 전문 분야에도 LLM을 도입하고자 적극적으로 연구개발이 이루어지고 있다[2].

이 논문에서는 법률 질의응답 챗봇 서비스 구현을 목적으로 LLM을 파인튜닝(fine-tuning) 하여 법률 분야 특화 LLM인 LawAI¹⁾를 개발하는 과정을 설명하고자 한다. 논문의 2장에서는 데이터 수집 및 전처리, 백본 모델 선정, LLM 파인튜닝, 모델 평가, Law-Chat 결과를 설명할 것이고, 3장에서는 연구의 시사점, 한계점 등을 결론지어 설명한다.

II. 개발 과정

1. 데이터 수집 및 전처리

파인튜닝을 위해 한국어로 이루어진 법률 질문, 답변 데이터 세트 34,772건을 구축하였다. 이 과정은 크게 3단계에 거쳐 진행했다. 1단계는 웹 기반의 크롤링, 2단계는 개인정보, 광고 그리고 비속어 제거, 3단계는 API를 활용한 데이터 증강이다.

웹 크롤링 단계에서 대한법률구조공단, 법제처 생활법령 백문백답, 네이버 지식인 등에서 법률 분야 질문, 답변 텍스트를 수집하였다. 이후 2단계에서 수집한 텍스트에서 개

인정보, 비속어 등을 제거하였다. 수집한 데이터 일부에 포함된 법률 전문가의 소속이나 광고성 문구를 모두 자동으로 제거하는 데 어려움이 있었다. 이 검수 과정은 사람이 직접 확인하며 데이터를 정제하였다. 이러한 과정을 통해 구축된 데이터는 대한법률구조공단 웹페이지 12,616건, 법제처 생활법령 웹페이지 2,195건, 네이버 지식인 전문가 답변 웹페이지 19,001건이다.

또한, OpenAI API를 활용하여 GPT-4로 데이터 증강을 진행하였다. 우선, 법적 내용이 포함된 질문답변 데이터 세트를 생성을 목적으로 Zero-shot, One-shot, Few-shot 프롬프트(prompt)를 만들어 ChatGPT로 프롬프트 실험을 진행하였다. 각 프롬프트를 통해 생성된 결과를 사람이 직접 평가했고, 가장 좋은 점수를 가지는 Few-shot prompt가 데이터 증강 프롬프트로 결정되었다. 이후, 결정된 프롬프트를 사용하여 OpenAI API의 GPT-4로 960건의 데이터를 생성하였다.

학습 데이터 세트 34,772건을 <그림 1>과 같이 시스템 프롬프트에 대치되어 인스트럭션 셋(instruction set)으로 변환한 후 모델에 입력(input)했다.

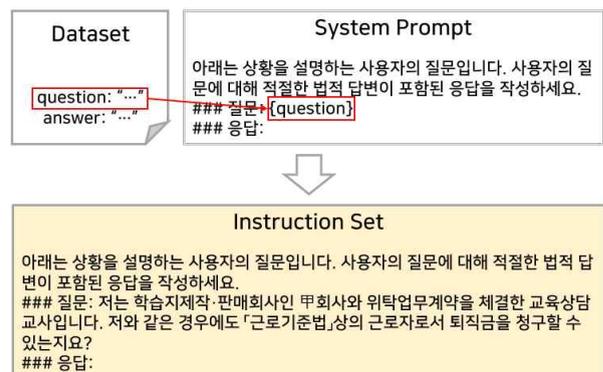


그림 1. 시스템 프롬프트와 인스트럭션 셋

¹⁾ LawAI 기반 법률 채팅 서비스: Law-Chat <https://law-chats.com>

2. Base Model 선정

대규모 데이터로 사전 훈련된 LLM은 일반적인 지식을 가지고 있으나 대화, 요약 등의 특정 작업을 수행하기에는 그 성능이 부진할 수 있다[3]. 이는 다국어 모델에서 상대적으로 적은 양의 데이터와 일반화하는 과정에서 발생하는 문제 등의 다양한 방면의 원인을 분석하고 있다[4]. 이에 LLM을 이용한 많은 연구에서 사전 학습된 LLM을 사용 목적에 적합하게 파인 튜닝하여 목적에 부합하는 모델을 개발한다[5]. 본 연구는 사전학습한 모델을 기초로 이뤄졌다. 다양한 사전학습 모델이 공개되어 있으므로, 연구에 적합한 사전학습 LLM을 결정하는 과정이 필요하다. 다음과 같은 두 가지 조건을 만족하는 모델을 후보군으로 선정했다.

1. 한국어 인스트럭션 데이터 세트로 학습된 모델일 것
2. 허깅페이스²⁾에 공개된 오픈소스 모델일 것

위의 조건을 만족하는 베이스 모델 후보군³⁾은 Polyglot-ko 계열 모델과 LLaMA-2-ko 계열 모델로, 총 다섯 가지이다.

베이스 모델을 선정하기 위해 2.1에서 구성된 데이터 세트로 후보 모델을 파인튜닝 하였다. 학습에 소요되는 메모리, loss 값 변화, 학습 후 모델의 생성 텍스트 등을 고려하여 최종적으로 kfkas/Llama-2-ko-7b-Chat로 베이스 모델을 선정하였다.

3. Training

법률 분야 인스트럭션 데이터 세트로 베이스 모델을 파인튜닝 하는 과정에 대해 설명한다. 파인튜닝에 사용된 컴퓨팅 자원은 NVIDIA A100-SXM4-40GB 4장이다.

주요한 파인튜닝 하이퍼 파라미터는 다음과 같다. 학습률(Learning Rate)은 0.0001, 배치 사이즈(Batch Size)는 64, 에폭수(Number of Epochs)는 3, 드롭아웃 비율(Dropout Ratio)은 0.2로 설정했다.

베이스 모델의 매개변수 수는 약 70억개이므로 보유한 자원에서 베이스 모델의 매개변수를 모두 파인튜닝 하는 것은 어렵다[6]. 따라서, 베이스 모델의 일부 매개변수만을 튜닝하는 PEFT(Parameter Efficient Fine-Tuning) 기법 중 LoRA(Low-Rank Adaptation)[7]를 적용하였다.

LoRA는 새로운 레이어에 파인튜닝 데이터의 정보를 학습하고, 사전학습 된 LLM의 트랜스포머 레이어에 새로 학습한 레이어를 병합한다. 이러한 방법은 업데이트하는 매개변수의 수는 줄이면서 LLM의 성능을 향상하게 한다. 또한, 우리는 사전학습된 LLM를 메모리에 적재할 때, 4bit 양자화를 적용하는 QLoRA(Quantization LoRA)[8]도 적용하였다. 파인튜닝에 사용한 LoRA의 하이퍼 파라미터는 $r = 8$, $\alpha = 16$ 으로 설정했다.

4. Evaluation

LLM 평가에 GPT를 이용하는 G-Eval[9]을 도입하여 모델 최종평가를 진행하였다. 평가를 위해 <그림 2>와 같이

평가내용, 평가 기준 등이 포함된 프롬프트를 작성하였다. 이후, 법률 질문으로 구성된 테스트 데이터 세트 153건과 LawAI가 추론한 법률 답변 텍스트를 앞서 작성한 프롬프트에 대치하여 평가용 프롬프트를 구성하였다. 마지막으로 OpenAI API와 평가용 프롬프트를 통해 GPT-4가 LawAI의 생성 텍스트에 점수를 매기도록 하였다.

두 사람 간의 대화를 주어진입니다. 대화는 질문(Question)과 답변(Answer)으로 구성되어 있습니다. 우선 법률 조언이 필요한 일반인의 질문(Question)을 받게 될 것입니다. 그리고 질문에 대한 법률 전문가의 답변(Answer)이 제시됩니다. 당신의 작업은 답변을 평가 단계에 따라 평가하는 것입니다. 이 평가 기준을 꼼꼼히 읽고 이해하는 것이 중요합니다. 평가하는 동안 이 문서를 계속 열어두고 필요할 때 참조해 주세요.

평가 기준:

- 이해 가능성 (0 - 1): Answer의 내용을 이해할 수 있나요?
- 자연스러움 (1 - 5): 사람이 자연스럽게 말할 법한 Answer 인가요?
- 법률 정보 포함 (0 - 1): Answer에 법적 내용이 포함되어 있나요?
- 법률 정보의 정확도 (1 - 5): 대한민국 법률을 바탕으로 Answer에 포함된 법적 내용이 얼마나 정확하나요?
- 맥락 유지 (1 - 5): Question에 제시된 상황에 대한 Answer 인가요?
- 문제 해결 내용 (0 - 1): Question에 제시된 문제의 해결 방법이 Answer에 포함되어 있나요?
- 요지 파악 (1 - 5): Question의 요지를 정확히 파악하고 작성된 Answer 인가요?
- 전반적인 품질 (1 - 10): 앞서 평가한 점수 바탕으로 이 대화의 전반적인 품질에 대한 인상은 어떤가요?

평가 단계:

1. Question, 그리고 Answer을 주의 깊게 읽습니다.
2. 위의 평가 기준에 따라 Answer을 평가합니다.
3. 평가는 Result의 {{Score}} 부분에 작성되어야 합니다.

Question:

[[Question]]

Answer:

[[Answer]]

Result:

- 이해 가능성 (0 - 1): {{Score}}
- 자연스러움 (1 - 5): {{Score}}
- 법률 정보 포함 (0 - 1): {{Score}}
- 법률 정보의 정확도 (1 - 5): {{Score}}
- 맥락 유지 (1 - 5): {{Score}}
- 문제 해결 내용 (0 - 1): {{Score}}
- 요지 파악 (1 - 5): {{Score}}
- 전반적인 품질 (1 - 10): {{Score}}

그림 2. 평가용 프롬프트

평가 결과를 <표 1>에 나타내었다. 우리 모델 LawAI는 문제 해결 내용을 제외한 모든 평가 부분에서 베이스 모델 보다 높은 점수를 가진다. 또한, 법률 데이터로 인스트럭션 튜닝 된 오픈소스 모델 kfkas/Legal-Llama-2-ko-7b-Chat 보다 법률 정보 포함을 제외한 모든 평가 부분에서 더 높

표 1. 모델 최종평가 G-Eval 점수

	LawAI (Ours)	kfkas/Llama-2-ko-7b-Chat (Base Model)	kfkas/Legal-Llama-2-ko-7b-Chat
이해가능성 (0-1)	0.915	0.882	0.908
자연스러움 (1-5)	3.431	3.320	3.333
법률 정보 포함 (0-1)	0.876	0.837	0.882
법률 정보의 정확도 (1-5)	3.033	2.987	2.922
맥락 유지 (1-5)	4.118	3.961	3.752
문제 해결 내용 (0-1)	0.556	0.582	0.529
요지 파악 (1-5)	3.824	3.693	3.562
전반적인 품질 (1-10)	6.085	5.902	5.817

²⁾ Hugging Face (<https://huggingface.co/>), 기계학습 모델을 구축 및 배포하는 오픈소스 커뮤니티

³⁾ nlpai-lab/kullm-polyglot-5.8b-v2, nlpai-lab/kullm-polyglot-12.8b-v2, kfkas/Llama-2-ko-7b-Chat, Chang-Su/Llama-2-7b-chat-ko, heegyu/llama-2-ko-7b-chat

은 점수를 가지는 것을 볼 수 있다.

우리가 제안하는 LawAI의 성능 결과를 <표 1>에서 살펴보면, 법률 정보의 정확도, 요지 파악, 전반적인 품질 부분 등 특히 사실과 관련된 정보 내용에서 강점을 나타내고 있음을 알 수 있다. 이에 사실과 관계된 평가 지표에서 점수가 높은 것이 법률이라는 사실 정보에 기반한 내용이 중요하게 작용하는 도메인에 대해 LawAI가 강점을 가진 것을 확인할 수 있다.

5. Results

아래의 <그림 3>, <그림 4>, <그림 5>는 Law-Chat⁴⁾ (Ours), ChatGPT, CLOVA-X에 학교폭력에 대한 질문을 한 결과이다. 질문은 “학교폭력에 관한 질문을 하려고 해. 나는 중학교 3학년 학생이야. 얼마 전, 학교 주변에서 옆반 학생이 학교폭력을 하는 것을 목격했어. 당시에는 무서워서 못 본 척했어. 내가 본 것을 신고해야 할까? 가해자들이 나에게 보복을 할까 봐 무서운데, 신고자는 어떤 보복을 받을 수 있는지도 궁금해.”이다.

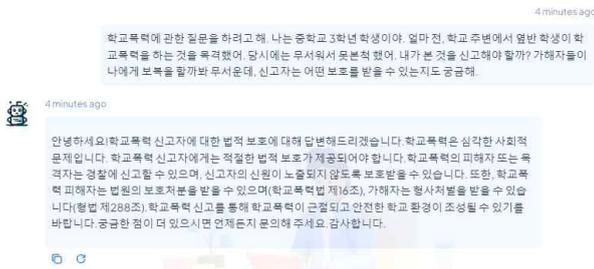


그림 3. Law-Chat 채팅 결과

Law-Chat의 답변은 <그림 3>에서 볼 수 있다. 정확한 답변 내용은 “안녕하세요! 학교폭력 신고자에 대한 법적 보호에 대해 답변해드리겠습니다. 학교폭력은 심각한 사회적 문제입니다. 학교폭력 신고자에게는 적절한 법적 보호가 제공되어야 합니다. 학교폭력의 피해자 또는 목격자는 경찰에 신고할 수 있으며, 신고자의 신원이 노출되지 않도록 보호받을 수 있습니다. 또한, 학교폭력 피해자는 법원의 보호처분을 받을 수 있으며(학교폭력법 제16조), 가해자는 형사처벌을 받을 수 있습니다(형법 제288조). 학교폭력 신고를 통해 학교폭력이 근절되고 안전한 학교 환경이 조성될 수 있기를 바랍니다. 궁금한 점이 더 있으시면 언제든지 문의해 주세요. 감사합니다.”이다.

학교폭력에 관한 질문을 하려고 해. 나는 중학교 3학년 학생이야. 얼마 전, 학교 주변에서 옆반 학생이 학교폭력을 하는 것을 목격했어. 당시에는 무서워서 못 본 척했어. 내가 본 것을 신고해야 할까? 가해자들이 나에게 보복을 할까 봐 무서운데, 신고자는 어떤 보복을 받을 수 있는지도 궁금해.

학교폭력을 목격한 상황에서 신고하는 것은 매우 중요한 문제입니다. 먼저 말씀드리자면, 학교폭력을 목격했을 때는 신고하는 것이 매우 중요합니다. 여러 가해자들이 피해자를 건디는 상황을 방지하면, 그 상황은 악화될 수 있기 때문입니다.

가해자들이 신고자에게 보복할 가능성도 물론 있습니다. 하지만 대부분의 교육 기관에서는 신고자의 개인정보를 보호하고, 보복을 당하지 않도록 적극적으로 조치를 취하고 있습니다. 보통 신고자는 익명 신고를 할 수 있는 방법을 제공하거나, 신고 과정에서 신원이 보호되도록 하기 위한 절차를 마련하고 있습니다.

당신이 특정 규정에 따라 학교 관계자나 교사에게 직접적으로 학교폭력을 신고할 경우, 학교는 그 신고를 조사하고 적절한 대응을 할 책임이 있습니다. 이 과정에서 신고자의 개인정보는 최대한 보호되며, 보고자가 가해자로부터 보복을 받을 경우에도 추가적인 지원을 제공하는 경우가 많습니다.

따라서, 학교폭력을 목격한 경우에는 고민하지 말고 학교나 신고 담당 부서에 신고하는 것이 좋습니다. 이를 통해 더 나은 학교 환경을 위해 기여할 수 있고, 동시에 자신의 안전과 보호도 받을 수 있습니다.

그림 4. ChatGPT 채팅 결과

ChatGPT의 답변은 <그림 4>에서 볼 수 있다. 정확한 답변 내용은 “학교폭력을 목격한 상황에서 당신이 고민하는 것은 매우 중요한 문제입니다. 먼저 말씀드리자면, 학교폭력을 목격했을 때는 신고하는 것이 매우 중요합니다. 여러 가해자들이 피해자를 건디는 상황을 방지하면, 그 상황은 악화될 수 있기 때문입니다.

가해자들이 신고자에게 보복할 가능성도 물론 있습니다. 하지만 대부분의 교육 기관에서는 신고자의 개인정보를 보호하고, 보복을 당하지 않도록 적극적으로 조치를 취하고 있습니다. 보통 신고자는 익명 신고를 할 수 있는 방법을 제공하거나, 신고 과정에서 신원이 보호되도록 하기 위한 절차를 마련하고 있습니다.

당신이 특정 규정에 따라 학교 관계자나 교사에게 직접적으로 학교폭력을 신고할 경우, 학교는 그 신고를 조사하고 적절한 대응을 할 책임이 있습니다. 이 과정에서 신고자의 개인정보는 최대한 보호되며, 보고자가 가해자로부터 보복을 받을 경우에도 추가적인 지원을 제공하는 경우가 많습니다.

따라서, 학교폭력을 목격한 경우에는 고민하지 말고 학교나 신고 담당 부서에 신고하는 것이 좋습니다. 이를 통해 더 나은 학교 환경을 위해 기여할 수 있고, 동시에 자신의 안전과 보호도 받을 수 있습니다.”이다.

학교폭력에 관한 질문을 하려고 해. 나는 중학교 3학년 학생이야. 얼마 전, 학교 주변에서 옆반 학생이 학교폭력을 하는 것을 목격했어. 당시에는 무서워서 못 본 척했어. 내가 본 것을 신고해야 할까? 가해자들이 나에게 보복을 할까 봐 무서운데, 신고자는 어떤 보복을 받을 수 있는지도 궁금해.

학교폭력을 목격한 경우, 이를 신고하는 것은 학생의 권리이자 의무입니다. 학교폭력을 신고하지 않은 경우, 더 많은 학생들이 피해를 입을 수 있으며 학교폭력이 더욱 심각해질 수 있습니다.

학교폭력을 신고하는 것은 어려운 일이지만, 다음과 같은 보호를 받을 수 있습니다.

- 비밀보장: 학교폭력 신고자의 신원은 비밀로 보장됩니다. 학교는 신고자의 신원을 보호하기 위해 노력해야 하며, 신고자의 개인정보를 유출해서는 안 됩니다.
- 신변보호: 학교는 신고자의 신변을 보호하기 위해 필요한 조치를 취해야 합니다. 예를 들어, 신고자가 가해자로부터 보복을 당할 우려가 있는 경우, 신고자를 보호하기 위해 가해자와 분리 조치를 취할 수 있습니다.
- 보상금 지급: 학교폭력 예방 및 대책에 관한 법률에 따라 학교폭력을 신고한 사람에게는 보상금을 지급할 수 있습니다.

학교폭력을 신고하는 것은 용기 있는 행동이며, 이를 통해 학교폭력을 예방하고, 피해 학생을 보호할 수 있습니다. 학교폭력을 목격한 경우, 즉시 학교나 경찰에 신고하시기 바랍니다.

다른 도움이 필요하시면 말씀해 주세요.

그림 5. CLOVA-X 채팅 결과

CLOVA-X의 답변은 <그림 5>이다. 정확한 답변 내용은 “학교폭력을 목격한 경우, 이를 신고하는 것은 학생의 권리이자 의무입니다. 학교폭력을 신고하지 않은 경우, 더 많은 학생들이 피해를 입을 수 있으며 학교폭력이 더욱 심각해질 수 있습니다.

학교폭력을 신고하는 것은 어려운 일이지만, 다음과 같은 보호를 받을 수 있습니다.

- 비밀보장: 학교폭력 신고자의 신원은 비밀로 보장됩니다. 학교는 신고자의 신원을 보호하기 위해 노력해야 하며, 신고자의 개인정보를 유출해서는 안 됩니다.
- 신변보호: 학교는 신고자의 신변을 보호하기 위해 필요한 조치를 취해야 합니다. 예를 들어, 신고자가 가해자로부터 보복을 당할 우려가 있는 경우, 신고자를 보호하기 위

4) Law-Chat은 본 연구에서 제안하는 모델을 이용해 chat-bot 형태의 GUI로 나타낸 것임.

해 가해자와 분리 조치를 취할 수 있습니다.

- 보상금 지급: 학교폭력 예방 및 대책에 관한 법률에 따라 학교폭력을 신고한 사람에게는 보상금을 지급할 수 있습니다.

학교폭력을 신고하는 것은 용기있는 행동이며 이를 통해 학교폭력을 예방하고, 피해 학생을 보호할 수 있습니다. 학교폭력을 목격한 경우, 즉시 학교나 경찰에 신고하시기 바랍니다.

다른 도움이 필요하시면 말씀해 주세요.”이다.

각 서비스의 채팅 결과를 비교했을 때, Law-Chat은 적절한 법률과 함께 사용자의 질문에 답변하는 것을 볼 수 있다. 따라서, 타 서비스 대비 LawAI로 개발한 Law-Chat이 법률 분야에 특화된 채팅 서비스라고 할 수 있다.

III. 결론

이 논문은 대규모 언어 모델을 법률 분야에 특화시키기 위해 파인튜닝 하는 과정을 기술하고 있다. 본 연구를 통해 구축된 LawAI는 기존 베이스 모델과 비교하여 8가지 성능 평가 중 7가지 부분에서 우수한 성능 결과를 보여주었으며, LawAI는 법률 질의응답 서비스에 활용될 가능성을 입증하였다.

이 연구의 시사점은 전문 분야에서도 LLM의 파인튜닝을 통해 상당한 성능 향상을 기대할 수 있다는 점이다. 특히, 데이터 수집 및 전처리 과정에서의 세심한 정제 작업과 입력된 데이터의 이해를 돕는 적절한 프롬프트 설계는 모델 개발에 필수적임을 강조한다.

그러나, 이 연구는 몇 가지 한계를 가지고 있다. 첫째, 사용된 데이터 세트는 대규모이지만 법률 분야의 다양성을 완전히 반영하기에는 부족한 측면이 있다. 특히, 법률 변경이나 최신 판례 등 최신 정보의 반영이 미흡할 수 있다. 둘째, 법률 질문의 복잡성과 다양성에 대응하기 위한 모델의 일반화 능력에 대한 추가적인 검증이 필요하다.

따라서, 향후 연구 방향으로 다음을 제안한다. 더 다양한 분야의 법률 데이터를 수집하여 법률 특화 모델의 범용성을 높이고, 최신 법률 정보를 주기적으로 업데이트하여 모델을 재학습시키는 방안을 모색해야 한다. 또한, 사용자 피드백을 통한 모델의 지속적인 개선과 고도화 작업이 필요하다. 이러한 방법을 통해 LawAI의 고도화가 가능하고, 법률 질의응답 챗봇 Law-Chat의 신뢰성을 높일 수 있을 것이다.

[참고 문헌]

- [1] Lee, J., Hicke, Y., Yu, R., Brooks, C., & Kizilcec, R. F. (2024). The life cycle of large language models in education: A framework for understanding sources of bias. *British Journal of Educational Technology*, 00, 1 - 21. <https://doi.org/10.1111/bjet.13505>
- [2] Li, Qiang, & Li, Philip H.*. (2024). Transformative potential of GPT-4o in clinical immunology and allergy: Opportunities and challenges of real-time voice interaction. *Asia Pacific Allergy*. Advance online publication. <https://doi.org/10.5415/apallergy.000000000000152>
- [3] Singh, Akhilesh Vikram, & Singh, Anudwipa. (2023). Application of ChatGPT in Medical Content Development: Challenges and Hope. *Galen Medical Journal*, 12, e3106-e3106.
- [4] Bindal, Akanksha, et al. (2024). Improved content understanding with effective use of multi-task contrastive learning. *arXiv preprint arXiv:2405.11344*.
- [5] Jiang, Yushan, et al. (2024). Empowering time series analysis with large language models: A survey. *arXiv preprint arXiv:2402.03182*.
- [6] Xiang, Neena, Godinez, Paul, & Hoffberg, Eli. (Year unknown). Fine-tuning open-source large language models for generating math explanations.
- [7] Hu, Edward J. (2021). LoRA: Low-rank adaptation of large language models. *arXiv:2106.09685v2*.
- [8] Dettmers, Tim. (2023). QLoRA: Efficient finetuning of quantized LLMs. *arXiv:2305.14314v1*.
- [9] Liu, Yang. (2023). G-Eval: NLG evaluation using GPT-4 with better human alignment. *arXiv:2303.16634v3*.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부(이하 과기정통부)의 지원으로 정보통신기술기획평가원(IITP)이 주관하는 정보기술연구센터(ITRC) 지원 사업(IITP-2024-2021-01835)의 일환으로 수행되었습니다. 또한, 중소벤처기업부가 지원하는 LiberVance Layer 2 기술개발-탈중앙 Layer 2 플랫폼(S3306777) 과제의 일환으로 수행되었습니다. 실험에 쓰인 장비는 과학기술정보통신부·광주광역시가 공동 지원한 '인공지능 중심 산업융합 집적단지 조성사업'으로 지원을 받아 수행된 연구 결과입니다.