

LoRA 기반 파인튜닝을 활용한 초등 경제 교육용 대규모 언어모델 개발 및 성능 평가

이태오¹, 김태국^{2*}

¹국립부경대학교 데이터공학과 학생, ²국립부경대학교 컴퓨터-인공지능공학부 교수

Development and Evaluation of a LoRA-Fine-Tuned Large Language Model for Elementary Economic Education

Tae-O Lee¹, Tae-Kook Kim^{2*}

¹Student, Department of Data Engineering, Pukyong National University

²Professor, School of Computer and Artificial Intelligence Engineering, Pukyong National University

요약 본 연구는 초등학생의 경제 개념 이해를 돕기 위한 대화형 학습 모델을 구축하고, 대규모 언어모델(LLM)의 교육적 실용성을 검증하는 데 목적이 있다. 이를 위해 초등학생의 일상 경험과 밀접한 경제 상황을 반영한 1,000개의 질의응답 데이터셋을 자체 구축하였으며, LLaMA 3.1 8B Instruct 모델에 LoRA(Low-Rank Adaptation) 기반 파인튜닝을 적용하여 경제 교육에 특화된 모델을 개발하였다. 실험 결과, 파인튜닝된 모델은 기존 모델 대비 BLEU 점수가 0.18에서 5.67로 약 31배 향상되었으며, 의미적 유사성을 측정하는 BERTScore(F1) 역시 0.6642에서 0.7743으로 증가하여 생성 문장의 품질과 정확도가 모두 개선되었음을 확인하였다. 이는 소규모 데이터셋과 제한된 GPU 자원 환경에서도 파라미터 효율적 미세조정(PEFT) 기법을 통해 도메인 특화 모델을 효과적으로 구축할 수 있음을 시사한다. 본 연구는 LLM을 활용한 초등 경제 교육 시스템 설계의 실증적 토대를 마련하였으며, 향후 사용자 평가 및 데이터 고도화를 통해 개인화된 맞춤형 경제 교육 도구로 발전할 수 있는 가능성을 제시한다.

주제어 : 대규모 언어모델, 경제 교육, 파인튜닝, LoRA, 대화형 학습

Abstract This study aims to develop a conversational learning model designed to support elementary students' understanding of economic concepts and to empirically validate the educational feasibility of Large Language Models (LLMs). To this end, a custom dataset comprising 1,000 question-and-answer pairs was constructed, capturing everyday economic scenarios closely related to the lived experiences of elementary students. A domain-specific model for economic education was then developed by applying Low-Rank Adaptation (LoRA)-based fine-tuning to the LLaMA 3.1 8B Instruct model. Experimental results demonstrated that the fine-tuned model achieved a substantial performance gain, with the Bilingual Evaluation Understudy (BLEU) score increasing from 0.18 to 5.67—approximately a 31-fold improvement over the base model. Furthermore, the BERTScore (F1), which measures semantic similarity, rose from 0.6642 to 0.7743, confirming enhanced quality and accuracy in sentence generation. These findings suggest that Parameter-Efficient Fine-Tuning (PEFT) enables the effective construction of domain-specific language models even within environments characterized by compact datasets and limited GPU resources. This study establishes an empirical foundation for designing LLM-based economic education systems for primary learners and highlights the potential for evolving into personalized, adaptive educational tools through future user-centered evaluations and dataset expansion.

Key Words : Large Language Model (LLM), Economic Education, Fine-tuning, LoRA, Conversational Learning

이 논문은 4단계 두뇌한국21 사업 (4단계 BK21 사업)에 의하여 지원되었음.

*교신저자 : 김태국(king@pknu.ac.kr)

접수일 2025년 12월 24일

수정일 2026년 01월 19일

심사완료일 2026년 02월 08일

1. 서론

현대 사회에서 경제적 문제는 개인의 일상생활과 밀접하게 연관되어 있으며, 경제적 의사결정 능력은 시민으로서의 필수 역량이다. 특히 초등학생 시기는 경제 개념이 점차 체계화되고 생활 경험을 통해 개념이 형성되는 시기이므로, 특히 중요하다. 그러나 기존 초등 경제 교육 연구에서는 경제 개념 이해도가 단편적인 지식 축적에 머무르는 경향과 객관적 이해 수준과 주관적 이해 수준 사이의 괴리 현상이 보고되었다. 예를 들어, 김기환은 초등학교의 경제 개념이 주관적으로는 이해한다고 판단하지만, 객관적인 평가 정답률은 상대적으로 낮은 경향을 확인했다[1]. 이는 전통적 지식 중심 학습이 학생의 경제적 사고 형성에 한계를 지닌다는 점을 시사한다.

최근 인공지능의 급부상으로 다양한 분야에서 이를 활용하려는 시도가 활발히 이루어지고 있다. 예를 들어 컴퓨터 비전과 딥러닝 기반 인공지능 모델은 IoT 환경과 결합되어 가정 내 화재와 같은 위험 상황을 조기에 탐지하는 시스템으로 발전하고 있으며, 이는 실생활 데이터를 활용한 지능형 판단의 가능성을 보여준다[2-4]. 또한 대규모 데이터를 기반으로 한 인공지능 분석 기법은 교육 연구 분야에서도 활용되어, 국제 대학 평가 지표나 연구 네트워크 구조를 체계적으로 분석하고 지역 간 불균형 문제를 정량적으로 규명하는 데 기여하고 있다[5,6]. 더 나아가 실내 위치 인식과 같은 IoT 핵심 기술의 발전은 인공지능 서비스가 사용자의 공간적·환경적 맥락을 보다 정밀하게 인식할 수 있는 기반을 제공하며, 이는 다양한 맞춤형 서비스 및 지능형 시스템 설계로 확장되고 있다[7]. 이러한 연구들은 인공지능이 단순한 자동화 도구를 넘어, 실제 환경에서 생성되는 대규모 데이터와 결합하여 복합적인 문제를 이해하고 지원하는 핵심 기술로 자리 잡고 있음을 시사한다.

특히 대규모 언어모델(Large Language Model, LLM)의 발전은 전통적 교육의 한계를 보완할 수 있는 새로운 교육 도구의 가능성을 제시한다. 최근 디지털 콘텐츠 및 서비스 분야 연구에서는 사용자 특성이나 이용 맥락에 따라 정보를 맞춤형으로 제공하는 접근이 제안되고 있으며[8,9], 이러한 흐름은 학습자 수준에 따라 설명 방식이 달라질 수 있는 LLM 기반 교육 도구의 필요성에도 맞닿아 있다. LLM 기반 대화형 시스템을 활용하면 학생들이 교과서의 정답을 암기하는 방식이 아니라, 일상 대화 형식으로 경제 개념을 탐구할 수 있다. 예를 들어 “주식이 뭐야?”, “카드사는 왜 생긴 거야?”와 같은 질문을 스

스로 던지고, 챗봇과의 상호작용을 통해 답을 찾아가는 과정에서 경제 개념을 보다 체감적인 맥락 속에서 이해할 수 있다. 이는 학생 개인의 관심사와 경험을 반영한 맞춤형 설명이 가능하다는 점에서, 기존 교과 중심 교육과 뚜렷한 구분이 된다.

또한 대화형 학습 환경은 단기적인 지식 습득을 넘어, 지속적인 학습 동기와 올바른 경제 태도 형성에도 기여할 수 있다. 학생이 생활 속에서 마주치는 가격 변화, 저축·소비 선택, 리스크 감수 등에 대해 질문을 던지고, 그 때마다 챗봇과 상호 작용하며, 의미를 재구성하는 경험은, 경제를 단지 시험 과목이 아니라 생활을 이해하는 언어로 인식하게 만드는 데 도움을 준다. 특히 청소년기에는 첫 금융 거래, 첫 아르바이트, 첫 투자 경험 등이 축적되는 시기라는 점에서, 이 시기에 형성된 경제 태도와 금융 문해력은 성인기 재무·투자 태도와도 밀접한 관련을 갖는다.

이런 관점에서 볼 때, 학생이 자발적으로 질문하고 탐구하는 대화 기반 경제 학습 도구는 단순한 보조 교재를 넘어 장기적인 금융 문해력과 책임 있는 경제 태도 형성에 기여할 수 있는 인프라로서 의미를 가진다. 이에 본 연구는 한국어 환경에서 초등학교 수준의 경제 개념 이해를 지원하기 위해 경제 특화 LLM 챗봇을 개발하고, 그 교육적 가능성을 탐색하는 것을 목적으로 한다.

2. 관련 연구

2.1 거대 언어모델

대규모 언어모델(Large Language Model, LLM)은 수십억 개 이상의 파라미터를 가진 거대 신경망으로, 대량의 텍스트 데이터를 활용해 일반적인 언어 패턴과 지식을 사전 학습한 모델을 의미한다[9]. 대표적인 예로 GPT 계열, PaLM, LLaMA 등이 있으며, 모두 트랜스포머(Transformer) 구조를 기반으로 대규모 말뭉치에서 자기 지도 학습 방식으로 학습된다. 이러한 LLM은 한 번 사전학습이 끝나면, 추가적인 목적 설정이나 특정 분야에 대해 별도의 구조 변경 없이도 프롬프트 입력만으로 다양한 자연어 생성·이해 과제를 수행할 수 있다는 점에서 기존의 목적별 모델과 구별된다. LLM에 대한 최근 종합 논문들은 사전학습, 적응 튜닝, 활용, 성능 평가라는 네 축에서 기술 발전이 이루어지고 있음을 정리하며, 특히 교육, 코딩, 의료, 금융 등 특정 분야 특화 LLM 응용이 폭발적으로 증가하고 있음을 보여준다[10,11].

교육 분야에서도 LLM 기반 챗봇은 개별 학습자의 질문에 맞춤형 피드백을 제공하는 도구로 주목받고 있다. 최근 LLM 에이전트를 활용해 피드백 코멘트 생성, 과제 설계, 튜터링 등을 지원하는 연구들이 등장하고 있으며[12], 국내에서도 Clinical Performance Examination(CPX) 연습을 위한 LLM 기반 챗봇을 개발하여 의학교육에 적용한 사례가 보고된 바 있다[13]. 이러한 흐름은 LLM이 단순 질의응답 도구를 넘어, 학습자가 스스로 질문을 생성하고 개념을 재구성하는 대화형 학습 파트너로 기능할 수 있음을 보여준다. 본 연구에서 다루는 초등 경제 개념 챗봇 역시, 이러한 LLM 기반 교육용 에이전트의 한국어·경제 교육 특화 버전으로 위치 지을 수 있다.

2.2 파인튜닝, LoRA의 개념

LLM은 범용적인 언어 능력을 갖추고 있지만, 특정 분야(예: 초등 경제 교육)에 맞추어 세밀하게 조정되지 않은 상태에서는 원하는 수준의 응답 품질을 보장하기 어렵다. 이 때문에 사전 학습된 LLM을 특정 목적이나 분야에 맞게 추가 학습시키는 파인튜닝이 널리 활용된다. 그러나 GPT-3, LLaMA 3과 같은 LLM은 수십억~수천억 개의 파라미터를 포함하고 있기 때문에, 전 파라미터를 모두 업데이트하는 전통적인 파인튜닝 방식은 GPU 메모리와 연산량 측면에서 매우 큰 부담을 초래한다. 이러한 문제를 해결하기 위해 최근에는 파라미터 효율적 파인튜닝(PEFT)이 활발히 연구되고 있으며, 사전학습 모델의 대부분은 고정된 채, 적은 수의 추가 파라미터만 학습하는 방식으로 특정 분야 적용을 수행한다[14,15]. 그 중, Low-Rank Adaptation(LoRA)는 Transformer의 특정 가중치 행렬에 대해, 저랭크 행렬을 추가로 학습시키는 방식으로 모델을 적응시키는 대표적인 PEFT 기법이다. Hu 등(2021)은 사전 학습된 LLM의 가중치를 그대로 고정된 상태에서, 각 레이어의 선형 변환에 대해 저랭크(rank가 매우 낮은) 행렬 쌍을 삽입해 학습하는 LoRA를 제안하였다[16]. 이를 통해 GPT-3 175B 모델을 예로 들었을 때, 전통적인 풀 파인튜닝 대비 학습해야 하는 파라미터 수를 약 10,000배까지 줄이고, GPU 메모리 사용량을 약 3배 절감하면서도 RoBERTa, DeBERTa, GPT-2, GPT-3 등의 여러 벤치마크에서 동등하거나 더 나은 성능을 달성할 수 있음을 보였다. 이후 LoRA는 LLaMA, BLOOM 등 다양한 공개 LLM의 특정 분야 적용에 널리 활용되고 있으며, PEFT 서베이 논문들에서도 어댑터 튜닝, 프롬프트 튜닝과 함께 핵심 기법으로 정리되고 있다. 본 연구는 이러한 LoRA의 장점을 활용하여,

사전 학습된 LLaMA 3.1 8B Instruct 모델의 본래 파라미터는 그대로 유지한 채, 초등 경제 Q&A 데이터셋에 특화된 저랭크 적응 파라미터만을 학습시켰다. 이와 같은 접근은 (1) 상대적으로 소규모 데이터셋으로도 안정적인 학습이 가능하고, (2) 단일 GPU 환경에서도 실용적인 자원 소모로 파인튜닝을 수행할 수 있으며, (3) 필요시, 여러 버전의 도메인 특화 모델(예: 초등 경제, 중등 경제, 금융 소비자 보호 등)을 가벼운 LoRA 어댑터 교체만으로 관리할 수 있다는 점에서 교육용 LLM 개발 시 높은 실용성을 가진다.

3. 연구 방법 및 결과

3.1 연구 방법

3.1.1 데이터셋 구축

본 연구에서는 초등학생의 일상 경험을 경제 개념과 자연스럽게 연결할 수 있도록 구성된 1,000개의 질의응답 데이터셋을 구축하였다. 데이터는 사전에 정의된 경제 교육 목적에 부합하도록 ChatGPT를 활용하여 생성하였으며, 용돈 관리, 소비 선택, 저축과 투자, 협력의 필요성, 자원 활용, 가격 변화와 같은 생활 밀착형 경제 주제를 중심으로 질문을 설계하였다. 이러한 접근은 전통적인 교과서 중심의 경제 교육이 추상적 개념 위주로 구성되어 학습자의 흥미 유지가 어렵다는 문제를 해결하고, 학생이 자기 경험과 연결해 경제 원리를 이해할 수 있도록 지원하기 위함이다.

데이터는 instruction-input-output 구조의 JavaScript Object Notation Lines(JSONL) 형식으로 구축되었으며, instruction은 모델이 이해해야 할 질문, input은 추가 맥락 또는 부가 정보(본 연구에서는 비워둠), output은 초등학생 수준에 적합한 설명형 응답으로 구성되어 있다. 데이터의 문장은 경제 용어 사용을 최소화하되 경제 개념적 의미는 유지하는 방향으로 작성되었다. 예를 들어 '기회비용'이라는 용어 대신 '하나를 사면 다른 것을 못 살 수도 있어요'와 같이 이해가 가능한 방식으로 개념을 설명하였다. 이를 통해 본 연구 데이터셋은 경제 개념의 기초 구조를 유지하면서도, 도메인 지식이 부족한 학습자도 자연스럽게 경제적 사고 과정을 습득할 수 있도록 설계되었다. 예를 들어, 아래의 [Fig. 1]은 학습 데이터 중 하나로, 질문으로 "물건을 사려면 왜 돈을 내야 해?"("Why do we have to pay money to buy

things?”) 라고 하면 답변으로 적당한 것으로 “돈은 우리가 물건이나 서비스를 받을 때 그 값으로 내는 거예요. 그래야 물건을 만든 사람도 다른 걸 살 수 있답니다!” (“Money is what we give when we get things or services, so the people who made them can buy other things too!”) 가 되는 것을 보여준다.

```
{
  "instruction": "물건을 사려면 왜 돈을 내야 해?",
  "input": "",
  "output": "돈은 우리가 물건이나 서비스를 받을 때 그 값으로 내는 거예요. 그래야 물건을 만든 사람도 다른 걸 살 수 있답니다!"
}
```

[Fig. 1] A sample of training data

또한 질문은 특정 경제 개념에 종속되지 않도록 다양한 주제를 포함하여 균형 있게 구성하였다. 전체 데이터셋은 경제 원리(희소성, 선택, 기회비용), 재화·서비스 이해, 금융과 화폐, 협력과 분업, 자원과 환경, 가격과 시장의 여섯 범주로 분류할 수 있으며, 각 범주가 전체 데이터의 10~25% 수준으로 분포하여 특정 영역에 치우치지 않도록 관리하였다. 이러한 구성은 향후 다양한 경제 학습 시나리오에서 모델이 안정적으로 질문에 대응할 수 있는 기반을 제공한다.

3.1.2 사전학습 기반 모델: LLaMA 3.1 8B Instruct

본 연구에서 사용한 기반 모델은 Meta가 공개한 LLaMA 3.1 8B Instruct 모델로, 약 80억 개(8B)의 파라미터를 포함한 중형 규모의 LLM이다. LLaMA 계열 모델은 트랜스포머 구조를 기반으로 대규모 웹 텍스트, 문서, 코드, 질의응답 데이터 등을 활용하여 사전학습 되었으며, 특히 LLaMA 3.x 모델군은 이전 버전 대비 한국어를 포함한 다국어 처리 성능과 추론 능력이 강화된 것이 특징이다. 이러한 특성은 본 연구와 같이 비영어권 언어 기반의 교육 목적 연구에서 활용 가능성을 높여준다.

본 연구에서 활용한 LLaMA 3.1 8B Instruct 모델은 사전 학습된 LLaMA 3.1 모델에 추가적으로 instruction-following 튜닝이 적용된 버전으로, 사용자의 명령이나 질문을 이해하고 맥락에 맞는 응답을 생성하는 데 최적화되어 있다. 일반적인 사전학습 언어 모델이 확률적 다음 토큰 예측에 초점을 맞춘 것과 달리, Instruct 계열 모델은 인간 피드백 기반 강화학습(RLHF, Reinforcement Learning from Human Feedback), supervised fine-tuning(SFT), preference tuning 등을 통해 사용

자 지시형 자연어 생성 능력을 강화하였다. 이는 초등학생의 질문에 직관적인 방식으로 응답해야 하는 본 연구 목적과 부합한다.

LLaMA 3.1 8B는 파라미터 수 대비 높은 성능을 제공하여, GPT-3.5 기반 상용 엔진 대비 연산 자원이 적게 소요되면서도 다양한 자연어 처리 작업에서 경쟁력 있는 결과를 보여준다. 특히 GPU 1~2장 수준에서도 가용한 모델 크기는, 교육용 데이터셋과 같이 상대적으로 소규모의 도메인 데이터로 반복적 실험과 파인튜닝을 수행해야 하는 연구 환경에서 실질적인 이점을 제공한다. 또한 공개된 모델 가치와 허용적인 라이선스 정책은 연구자들이 모델 구조를 변경하거나 추가 실험을 수행할 수 있는 개방성을 확보하게 하여, 국내외 연구 환경에서 활용도가 높다.

3.1.3 파인튜닝 도구 및 설정: TorchTune, LoRA

본 연구에서는 LLaMA 3.1 8B Instruct 모델을 초등 경제 도메인에 적합하도록 확장하기 위해, Meta가 공개한 TorchTune 프레임워크를 활용하여 파인튜닝을 수행하였다. TorchTune은 PyTorch 기반의 대규모 언어 모델 파인튜닝을 위해 설계된 경량화 틀체임으로, 모델 다운로드, LoRA 구성, 학습 파이프라인 설정 등이 통합적으로 제공되어 실험 환경 구축에 용이하다. 특히 기존 Hugging Face 기반 파인튜닝 환경 대비 설정 복잡도가 낮으며, 모델 구조 변경 없이도 다양한 어댑터 기반 실험을 수행할 수 있는 점에서 본 연구의 목적에 부합하였다.

파인튜닝은 LoRA기법을 적용하여 수행하였다. LoRA는 사전 학습된 모델의 파라미터 전체를 업데이트하지 않고, 특정 선형 변환 가중치에 대해 저랭크 행렬을 추가로 학습시키는 방식으로 모델을 도메인에 적응시키는 기법이다. 이를 통해 전체 파라미터를 재학습하는 기존 방식 대비 GPU 메모리 요구량을 크게 줄일 수 있으며, 소규모 데이터셋에서도 과적합의 위험을 완화하면서 도메인 특화 성능을 확보할 수 있다.

본 연구에서는 Query(q_proj), Value(v_proj), Output(output_proj) projection 가중치 및 MLP 계층을 대상으로 LoRA를 삽입하였다. 이는 모델의 주요 표현 구조가 위치한 self-attention 모듈과 MLP 영역에 적응 파라미터를 부여함으로써, 경제 개념 질문에 대한 문장 생성 과정에서 의미적 연결성과 문맥 반응성을 높이기 위함이다. LoRA의 저랭크 차원은 8, scaling factor는 16으로 설정하였으며, dropout은 적용하지 않았다. 이러한 설정은 소규모 도메인 데이터셋에서 파인튜닝 안정성

을 확보하는 데 적합한 구성으로 알려져 있으며, 본 연구에서도 동일한 효과를 확인하였다. 이외의 실험 환경은 아래의 <Table 1>과 같다.

<Table 1> Experimental environment

| Name | Value |
|------------------|-----------------|
| Operation System | Linux 24.04 |
| GPU | NVIDIA RTX 4090 |
| Form of Dataset | JSONL |
| Optimizer | AdamW |
| Learning rate | 3e-4 |
| Epoch | 3 |

3.1.4 평가 지표

본 연구에서는 파인튜닝된 모델의 성능을 정량적으로 평가하기 위해 BLEU(Bilingual Evaluation Understudy) 점수와 BERTScore를 활용하였다. 두 지표는 모두 자연어 생성 모델 성능 평가에 널리 사용되는 지표지만, 측정 관점이 서로 다르기 때문에 병행 사용함으로써 생성 문장의 표면적 유사도와 의미적 유사도를 균형 있게 분석하고자 하였다. BLEU는 기계 번역 성능 평가를 위해 개발된 N-gram 기반 정량 평가 지표로, 생성된 문장과 정답 문장 간의 단어 조합 유사도를 계산하여 모델의 출력 품질을 평가한다. BLEU는 문장의 표면적 형태 유사성을 측정하기 때문에, 답변 문장 구조나 단어 배열이 정답과 얼마나 유사하게 생성되는지를 평가하는 데 유용하다. 본 연구에서는 경제 개념 설명문이 비교적 짧고 특정 핵심 단어 (예: 저축, 가격 등) 등의 유지 여부가 중요한 특성이 있어, BLEU 지표가 모델의 기본적인 생성 안정성을 확인하는 데 적합하다고 판단하였다. BLEU는 단순한 N-gram 일치 여부를 기반으로 하는 반면, BERTScore는 사전 학습된 언어 모델(BERT 계열)의 임베딩 공간을 활용하여 문장 간 의미적 유사도를 평가하는 지표이다. 즉, 단어가 다르게 표현되더라도 의미가 동일하게 전달되었다면 높은 점수를 받을 수 있다는 점에서, 경제 개념 설명처럼 표현 방식이 유연한 문장에 적합하다. BERTScore는 Precision, Recall, F1 지표로 구성되며 본 연구에서는 모델 출력이 정답의 의미를 얼마나 정확하게 반영하는가를 판단하기 위해 세 가지 지표를 모두 사용하였다. 특히 본 연구의 데이터셋은 문장의 자연스러운 서술과 개념적 연결성이 중요하므로, 의미 기반 유

사도 측정을 제공하는 BERTScore는 BLEU 지표를 보완하는 평가 기준으로서 타당하다.

3.2 연구 결과

본 연구에서는 사전학습된 LLaMA 3.1 8B Instruct 모델(Base LLaMA)과 본 연구에서 구축한 초등 경제 Q&A 데이터셋을 활용하여 LoRA 기반 파인튜닝을 수행한 모델(Fine-Tuned LLaMA)의 성능을 비교하였다. 성능 평가는 BLEU 점수와 BERTScore를 활용하여 정량적으로 실시하였다. <Table 2>는 두 모델의 출력 문장에 대해 BLEU 및 BERTScore를 측정된 결과를 나타낸다. 먼저 BLEU 점수는 기본 모델이 0.18에 머무른 반면, 파인튜닝된 모델은 5.67을 기록하여 형태적 표현 유사도 측면에서 약 31배 이상 향상된 성능을 보였다. 이는 파인튜닝을 통해 모델이 경제 개념 중심 문장 구조에 보다 적절히 적응했음을 의미한다. BERTScore 역시 모든 세부 지표에서 향상이 확인되었다. 기본 모델은 Precision 0.6414, Recall 0.6910, F1 0.6642를 기록하였다. 반면 파인튜닝된 모델은 Precision 0.7772, Recall 0.7719, F1 0.7743으로 개선되어, 단순 단어 배열 차원이 아닌 의미적 유사도 측면에서도 높은 성과를 보였다. 특히 F1 점수의 상승 폭은 모델이 경제 개념의 핵심 내용을 보다 정확하게 전달하는 방향으로 학습되었음을 시사한다.

<Table 2> Base vs Fine-Tuned

| Model | BLEU | BERT Score (P) | BERT Score (R) | BERT Score (F1) |
|------------------|------|----------------|----------------|-----------------|
| Base LLaMA | 0.18 | 0.6414 | 0.6910 | 0.6642 |
| Fine-Tuned LLaMA | 5.67 | 0.7772 | 0.7719 | 0.7743 |

4. 결론

본 연구는 대규모 언어모델(LLM)을 기반으로 초등학생의 경제 개념 이해를 돕기 위한 한국어 경제 교육용 챗봇 모델을 개발하고, LoRA 기반 파인튜닝 기법의 효과를 실증적으로 평가하였다. 이를 위해 1,000개의 초등 경제 개념 질의응답 데이터셋을 구축하여 LLaMA 3.1 8B Instruct 모델에 적용한 결과, 파인튜닝 수행 전과 후의 차이가 명확하게 관찰되었다. 특히 BLEU 점수는

0.18에서 5.67로 크게 향상되었으며, 의미 기반 유사도를 측정하는 BERTScore 역시 F1 기준으로 0.6642에서 0.7743으로 상승하였다.

이러한 결과는 LoRA 기반 파인튜닝이 모델 전체를 재학습하지 않고도 도메인 지식을 효과적으로 적용할 수 있음을 시사한다. 즉, 소규모 데이터셋과 제한된 GPU 자원만으로도 경제 개념에 특화된 반응 품질을 확보할 수 있다는 점에서 본 연구의 접근은 교육 현장 적용 가능성을 시사한다. 특히 본 연구에서 구축한 데이터셋이 일상 경험과 경제 원리를 연결하는 방식으로 설계되었음을 고려할 때, 파인튜닝된 모델이 단순한 정의 암기형 답변이 아니라 의미적 맥락을 유지하면서도 초등 교육 수준에 맞춘 설명을 제공할 수 있게 되었다는 점은 중요한 성과라 할 수 있다.

종합적으로 볼 때, 본 연구는 도메인 특화 교육용 LLM 개발에 있어 LoRA 기반 파인튜닝이 현실적이고 효율적인 접근 방법임을 입증하였으며, 경제 교육 분야에서 LLM 활용의 가능성을 적극적으로 제시한다. 향후 데이터 확장, 사용자 평가, 난이도 조절 알고리즘 개발 등이 병행된다면, 본 연구에서 제시한 접근법은 개인화된 경제 교육 도구로서 실질적 보급 가능성을 갖출 것으로 기대된다.

REFERENCES

- [1] K.H.Kim, "A Study on the Subjective Understanding of Elementary School Students' Economic Concepts," *Theory and Research in Citizenship Education*, Vol.55, No.1, pp.1-21, 2023.
- [2] B.Peng, T.K.Kim, "YOLO-HF: Early Detection of Home Fires Using YOLO," *JIEEE Access*, Vol.13, pp.79451-79466, 2025.
- [3] S.B.Park, Y.J.Jeong, D.E.Lee, T.K.Kim, "A Study on the Elevator System Using Real-time Object Detection Technology YOLOv5," *Journal of Internet of Things Convergence*, Vol.10, No.2, pp.103-108, 2024.
- [4] E.S.Oh, S.R.Gwon, J.M.Oh, B.Peng, T.K.Kim, "Implementation of a real-time public transportation monitoring system," *Journal of Internet of Things Convergence*, Vol.10, No.4, pp.9-19, 2024.
- [5] T.Kim, T.K.Kim, "Systemic Analysis of the QS International Research Network Indicator Using Big Data: Regional Inequalities and Recommendations for Improved University Rankings," *JIEEE Access*, Vol.13, pp.111335-111353, 2025.
- [6] T.Kim, T.K.Kim, "Global Inequality in Research: A Quantitative Analysis of 1628 Institutions in THE World University Rankings 2025," *JIEEE Access*, Vol.13, pp.161262-161278, 2025.
- [7] T.K.Kim, "Technology Trends of Indoor Localization Scheme," *Journal of Internet of Things Convergence*, Vol.11, No.2, pp.173-177, 2025.
- [8] T.K.Kim, "Mobile Content Delivery Scheme Based on User Characteristics Categorization," *Advanced Science Letters*, Vol.23, No.10, pp.9917-9920, 2017.
- [9] T.K.Kim, E.J.Kim, "View pattern-based adaptive streaming strategy for mobile content delivery services," *Multimedia Tools and Applications*, Vol.75, No.20, pp.12693-12704, 2016.
- [10] W.X.Zhao, et al, "A survey of large language models," arXiv preprint arXiv:2303.18223, Vol.1, No.2, pp.1-124, 2023.
- [11] S.Minaee, et al, "Large language models: A survey," arXiv preprint arXiv:2402.06196, 2024.
- [12] Z.Chu, et al, "LLM agents for education: Advances and applications," arXiv preprint arXiv:2503.11733, 2025.
- [13] J.D.Kim, et al, "Development of an LLM-based CPX practicing chatbot for Korean Medicine education: implementation of automated scoring and feedback generation framework," *Journal of Korean Medicine*, Vol.45, No.4, pp.215-230, 2024.
- [14] Z.Han, et al, "Parameter-efficient fine-tuning for large models: A comprehensive survey," arXiv preprint arXiv:2403.14608, 2024.
- [15] L.Wang, et al, "Parameter-efficient fine-tuning in large language models: a survey of methodologies," *Artificial Intelligence Review*, Vol.58, No.8, pp.1-35, 2025.
- [16] E.J.Hu, et al, "Lora: Low-rank adaptation of large language models," *ICLR*, Vol.1, No.2, 2022.

이 태 오(Tae-O Lee)

[준회원]



- 2020년 2월 ~ 2025년 9월 :
국립부경대학교 컴퓨터공학과
(공학사)
- 2025년 9월 ~ 현재 : 국립부경대
학교 데이터공학과 석사과정

<관심분야>

머신러닝, 딥러닝

김 태 국(Tae-Kook Kim)

[종신회원]



- 2004년 8월 : 고려대학교
전기전자전파공학부(공학사)
- 2006년 8월 : 고려대학교
메카트로닉스학과(공학석사)
- 2014년 8월 : 고려대학교
모바일솔루션학과(공학박사)
- 2016년 3월 ~ 2022년 2월 :
동명대학교 AI학부 교수
- 2022년 3월 ~ 현재 : 국립부경대학교 컴퓨터·인공지능
공학부 교수

<관심분야>

사물인터넷(IoT), 콘텐츠 전송 네트워크(CDN), 이동성, 인
공지능(AI), 빅데이터(big data), 모바일 서비스