

Groq API를 이용한 경제 챗봇 구현

이태오¹, 김태국^{2*}

¹국립부경대학교 데이터공학과 학생, ²국립부경대학교 컴퓨터·인공지능공학부 교수

Implementing a Finance Chatbot using Groq API

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 기존의 대규모 언어 모델(Large Language Models, LLMs) 기반 금융 챗봇은 고정된 파라미터 내의 지식에 의존하므로, 변동성이 큰 실시간 시장 데이터를 반영하는 데 한계가 있다. 또한, 외부 정보를 통합하는 기존 시스템들은 질의의 특성과 무관하게 검색 및 도구 호출을 수행함으로써 불필요한 계산 비용을 발생시키고 응답 지연을 초래하는 경향이 있다. 본 연구에서는 이러한 효율성 문제를 해결하기 위해 사용자 질의의 의도를 파악하여 외부 데이터 접근 여부를 동적으로 결정하는 '조건부 도구 증강 생성(Conditional Tool-Augmented Generation)' 아키텍처를 제안한다. 제안된 시스템은 Groq Cloud의 LPU(Language Processing Unit) 환경에서 LLaMA 3.1 70B 모델을 기반으로 구축되었다. 시스템은 실시간 금융 정보가 필요한 질의에 대해서는 Yahoo Finance API를 호출하여 최신 주가 및 시계열 데이터를 확보하며, 일반적인 경제 개념 및 원리에 대한 질의는 퓨샷 프롬프팅(Few-shot Prompting)을 통해 모델 내부 지식으로 처리한다. 아울러 대화의 연속성과 맥락 유지를 위해 대화 메모리 모듈을 통합하였다. 실험 결과, 제안한 조건부 호출 전략은 정보의 최신성을 보장함과 동시에 무분별한 외부 API 호출을 최적화하여 전체적인 응답 성능과 비용 효율성을 크게 향상시켰다.

주제어 : 대규모 언어 모델, Groq, 경제 챗봇, LLaMA, 주식 데이터, Yahoo Finance, 사물인터넷

Abstract Conventional financial chatbots based on Large Language Models (LLMs) face limitations in reflecting volatile real-time market data due to their reliance on static, pre-trained parameters. Furthermore, existing systems that integrate external information often perform searches and tool invocations regardless of the specific nature of the query, leading to unnecessary computational overhead and increased response latency. To address these efficiency challenges, this study proposes a "Conditional Tool-Augmented Generation" architecture that dynamically determines the necessity of external data access by identifying user query intent. The proposed system was implemented using the LLaMA 3.1 70B model within the Groq Cloud's LPU (Language Processing Unit) environment. For queries requiring real-time financial information, the system retrieves the latest stock prices and historical time-series data via the Yahoo Finance API; meanwhile, queries regarding general economic concepts are processed through the model's internal knowledge using few-shot prompting techniques. Additionally, a dialogue memory module is integrated to maintain contextual consistency throughout the conversation. Experimental results demonstrate that the proposed conditional invocation strategy ensures the freshness of information while optimizing indiscriminate external API calls, thereby significantly enhancing overall response performance and cost-efficiency.

Key Words : Large Language Mode (LLM), Groq, Economic Chatbot, LLaMA, Stock Data, Yahoo Finance, Internet of Things (IoT)

1. 서론

최근 인공지능 기술의 고도화는 객체 인식, 상황 감지, 대규모 데이터 분석 등 다양한 응용 분야에서 실질적인 성과를 창출하고 있다. 특히 딥러닝 기반 시각 인식 기술은 화재 조기 탐지와 같은 안전 분야에서 활용 가능성을 입증하고 있으며[1], 동시에 연구 성과와 학술 영향력을 정량적으로 분석하려는 시도 역시 빅데이터와 인공지능 기법을 중심으로 확장되고 있다[2,3]. 이러한 흐름은 IoT 환경과 실내 위치 인식 기술을 포함한 지능형 시스템 전반으로 이어지며[4], 인공지능 기반 서비스 설계와 응용 연구의 필요성을 더욱 부각시키고 있다.

이러한 인공지능 응용 기술 중, 대화형 에이전트는 자연어 이해·생성 기술의 고도화와 함께 금융, 교육, 헬스케어, 공공 서비스 등 다양한 영역에서 핵심 사용자 인터페이스로 확산되고 있다. 특히 웹 기반 애플리케이션 환경에서 대화형 UI는 사용자의 질의 의도를 빠르게 수집하고, 복잡한 정보 탐색 과정을 단순화하며, 개인화된 상호작용을 제공할 수 있다는 점에서 실무적 가치가 크다. 이러한 흐름을 체계적으로 정리한 연구들은 대화형 에이전트가 “정보 접근성 향상”과 “의사결정 보조”라는 두 축에서 지속적으로 확장되고 있음을 보여준다[5].

금융 영역에서의 대화형 시스템은 특히 높은 수요를 갖는다. 금융 의사결정은 정보 비대칭성과 시간 민감성이 크고, 사용자가 요구하는 정보(가격, 추세, 비교, 맥락 설명 등)가 이질적이며, 결과에 대한 신뢰와 설명가능성 요구도 높다. 국내 연구에서도 금융 서비스에서 챗봇의 품질·페르소나·개인화 수준 등이 사용자 신뢰와 이용 의도에 유의미한 영향을 미친다는 점이 보고되었다[6,7]. 다만 기존 금융 챗봇은 (i) 정형화된 FAQ 중심 응답, (ii) 제한적 시나리오 기반 대화, (iii) 실시간 데이터 연동의 제약 등으로 인해 “실시간 시장 데이터 기반 질의”와 “경제 맥락 설명형 질의”를 하나의 일관된 사용자 경험으로 통합하는 데 한계가 있었다.

한편, 대규모 언어 모델(Large Language Model, LLM)은 금융 텍스트 이해 및 요약, 감성 분석, 정보 추출 등에서 기존 기법 대비 우수한 성능과 범용성을 보이며 금융 응용 가능성을 확대하고 있다. 예를 들어 FinBERT는 금융 도메인 지식에 특화된 언어 모델을 구성하여 금융 텍스트에서의 정보 추출 및 요약 성능 향상을 보고하였다[8]. 또한 LLM을 금융 데이터/뉴스 해석 및 트레이딩 의사결정 보조에 활용하려는 연구도 증가하고 있으며, 실제로 LLM 기반 감성 신호를 활용한 거래 전략의

가능성을 실증적으로 다룬 연구가 보고된 바 있다[9].

이러한 문제의식에 기반하여 본 연구는 실시간 금융 데이터 접근과 경제 지식 기반 대화를 하나의 사용자 경험으로 통합하는 경제 챗봇을 설계·구현하는 것을 목적으로 한다. 본 논문의 학술적 기여는 다음과 같이 정리할 수 있다.

첫째, 본 연구는 실시간 주가 조회 및 비교와 같은 정량적 금융 질의와 일반적인 경제 개념 설명을 단일 챗봇 인터페이스 내에서 처리할 수 있는 통합 구조를 제안한다. 이를 통해 기존 금융 챗봇의 기능 분절 문제를 완화하고, 보다 일관된 사용자 경험을 제공한다.

둘째, 언어 모델이 질의의 특성에 따라 외부 금융 데이터 소스의 활용 여부를 판단하는 도구 기반 외부 정보 활용 구조(tool-augmented generation)를 구현함으로써, 전통적인 문서 검색 중심 RAG 방식과 차별화된 금융 응용 사례를 제시한다.

셋째, few-shot prompt engineering을 활용하여 주식과 직접적인 관련이 없는 일상 대화에서도 경제적 맥락을 자연스럽게 유도하도록 설계함으로써, 금융 챗봇을 경제 교육 및 의사결정 보조 도구로 확장할 가능성을 제시한다.

마지막으로, 본 연구는 Groq Cloud와 LLaMA 3.1 70B를 활용한 실제 구현 사례를 상세히 기술하여, 향후 LLM 기반 금융 대화형 시스템 연구를 위한 재현 가능한 참고 모델을 제공한다.

2. 관련 연구

기존 연구는 금융 도메인에서 거대 언어 모델(LLM)의 활용을 강화하기 위해 크게 (1) 금융 특화 LLM 자체를 구축하거나 (2) 외부 금융 지식/데이터를 결합해 질의응답을 고도화하는 방향으로 진행되어 왔다. 먼저 BloombergGPT는 대규모 금융 텍스트를 포함한 데이터로 모델을 학습하여 금융 도메인 과제에서의 성능을 체계적으로 보고함으로써 금융 특화 언어 모델 구축의 유효성을 제시하였다[10]. 또한 FinGPT는 금융 분야에서의 데이터 수집·정제 파이프라인과 경량 적응(LoRA 등)을 포함한 오픈소스 프레임워크를 제안하여, 금융 도메인에서 LLM을 구축·활용하기 위한 재현 가능한 개발 방법론을 제공하였다[11]. 한편, 외부 지식 결합 관점에서는 금융 보고서/문서 질의응답에서 RAG 기반 접근을 평가·비교하여, 검색 품질 및 증거 제시가 답변 정확도와

신뢰도에 중요한 영향을 미친다는 점을 실증적으로 논의한 연구도 보고되었다[12]. 선행 연구는 금융 특화 모델링과 외부 정보 결합을 통해 정확도·최신성·신뢰도를 개선하려는 흐름으로 발전해 왔으며, 본 연구는 이러한 맥락에서 실시간 금융 데이터(API)를 질의 유형에 따라 조건적으로 호출하는 도구 기반 결합 방식을 구현 관점에서 제시한다. 본 연구는 LLM 기반 구조를 채택하고 있으므로, 해당 기술의 이론적 배경과 발전 과정을 간략히 정리한다.

2.1 대규모 언어 모델(LLM, Large Language Model)

LLM은 대량의 텍스트 데이터를 학습하여 인간 언어를 이해하고 생성할 수 있도록 설계된 인공지능 모델이다. 이러한 모델은 수십억 개의 매개변수를 사용하여 문맥을 분석하고 자연어 처리 작업을 수행한다. LLM은 텍스트 생성, 번역, 요약, 질의응답, 코딩 지원 등 다양한 응용 분야에서 활용된다. 대표적인 LLM으로는 OpenAI의 GPT 시리즈, Google의 PaLM, Meta의 LLaMA 등이 있으며, 대부분 트랜스포머 구조를 기반으로 작동한다. LLM은 많은 양의 데이터와 컴퓨팅 자원을 필요로 하지만, 이를 통해 높은 수준의 언어 이해 및 생성 능력을 보여준다.

2.2 트랜스포머 구조

트랜스포머는 자연어 처리(NLP)와 기타 인공지능 작업에 널리 사용되는 딥러닝 아키텍처이다. 2017년 Vaswani 등[13]의 논문 "Attention Is All You Need"에서 처음 소개되었으며, 기존의 순환 신경망(RNN)과 컨볼루션 신경망(CNN)을 대체할 강력한 모델로 자리 잡았다. 트랜스포머는 어텐션 메커니즘, 특히 셀프 어텐션(self-attention)을 활용하여 입력 데이터의 모든 요소 간 관계를 동적으로 모델링한다. 이를 통해 문맥 정보를 효과적으로 포착할 수 있으며, 시퀀스 길이에 따른 순차 계산이 요구되지 않아 병렬 처리가 가능하다는 장점을 가진다. 이러한 특성으로 인해 기존 RNN 기반 모델에 비해 학습 속도와 성능 측면에서 우수한 결과를 보인다[14,15].

구조적으로 트랜스포머는 인코더와 디코더로 구성된다. 인코더는 입력 시퀀스를 임베딩한 후, 다중 헤드 셀프 어텐션과 위치별 피드 포워드 네트워크를 반복적으로 적용하여 입력 데이터의 중요한 특징을 추출한다. 디코더는 인코더의 출력과 이전 단계의 생성 결과를 바탕으로

로 다음 토큰을 예측함으로써 최종 출력 시퀀스를 생성한다. 이러한 인코더-디코더 구조는 기계 번역뿐만 아니라 질의응답, 요약, 대화 시스템 등 다양한 NLP 응용 분야에서 높은 성능을 입증하였다.

2.3 Groq Cloud

Groq은 AI 추론을 위한 고성능 하드웨어와 소프트웨어 플랫폼을 제공하는 기업으로, 특히 자체 개발한 언어 처리 장치(LPU)를 통해 빠르고 효율적인 AI 모델 실행을 지원한다. Groq Cloud는 이러한 기술을 클라우드 환경에서 활용할 수 있도록 하는 서비스로, 개발자와 기업이 대규모 언어 모델(LLM)을 비롯한 다양한 AI 애플리케이션을 손쉽게 배포하고 관리할 수 있게 한다. 특히, LLaMA 3.1, Mixtral, Gemma 등 다양한 오픈 소스 LLM을 지원하며, 초당 수백 토큰의 속도를 처리할 수 있는 성능을 제공한다. 본 연구에서는 LLaMA 3.1 70B 모델을 활용한다.

2.4 LLaMA 3.1

LLaMA 3.1은 Meta(구 Facebook)가 2024년 7월 23일에 공개한 최신 대규모 언어 모델로, 이전 버전인 LLaMA 2에 비해 성능과 기능이 크게 향상되었다. 8B, 70B, 405B 세 가지 크기의 모델로 제공되어, 다양한 용도와 환경에 맞게 선택할 수 있다. 본 논문에서는 70B를 사용했다. LLaMA 3.1의 다른 특징으로 한 번에 입력될 수 있는 토큰의 수가 128,000개로 확장되었다. 이로 인해 긴 문서나 복잡한 대화의 맥락을 효과적으로 처리할 수 있다. 또, LLaMA 3.1은 다국어 지원을 지원하여 한국어를 별도로 학습시키지 않아도 한국어 환경에서 사용할 수 있다.

2.5 Yahoo Finance

Yahoo Finance는 금융 뉴스, 데이터, 도구 및 분석을 제공하는 세계적으로 인기 있는 플랫폼으로, 개인 투자자부터 금융 전문가까지 다양한 사용자들에게 폭넓은 서비스를 제공한다. Yahoo의 포털 서비스의 일부로 시작되었으며, 현재는 실시간 시장 데이터, 금융 뉴스, 투자 관리 도구 등을 통해 금융 정보 접근성을 높이는데 기여하고 있다. 본 논문에서 기술하는 챗봇은 Yahoo Finance에서 실시간 주식 정보를 가져와 그것을 활용한다.

2.6 검색 증강 생성(Retrieval-Augmented Generation, RAG)

검색 증강 생성(Retrieval-Augmented Generation, 이하 RAG)은 대규모 언어 모델(LLM)이 외부 지식 베이스나 실시간 데이터 소스를 함께 활용함으로써, 보다 정확하고 신뢰도 높은 응답을 생성하도록 설계된 기법이다. 기존의 언어 모델은 학습 시점에 고정된 파라미터에 기반하여 응답을 생성하기 때문에, 최신 정보 반영이 어렵고 외부 데이터에 대한 직접적인 접근에 한계를 가진다. 이러한 한계를 보완하기 위해 RAG는 질의와 관련된 정보를 외부 데이터 소스에서 검색한 후, 해당 정보를 언어 모델의 입력으로 결합하여 응답을 생성하는 방식을 사용한다.

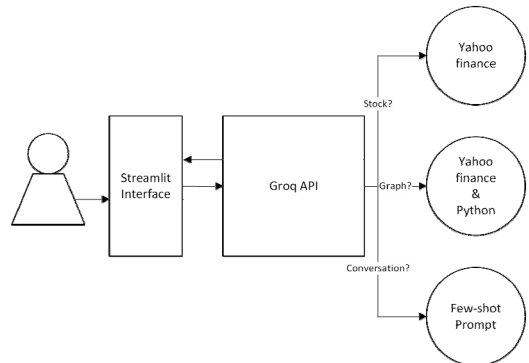
기술적으로, RAG는 언어 모델의 출력이 고정된 파라미터에만 의존하지 않고, 입력 질의에 따라 외부 지식 소스로부터 검색된 문서 집합에 조건화되어 생성되도록 설계된 프레임워크이다. 이를 수식적으로 표현하면, RAG는 사용자 질의 q 에 대한 관련 문서 집합 $D = d_1, d_2, \dots, d_k$ 를 검색한 후, 해당 문서들을 조건으로 하여 응답 y 를 생성하는 확률 분포 $p(y|q, D)$ 를 모델링하는 방식으로 정의될 수 있다. 이때 검색 단계는 언어 모델 외부에서 수행되며, 생성 단계는 검색 결과를 입력받아 조건부 텍스트 생성을 수행한다. 이러한 구조를 통해 RAG는 언어 모델의 파라미터 업데이트 없이도 최신 정보 반영과 도메인 특화 지식 활용이 가능하다는 특징을 가진다.

그러나, 본 연구에서 구현한 경제 챗봇은 전통적인 문서 검색 기반 RAG 방식을 직접적으로 적용한 구조는 아니며, 언어 모델이 질의의 특성에 따라 외부 정보 소스의 활용 여부를 조건적으로 결정하는 도구 기반 외부 정보 활용 구조(tool-augmented generation)로 설계된다. 구체적으로, 시스템은 사용자 질의가 실시간 금융 데이터와 같이 언어 모델의 내부 지식만으로는 충분히 처리하기 어려운 정보의 요청 여부를 먼저 판단한다. 이러한 판단은 언어 모델이 사전에 정의된 외부 함수 호출 가능성을 고려하여 수행되며, 외부 데이터 접근이 필요하다고 판단된 경우, 주식 가격 조회나 기간별 가격 비교와 같은 기능을 수행하는 함수가 호출되어 실시간 금융 데이터가 획득된다. 이후 획득된 외부 데이터는 언어 모델의 응답 생성 과정에 반영된다. 반면, 외부 정보 접근이 필요하지 않은 질의에 대해서는 추가적인 데이터 조회 없이 언어 모델이 자체적으로 경제 개념 설명, 맥락 기반 해석, 일반적인 경제 질의응답을 수행하도록 구성된다.

3. 구현 방법

본 장에서는 경제 챗봇을 구현하는 방법에 대해 논의한다. 본 연구에서 설계한 경제 챗봇은 두 가지 상황에 따라 차별화된 응답을 제공하도록 설계되었다. 첫째, 주식과 관련이 없는 일반적인 대화의 경우, 사용자의 경제적 관심을 유도하는 흥미로운 대화를 생성한다. 둘째, 특정 주식의 이름이 언급되거나, 주식명과 함께 특정 기간 정보 및 비교 요청과 관련된 대화가 감지될 경우, 외부 실시간 데이터를 활용하여 주식의 현재 시가를 제공하거나 서로 다른 주식을 비교하는 시각적 그래프를 제시한다. 이를 통해 사용자의 요구를 반영한 정교한 정보 제공과 대화 경험을 제공할 수 있다.

[Fig. 1]은 구현한 챗봇의 대략적인 구조를 나타낸다. 사용자는 Streamlit으로 만들어진 사용자 인터페이스를 통해 질의를 요청할 수 있다. 요청된 질의는 Groq API를 통해 사용하는 언어 모델인 LLaMA 3.1 70B로 전달되어 질의에 맞는 응답을 생성한다. 이때, 언어 모델은 질의의 의도를 파악하여 그것이 (1) 단순히 현재 주가를 묻는 것인지, (2) 두 개의 주식의 일정 기간 비교 그래프를 그려달라는 것인지, (3) 또는 주식과 관련이 없는 대화인지를 판단하여 응답을 생성한다. 생성된 응답은 Groq API를 통해 사용자 인터페이스에 보여지게 된다.



[Fig. 1] The structure of chatbot

3.1 도구 사용 여부 판단

사용자가 챗봇과 대화를 시도할 때, 챗봇은 먼저 대화의 의도를 파악해야 한다. 사용자의 요청이 특정 주식의 실시간 주가를 묻는 것인지, 혹은 일정 기간 동안 여러 주식의 주가를 비교해 달라는 것인지 분류하는 과정이 필요하다. 이를 구현하기 위해 두 개의 언어 모델을 사용한다. 첫 번째 언어 모델은 사용자와의 대화를 담당하는

챗봇 역할을 수행하며, 두 번째 모델은 대화의 의도를 판단하는 데 초점을 맞춘다.

판단을 담당하는 언어 모델은 대화 내용을 분석하여 도구의 사용 여부를 결정한다. 사용자가 요청한 내용을 기반으로, 특정 주식의 실시간 주가를 가져오는 함수를 호출하거나, 일정 기간 동안 주가 데이터를 비교하고 시각화하는 함수를 선택해 실행한다. 이러한 구조는 사용자의 다양한 요청에 효율적으로 대응하며, 대화형 인터페이스와 데이터 기반 분석 간의 유기적 연계를 가능하게 한다.

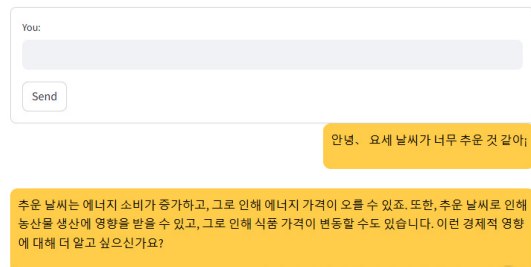
3.2 Few-shot prompt engineering

챗봇이 도구를 사용하지 않아도 된다고 판단하는 경우, 대화는 해당 기법을 활용하여 진행된다. Few-shot Prompt Engineering은 언어 모델에 입력값을 제공하기 전에, 모델이 수행해야 할 역할이나 대화의 방향을 미리 정의하여 원하는 방식으로 작동하도록 유도하는 기법이다. 이 방법을 통해 모델은 특정 맥락과 의도에 맞는 응답을 생성할 수 있도록 사전에 정보를 제공받는다.

이 기법을 활용하면 챗봇은 주식과 직접 관련이 없는 일상 대화에서도 경제와 관련된 주제를 자연스럽게 다룰 수 있다. 이를 통해 사용자가 경제에 대해 흥미를 느끼도록 유도하며, 대화의 유용성과 몰입도를 동시에 높인다. [Fig. 2]가 그 예시를 보여주는데, “안녕, 요새 날씨가 너무 추운 것 같아!” (Hi, It’s really cold lately!“)라고 질문을 하면, ”추운 날씨는 에너지 소비가 증가하고, 그로 인해 에너지 가격이 오를 수 있죠, 또한, 추운 날씨로 인해 농산물 생산에 영향을 받을 수 있고, 그로 인해 식품 가격이 변동할 수도 있습니다. 이런 경제적 영향에 대해 더 알고 싶으신가요?“ (”Cold weather increases energy consumption, which can lead to higher energy prices. In addition, cold conditions can

경제 심심이

안녕하세요! 경제 관련 질문을 해 주세요. 종료하려면 'QUIT'을 입력하세요.



[Fig. 2] Few-shot prompt engineering

affect agricultural production, potentially causing fluctuations in food prices. Would you like to learn more about these economic impacts?“)라고 대답하는 것을 볼 수 있다.

3.3 실시간 주가 데이터 응답

챗봇이 사용자의 대화에서 실시간 주가 데이터를 요구하는 요청을 감지하면, Yahoo Finance에서 데이터를 가져오는 함수를 호출하여 해당 정보를 제공한다. 이 과정에서 챗봇은 사용자 대화의 맥락을 이해하고, 주식명 또는 심볼(ticker)을 정확히 식별하여 검색한다. 이를 위해 자연어 처리 기법과 사전 정의된 주식명 데이터베이스를 활용하여 사용자가 입력한 표현을 표준화된 형식으로 변환한다.

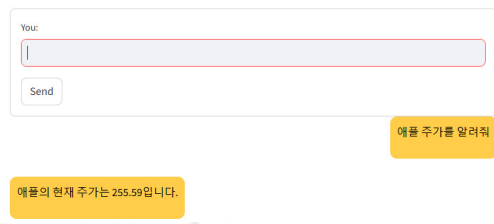
본 연구에서는 국내 주식이 아닌 미국 주식을 실험 대상으로 사용하였다. 이는 특정 국가의 금융 시장 특성을 분석하기 위함이 아니라, 실시간 금융 데이터 연동과 대규모 언어 모델 기반 대화형 시스템의 동작 구조를 검증하는 데 연구의 목적이 있기 때문이다. 미국 주식 시장은 Yahoo Finance와 같은 공개 금융 데이터 소스를 통해 종목 식별자, 가격 정보, 기간별 데이터가 안정적으로 제공되어 시스템 구현의 재현성과 일반화를 확보하는 데 유리하다.

예를 들어, 사용자가 “애플 주가를 알려줘.”와 같은 요청을 하면, 챗봇은 “애플”이 Apple Inc.의 주식명임을 인지하고, 해당 주식의 심볼인 “AAPL”을 추출한다. 그런 다음, Yahoo Finance API를 통해 실시간 주가 정보를 검색하여 사용자에게 반환한다.

이 과정은 대화형 인터페이스의 유연성을 유지하면서도, 정확하고 신속한 정보 제공을 가능하게 한다. 이를 통해 단순히 정보를 전달하는 것을 넘어, 금융 데이터 활용에 대한 실질적인 가치를 제공한다.

경제 심심이

안녕하세요! 경제 관련 질문을 해 주세요. 종료하려면 'QUIT'을 입력하세요.



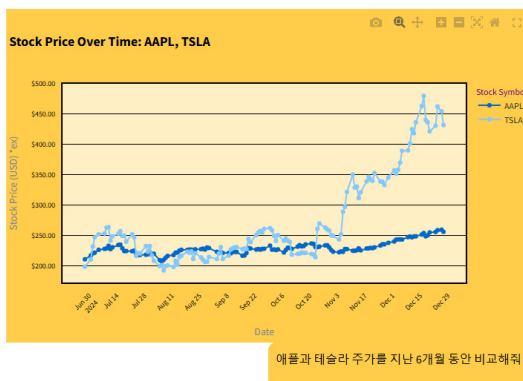
[Fig. 3] Real-time stock price response

3.4 주가 비교 그래프 작성

챗봇이 사용자의 대화에서 특정 주식들의 이름과 해당 주식들의 주가를 특정 기간 동안 비교하라는 요청을 감지하면, Yahoo Finance에서 해당 기간의 주가 데이터를 가져와 이를 시각화하는 함수를 호출한다. 이 과정에서 챗봇은 대화 내 텍스트에서 여러 회사 이름을 정확히 인식하고 이를 표준화된 주식 심볼로 변환하며, 동시에 사용자가 지정한 기간을 추출하여 데이터 검색에 활용한다.

예를 들어, [Fig. 4]에서는 사용자가 “애플과 테슬라 주가를 지난 6개월 동안 비교해 줘.”(“Compare Apple and Tesla stock prices over the past six months.”)라고 요청하면, 챗봇은 “애플”과 “테슬라”를 각각 Apple Inc.와 Tesla Inc.의 주식명으로 매핑하고, 그 심볼인 “AAPL”과 “TSLA”를 식별한다. 이어서 “지난 6개월”이라는 기간을 구체적인 날짜 범위로 변환한 뒤, Yahoo Finance API를 통해 해당 주식들의 일일 종가를 수집한다. 수집된 데이터는 데이터 처리 및 시각화 함수를 통해 그래프로 변환된다. 결과적으로, 챗봇은 사용자에게 직관적이고 이해하기 쉬운 그래프 형태로 데이터를 제공하며, 이를 통해 단순한 데이터 응답을 넘어 정보의 시각적 전달력을 극대화한다.

이 기능은 특히 다중 주식 분석 및 비교가 필요한 사용자에게 유용하며, 금융 데이터 분석에 대한 사용자 경험을 강화하는 데 기여할 수 있다. 또, 챗봇의 자연어 처리 능력과 데이터 시각화 기술의 결합은 복잡한 요청을 처리하고 사용자 만족도를 높이는 데 중요한 역할을 한다.



애플과 테슬라의 지난 6개월간 주가 비교 차트를 생성했습니다. 차트를 참고하여 두 회사의 주가를 비교해 보실 수 있습니다.

[Fig. 4] Comparing stock prices by showing Graph

4. 결론

본 연구에서는 Groq Cloud와 최신 언어 모델인 LLaMA 3.1을 활용하여 경제 챗봇을 설계하고 구현한 사례를 제시한다. 연구의 주요 목표는 금융 데이터와 경제 정보를 사용자에게 신속하고 효율적으로 제공하며, 사용자의 경제적 의사결정을 지원할 수 있는 실용적이고 사용자 친화적인 AI 시스템을 개발하는 것이었다. 이를 위해 챗봇은 두 가지 핵심 기능을 제공한다.

첫째, Yahoo Finance API를 활용하여 특정 주식의 실시간 가격을 조회하거나, 여러 주식의 기간별 가격 데이터를 비교하여 시각적으로 제시하는 기능을 구현한다. 이를 통해 사용자들이 금융 시장 데이터를 보다 직관적으로 이해하고 활용할 수 있도록 지원한다. 둘째, 일반적인 대화 요청에도 경제 관련 정보를 자연스럽게 녹여내어 사용자의 경제적 관심을 유도하는 기능을 제공한다. Few-shot Prompt Engineering 기법을 활용하여 챗봇의 대화 방향성을 설정하고, 사용자와의 상호작용을 통해 경제 지식 격차를 줄이고자 하였다.

챗봇의 설계 과정에서 사용된 도구 사용 여부 판단 시스템은 대화형 AI의 유연성과 기능적 효율성을 동시에 확보할 수 있음을 보여준다. 특히, 사용자 요청의 맥락을 분석하여 적절한 함수를 호출하거나 대화를 진행하는 구조는 다양한 응용 가능성을 열어준다. 또한, 실시간 데이터 접근성과 직관적인 시각화 기능은 사용자 경험을 대폭 향상시킨다.

본 연구는 경제 챗봇의 구현 사례를 넘어, 대규모 언어 모델이 실시간 금융 데이터 접근과 경제적 추론을 하나의 대화형 시스템으로 통합할 수 있음을 실증적으로 보여준다. 제안한 챗봇 구조는 사용자의 질의 맥락에 따라 외부 데이터 호출 여부를 동적으로 결정함으로써, 금융 도메인에서 도구 기반 외부 정보 활용 구조의 실질적 적용 가능성을 확인하였다. 학술적 관점에서 본 연구는 최신 대규모 언어 모델 기술을 실제 금융 챗봇 설계에 적용한 응용 연구로서, 이론 중심 연구와 실무 시스템 개발 사이의 거리를 연결하는 데 기여한다.

REFERENCES

- [1] B.Peng and T.K.Kim, "YOLO-HF: Early Detection of Home Fires Using YOLO," IEEE Access, Vol.13, pp.79451-79466, 2025.

- [2] T.Kim and T.K. Kim, "Systemic Analysis of the QS International Research Network Indicator Using Big Data: Regional Inequalities and Recommendations for Improved University Rankings," IEEE Access, Vol.13, pp.111335-111353, 2025.
- [3] T.Kim and T.K.Kim, "Global Inequality in Research: A Quantitative Analysis of 1628 Institutions in THE World University Rankings 2025," IEEE Access, Vol.13, pp.161262-161278, 2025.
- [4] T.K.Kim, "Technology Trends of Indoor Localization Scheme," Journal of Internet of Things Convergence, Vol.11, No.2, pp.173-177, 2025.
- [5] M.Allouch, A.Azaria and R.Azoulay, "Conversational Agents: Goals, Technologies, Vision and Challenges," MDPI Sensors, Vol.21, No.24, pp.8448, 2021.
- [6] S.H.Byun and C.H.Cho, "The Effect of the Anthropomorphism Level and Personalization Level on AI Financial Chatbot Recommendation Messages on Customer Response," The Korean Journal of Advertising and Public Relations, Vol.22, No.2, pp.4466-502, 2020.
- [7] E.B.Han, S.H.Han and J.H.Choi, "The Impacts of Comparison Options and Explanation Styles on Consumer Experience in AI Financial Chatbot Recommendations," The e-Business Studies, Vol.26 No.1, pp.3-23, 2025.
- [8] A.H.Huang, H.Wang and Y.Yang, "FinBERT: A Large Language Model for Extracting Information from Financial Text," Contemporary Accounting Research, Vol.40, No.2, pp.806-841, 2023.
- [9] K.Kirtac and G.Germano, "Sentiment trading with large language models," Finance Research Letters, 62, pp.105227, 2024
- [10] S.Wu, O.Irsoy, S.Lu, V.Dabrovolski, M.Dredze, S.Gehrmann, P.Kambadur, D.Rosenberg and G.Mann, "BloombergGPT: A Large Language Model for Finance," arXiv:2303.17564, 2024.
- [11] H.Yang, X.Y.Liu and C.D.Wang, "FinGPT: Open-Source Financial Large Language Models," arXiv:2306.06031, 2023.
- [12] I.Iaroshev, R.Pillai, L.Vaglietti and T.Hanne, "Evaluating Retrieval-Augmented Generation Models for Financial Report Question and Answering," Applied Sciences Vol.14, No.20, pp.9318, 2024.
- [13] A.Vaswani, N.Shazeer, N.Parmar, J.Uzkoreit, L.Jones, A.N.Gomez, L.Kaiser and I.Polosukhin, "Attention is all you need," Advances in neural information processing systems, 30, 2017.
- [14] J.Devlin, M.Chang, K.Lee and K.Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, Vol.1, 2019.
- [15] T.Wolf, L.Debut, V.Sanh, J.Chaumond, C.Delangue, A.Moi, P.Cistac, T.Rault, R.Louf, M.Funtowicz, J.Davison, S.Shleifer, P.von Platen, C.Ma, Y.Jernite, J.Plu, C.Xu, T.Le Scao, S.Gugger, M.Drame, Q.Lhoest and A.Rush, "Transformers: State-of-the-art natural language processing," Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations, pp.38-45, 2020.

이 태 오(Tae-O Lee)

[준회원]



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

〈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·인공지능공학부 교수

〈관심분야〉

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