

## Can MLLMs See the Market? Evaluating Pattern Reasoning in Real-World Financial Charts

Eun Hong Park\*, Ji Hoon Park\*, Ha Young Kim\*\*

\*Student, Graduate School of Information, Yonsei University, Seoul, Korea

\*\*Associate Professor, Graduate School of Information, Yonsei University, Seoul, Korea

### [Abstract]

In this paper, we propose an approach to evaluate the capability of Multimodal Large Language Models (MLLMs) to recognize and classify visual patterns in stock charts containing real-world visual noise. We apply three prompting strategies—Self-Consistency (SC), In-Context Learning (ICL), and Chain-of-Thought (CoT)—to assess their impact on classification accuracy, and introduce visual preprocessing using the Segment Anything Model (SAM) to analyze the effect of noise reduction. Through experiments on multiple MLLMs, we compare the individual and combined effects of these strategies. The results show that the SC prompting strategy and SAM preprocessing significantly enhance performance. This study serves as an early empirical investigation into the practical applicability of MLLMs for visual financial analysis, providing quantitative insights into how prompt design and visual preprocessing influence model performance.

▶ **Key words:** Multimodal, Financial Chart, Prompt Engineering, Self-Consistency, Chain-of-Thought, In-Context Learning

### [요 약]

본 논문에서는 MLLMs가 시각적 노이즈가 포함된 주가 차트에서 패턴을 인식하고 분류할 수 있는지를 검증하기 위해 다양한 프롬프트 전략과 시각 전처리 기법을 적용하고 그 효과를 비교·분석하였다. Self-Consistency(SC), In-Context Learning(ICL), Chain-of-Thought(CoT) 전략이 분류 정확도에 미치는 영향을 평가하고, Segment Anything Model(SAM)을 활용한 노이즈 감소 효과를 분석하였다. 여러 MLLMs를 대상으로 각 전략의 개별 및 조합 효과를 실험한 결과, SC 전략과 SAM 전처리가 성능 향상에 유의미한 영향을 미치는 것으로 나타났다. 본 연구는 MLLMs 기반 시각 패턴 분석의 실제 적용 가능성을 탐색한 초기 실증 연구로서, 프롬프트 설계와 시각 정보 전처리가 모델 성능에 미치는 구조적 영향을 정량적으로 분석했다는 점에서 의의를 지닌다.

▶ **주제어:** 다중모달리티, 금융 차트 분석, 프롬프트 엔지니어링, 자기 일관성, 연쇄 추론, 문맥 내 학습

- First Author: Eun Hong Park, Ji Hoon Park, Corresponding Author: Ha Young Kim
- \*Eun Hong Park (pehhh03@yonsei.ac.kr), Graduate School of Information, Yonsei University
- \*Ji Hoon Park (zzang9jihoon@yonsei.ac.kr), Graduate School of Information, Yonsei University
- \*\*Ha Young Kim (hayoung.kim@yonsei.ac.kr), Graduate School of Information, Yonsei University
- Received: 2025. 06. 02, Revised: 2025. 06. 16, Accepted: 2025. 06. 23.

## I. Introduction

금융 시장은 경제 지표, 기업 보고서, 글로벌 뉴스와 같은 텍스트뿐 아니라, 주가 차트, 기술 지표, 시계열 데이터 등 시각적·정량적 정보가 다양한 형태의 모달리티로 공존하는 복합적인 환경이다[1,2]. 이질적인 정보 간의 의미 관계를 통합적으로 이해하는 능력은 투자 판단의 핵심 요소로 작용하며, 정보 간 상호작용을 종합적으로 해석하는 고차원적 역량이 요구된다.

기술적 분석은 시각적 패턴을 기반으로 주가의 미래 흐름을 예측하는 대표적인 분석 기법으로, 매수·매도 타이밍 결정, 추세 전환 포착, 리스크 관리 등 실무에서 광범위하게 활용되고 있다[3,4]. 그중에서도 헤드 앤 숄더(Head & Shoulders)와 컵 앤 핸들(Cup & Handle)은 각각 하락 반전 및 상승 지속을 나타내는 대표적인 시각 패턴으로, 정량적 지표보다 더 직관적인 투자 판단 기준이 될 수 있다[5,6]. 그러나 이러한 패턴은 수학적으로 엄밀하게 정의하기 어렵고, 주관적 해석과 정성적 판단이 요구되기 때문에 자동화에 제약이 따른다[7]. 실제 차트에는 보조 지표, 주석, 배경 격자 등 다양한 시각적 노이즈가 포함되어 있어, 단순 규칙 기반 접근이나 단일 모달리티 기반의 컴퓨터 비전 (Computer Vision) 기술만으로는 복잡한 시계열 패턴을 안정적으로 인식하기 어렵다[8,9].

한편, MLLMs (Multimodal Large Language Models)는 서로 다른 모달리티를 입력으로 받아 사용자 질의에 응답하는 방식으로 최근 다양한 모델이 연구되고 있다 [10,11,12]. 이러한 MLLMs의 장점을 활용하여 차트 이미지를 기반으로 한 시각적 질의응답 (Visual Question Answering, VQ) 과제에 적용하려는 시도도 이루어졌다 [13]. 그러나 기존 차트 기반 MLLMs는 Fig. 1의 좌측과 같이 주로 이상적인 조건에서 생성된 정형화된 차트를 중심으로 학습되어 왔다[14]. 본 연구의 사례 연구 결과, 실제 금융 환경에서 자주 등장하는 Fig. 1의 우측과 같이 비정형적이고 복잡한 형태의 차트에서는 기존 모델들이 패턴 인식과 해석 능력에서 현저히 낮은 성능을 보였다 [11,12,13]. 현실적인 차트 이미지에 대해 기존 MLLMs는 단순 수치해석뿐 아니라 시각적 패턴의 식별에도 한계를 보이는 것이 확인되었다. 이러한 패턴을 인식하지 못하는 문제는 투자 판단의 핵심 전환 신호를 포착하지 못하거나 잘못된 해석을 유발할 수 있어 실전에서 큰 제약을 초래할 수 있다. 따라서 MLLMs가 시각적 패턴을 더욱 효과적으로 이해할 수 있도록 노이즈 환경에서도 추론을 유도할 수 있는 프롬프트 설계가 필수적이다[15,16,17].

이에 본 연구에서는 MLLMs가 복잡한 금융 차트 환경에서 시각적 패턴을 인식하고 추론을 수행하기 위해 최적의 프롬프트 전략은 무엇인지 탐구하고, 그 유효성을 실증적으로 검증하는 것을 목표로 한다. 이를 위해 노이즈가 포함된 실제 금융 차트 샘플을 수집하고 SC (Self-Consistency), ICL (In-Context Learning), CoT (Chain-of-Thought) 전략을 적용하여, 각 전략이 MLLMs의 패턴 인식 정확도에 미치는 영향을 비교·분석하였다[15,16,17]. 노이즈가 포함된 실제 차트와 플롯 차트를 모두 활용하고, 시각적 추론을 극대화할 수 있는 실질적인 프롬프트 설계를 중심으로 MLLMs의 잠재력을 탐색하고자 하였다.

본 연구의 주요 기여는 다음과 같다.

1. 실제 금융 시장을 반영한 노이즈가 포함된 차트 이미지 데이터셋을 구축하여 기존의 중심 벤치마크와 차별화된 실제 기반 실험 환경을 제공하였다.
  2. 사례 연구를 통해 MLLMs가 현실 차트에서 범하는 대표적인 오류 유형과 실패 요인을 도출하고, MLLMs의 시각적 해석의 한계를 규명하였다.
  3. 시각적 패턴 정의에 기반한 프롬프트를 설계하고, ICL, CoT, SC 등 다양한 프롬프팅 전략을 적용하여 MLLMs 추론 성능 향상 가능성을 실증적으로 입증하였다.
- 본 연구는 단순한 차트 질의응답을 넘어 MLLMs가 시각적 구조와 언어 정보를 통합적으로 활용하여 주가 패턴을 식별하고 해석할 수 있는지에 대한 가능성을 검토하며 향후 MLLMs가 금융 분석에서 수행할 수 있는 역할의 범위를 확장하는 데 기여하고자 한다.

## II. Related works

### 1. Visual Pattern Analysis in Financial Chart Understanding

금융 시장의 기술적 분석을 위해 정량적 수치 외 시각적 패턴은 시장의 추세와 심리를 파악하는 도구로 오랜 기간 활용되어 왔다[18]. 헤드 앤 숄더 (Head & Shoulders) 와 컵 앤 핸들 (Cup & Handle) 등 특정 차트 패턴은 가격 반전이나 추세 지속 여부를 판단하는 데 있어 높은 실용성과 설명력을 갖추고 있어 투자자들이 매수·매도 타이밍을 결정할 때 실질적인 근거로 활용된다[5,6]. 이러한 패턴은 일정한 구조를 가지지만 표현 방식은 차트 환경과 데이터의 시간 축 구성에 따라 다양하게 나타날 수 있다. 실제로는 패턴의 크기, 기울기, 대칭성, 기간 등에서 유의미한 차이

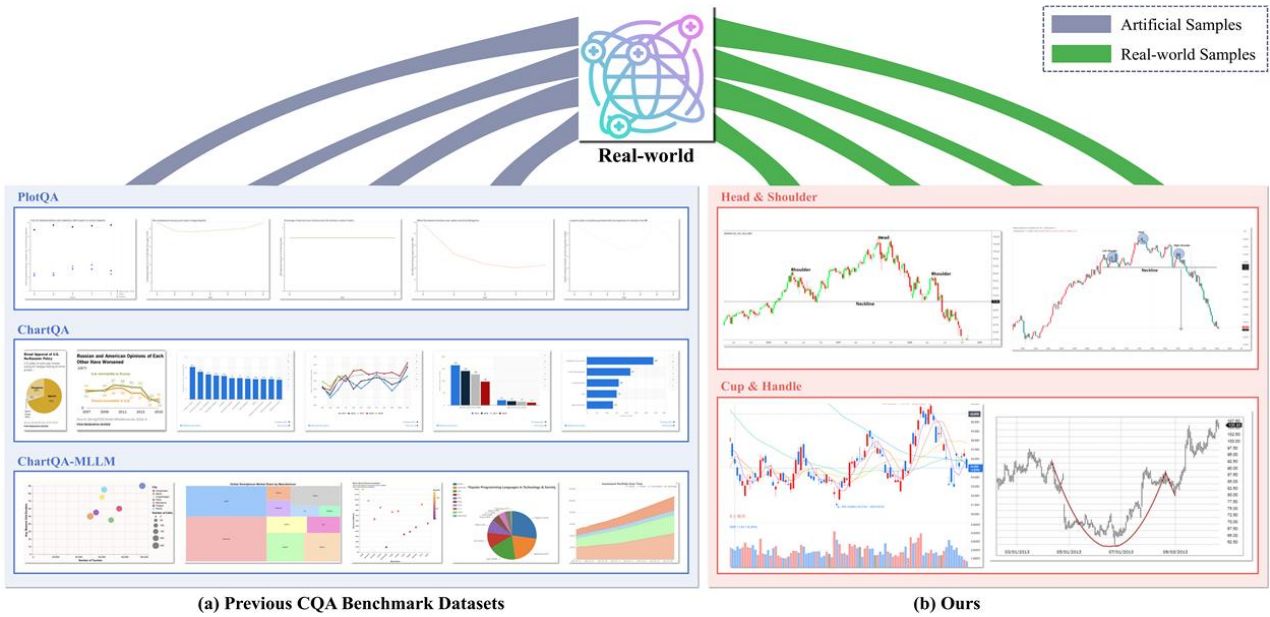


Fig. 1. Comparison between (a) existing CQA benchmark datasets and (b) our real-world financial pattern dataset. Previous datasets such as PlotQA, ChartQA primarily consist of infographic-like charts, whereas our dataset includes realistic, noisy stock charts with structural patterns.

가 존재하고 이를 정량적 기준으로 일관되게 구분하는 것에 어려움이 따른다[19]. 보조 지표, 거래량 창, 주석 등 다양한 노이즈가 포함된 실제 차트에서는 패턴의 경계가 모호해지고, 이는 규칙 기반 접근의 적용 가능성을 크게 제한한다[20].

한편, 최근 LLM 기반 금융 연구는 주로 텍스트 기반 정보를 중심으로 발전해 왔다. Jeong et al. 은 금융 문서, 뉴스, 리서치 데이터를 기반으로 LLM을 fine-tuning 하여 금융 분석 및 의사결정에 적용 할 수 있는 특화 모델을 구축하였다[21]. Che et al.은 재무제표, 산업 네트워크, 보고서 등을 포함하는 멀티모달 금융 데이터 셋을 구축하고 Attention 기반 통합 모델에 적용하여 분석 정밀도를 향상하였다[22]. 하지만 이들 연구는 대부분 텍스트 기반 데이터에 초점을 맞추고 있어, 실제 투자자들이 자주 활용하는 시각 중심의 차트 정보는 반영하지 못했다는 한계가 있다.

본 연구는 이러한 간극을 메우기 위해 실제 주식 시장에서 사용되는 노이즈 포함 차트 이미지와 구조적 패턴 정의를 통합한 멀티모달 접근법을 제시한다. 시각 정보와 함께 해당 패턴에 대한 정의와 해석 기준을 프롬프트로 제공함으로써 LLM이 패턴 존재 여부를, 프롬프팅을 통해 추론할 수 있도록 유도한다[15,16,17].

## 2. Multimodal LLMs for Financial Chart Reasoning

MLLMs는 시각적 정보와 언어 정보를 통합적으로 처리할 수 있어 다양한 도메인에서 응용되고 있으며 최근에는 차트 분석 및 수치 기반 질의응답에 활발히 사용되고 있다. 대표적으로 UniChart (Zhang et al., 2023)는 차트 이미지와 설명 텍스트를 함께 인코딩한 후 질의응답 및 수치 추론, 요약 등의 다운스트림 작업에 활용될 수 있는 범용 프레임워크를 제안했다[23]. 해당 모델은 차트 구조를 디코딩하는 차트 디렌더링(chart de-rendering)을 통해 차트의 시각 요소를 기계적으로 분해하고 이를 숫자 기반 추론으로 연결한다. ChartLlama (Han et al., 2024)는 GPT-4를 기반으로 생성된 차트 질의응답 데이터셋을 이용하여 fine-tuning된 MLLMs로 차트 요약, 주석 편집, 질의응답 등의 태스크를 수행한다[24]. 이 모델은 특히 다양한 프롬프트 시나리오에 강건하게 반응하며 차트-텍스트 변환 또는 차트 질의응답 같은 태스크에 강점을 보인다.

하지만 이러한 모델들은 시각적 패턴 인식 및 분류 문제에는 적용이 제한적이다. 본 연구는 이러한 한계를 보완하고자, MLLMs이 노이즈가 포함된 차트 환경에서 차트 패턴의 정의를 프롬프트로 명시적으로 제공하고 실제 노이즈가 포함된 차트 이미지에서 MLLMs가 시각 정보와 텍스트 지식을 조합해 시각 패턴 존재 여부를 추론하도록 설계하였다.

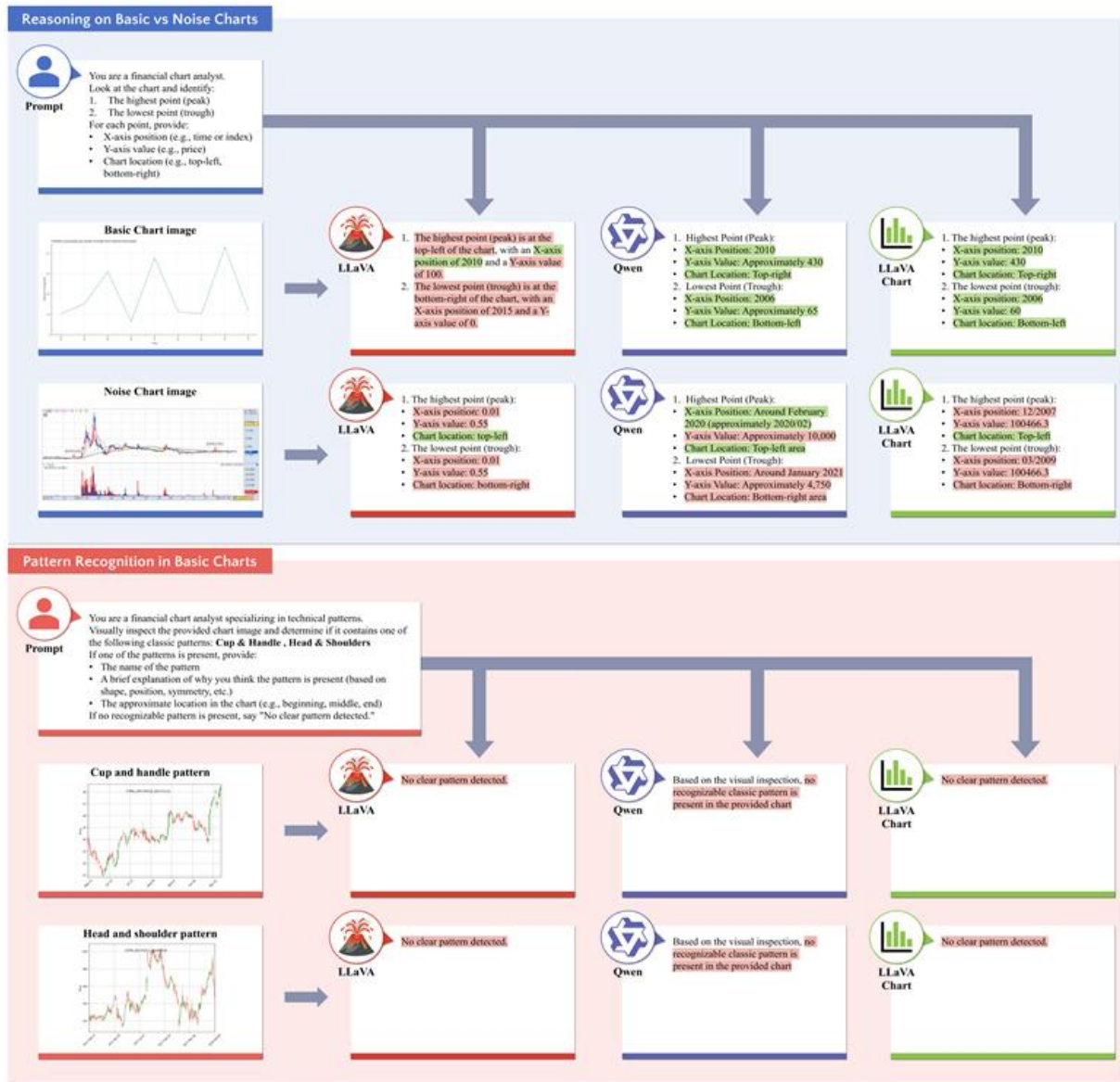


Fig. 2. Evaluation of MLLMs on Chart Reasoning (upper section) and Pattern Recognition (lower section) tasks. The basic chart image used in the reasoning task is sourced from the PlotQA dataset, while the noise-added chart image is captured from an actual HTS (Home Trading System) interface. Red highlights indicate incorrect responses, whereas green highlights denote correct answers.

### 3. Prompt Engineering for Visual Reasoning

프롬프트 엔지니어링 (Prompt Engineering) 은 LLM의 추론 능력을 효과적으로 끌어내는 핵심 전략으로, 복잡한 문제 해결 과정을 단계적으로 유도하는 방식인 CoT 기법이 추론 중심 태스크에서 우수한 성능을 보인다[16]. CoT는 모델이 최종 답을 도출하기까지의 사고 과정을 중간 단계로 언어화하도록 유도하는 방식으로, Kojima et al. 은 "Let's think step by step"이라는 단순한 문구 삽입만으로도 LLM이 수학, 논리, 추론 문제에서 성능이 향상됨을 입증하였다[25].

시각 정보가 포함된 멀티모달 환경에서도 CoT의 효과

는 점차 확장되고 있으며, Zhang et al. 은 차트 요약 및 차트 기반 질의응답 작업에서 few-shot 예시와 CoT를 결합한 프롬프트 구성을 통해 단순 수치 응답을 넘어 설명이 포함된 응답을 생성할 수 있음을 보여주었다[26].

SC 전략은 동일한 프롬프트에 대해 여러 번 추론을 반복 수행한 후, 가장 자주 등장하는 응답을 최종 예측값으로 선택하는 방식이다. 이는 복잡하고 추론 경로가 다양한 문제 상황에서 일관성과 신뢰성을 확보하는 데 유용하며, 최근에는 수학 문제 풀이, 지식 기반 질의응답 등 다양한 분야에서 활발히 활용되고 있다[17]. ICL은 프롬프트 내에 유사한 예시를 포함해, LLM이 새로운 입력을 추론할 때

이를 참조하도록 유도하는 전략으로 zero-shot 또는 few-shot 설정에서 일관되고 설명력 있는 응답을 생성하는 데 효과적이다[15]. ICL은 정형화되지 않은 패턴을 유추하는 태스크에서 우수한 성능을 보였으며, 이는 본 연구의 차트 패턴 분류와 같은 시각 추론 과제에서도 적합한 전략으로 작용할 수 있다.

본 연구는 CoT, ICL, SC 세 가지 프롬프트 전략을 차트 패턴 분류 문제에 적용하여 각 전략이 MLLMs의 추론 성능에 미치는 영향을 정량적으로 비교하였다[15,16,17]. 또한 세 가지 프롬프트 전략을 단독으로 적용하는 실험뿐만 아니라 전략 간 조합을 통한 다양한 프롬프트 구성 방식도 함께 실험함으로써, 프롬프팅 방식이 MLLMs의 차트 패턴 추론 성능에 미치는 영향을 더욱 다각적으로 분석하였다[27]. 이는 금융 영역에 시각 정보 해석과 언어 기반 추론을 결합한 실험적 접근이라는 점에서 기존 연구들과의 차별성을 갖는다.

### III. Case Study

본 연구는 MLLMs가 실제 금융 환경에서 사용되는 복잡한 차트 이미지 및 시각적 패턴을 얼마나 효과적으로 인식하고 추론할 수 있는지를 검증하기 위해 두 가지 사례 기반 실험을 설계하였다.

첫 번째 사례는 노이즈가 포함된 실제 차트 이미지에서 MLLMs의 기본적인 시각 추론 능력을 정성적으로 평가하는 데 초점을 맞추었으며, 두 번째 사례는 차트 내에 내재된 고차원적 시각 패턴을 모델이 인식하고 해석할 수 있는지 분석하였다.

실험에는 최근 MLLMs 분야에서 널리 활용되고 있는 LLaVA-1.5 및 Qwen2.5-VL 모델을 사용하였으며, 아울러 CQA 분야에서 SOTA(State-of-the-Art)를 기록한 ChartQA-MLLM 모델을 비교 대상으로 선정하였으며, 실험에 사용된 차트 이미지는 Fig. 2의 상단 좌측에 제시된 단순한 플롯 형태와 실제 사용자 환경의 복잡한 차트 이미지로 구분하여 구성하였다[28,29,30].

#### 1. Evaluation of Chart Reasoning Capabilities

본 절에서는 MLLMs의 기본적인 차트 이미지 추론 능력을 평가하기 위해 각 차트 이미지에 대해 최고점 및 최저점의 x, y 좌표와 해당 지점의 이미지 내 위치를 식별하도록 유도하는 프롬프트를 구성하였다.

실험 결과, 단순한 플롯 형태의 차트에 대해서는 모든 모델이 전반적으로 우수한 성능을 보였다. 반면 노이즈가 포함된 실제 차트의 경우 추론 정확도가 크게 저하되거나 명백한 오답을 생성하는 사례가 다수 확인되었다. 이는 MLLMs가 정제된 시각 정보에는 효과적으로 반응하나, 실제 환경의 복잡성과 시각적 노이즈를 포함하는 상황에서는 안정적인 추론을 수행하기 어렵다는 한계를 시사한다.

#### 2. Recognition of Chart Patterns by MLLMs.

다음으로 MLLMs가 금융 차트 내에 존재하는 구조적 시각 패턴을 인식할 수 있는지를 평가하였다. 본 실험에서는 노이즈를 제거한 차트 플롯 이미지를 기반으로 금융 차트 분석에서 널리 활용되는 대표적인 패턴인 컵 앤 핸들(Cup & Handle)과 헤드 앤 숄더(Head & Shoulders)를 모델이 인식할 수 있는지를 검증하였다. 각 모델에는 패턴 정의를 포함한 프롬프트를 제시하고, 해당 차트가 패턴에 부합하는지를 판별하도록 실험을 설계하였다[5,6].

그러나 실험 결과, Fig. 2 하단에 제시된 바와 같이 세 모델 모두 복잡한 패턴 인식 실패 또는 비논리적인 응답을 생성하였다. 이는 단순 수치나 위치 기반 추론과는 달리 고차원적 시각 패턴 이해는 현재의 MLLMs에 여전히 도전적인 과제임을 보여준다.

#### 3. Case Study Conclusion

두 사례 연구를 통해 도출된 주요 시사점은 다음과 같다. 첫째, MLLMs는 단순한 형태의 플롯 차트에 대해서는 기본적인 시각 추론이 가능하지만, 실제 환경의 복잡성을 반영한 차트에서는 성능이 급격하게 저하된다.

둘째, 현재의 MLLMs는 고차원적 시각적 구조를 기반으로 한 패턴 인식 과제에 대해 충분한 해석 능력을 갖추지 못하였다.

이러한 결과는 금융 차트 분석과 같은 도메인 특화 시각-언어 과제를 효과적으로 수행하기 위해서는 보다 정교한 프롬프트 설계 전략과 사전 처리 기법의 도입이 필수적임을 시사한다.

### IV. Models and Datasets

#### 1. MLLMs Model Selection

본 연구에서는 모델 학습에 사용된 데이터의 편향 가능성을 고려하고, 다양한 성능 특성을 비교 평가하기 위하여 세 가지 대표적인 MLLMs를 선정하였다.

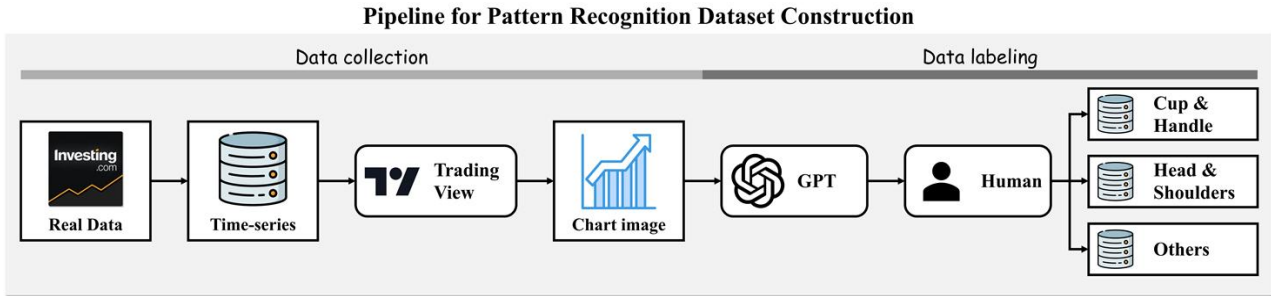


Fig. 3. Pipeline for Generating and Labeling Chart Images. To construct chart images containing noise and visual patterns, the proposed pipeline consists of two main stages: data collection and labeling.

### 1.1 Qwen2.5-VL-7B-Instruct (Qwen)

Qwen2.5-VL은 Alibaba에서 개발한 최신 버전의 비전-언어 통합 모델로, 고해상도 이미지 해석과 자연어 응답의 정밀도를 높이기 위한 설계가 적용되어 있다[29]. 7B 파라미터 기반의 지시 학습(Instruct) 기반 모델은 다양한 멀티모달 과제에서 안정적인 성능을 보여주며, 지시 기반 학습이 적용되어 프롬프트 기반 질의에 대한 반응성이 뛰어나다.

### 1.2 LLaVA-1.5-7b-hf (LLaVA)

LLaVA는 오픈 도메인 이미지 및 텍스트 데이터를 기반으로 학습된 다중모달 모델로, 다양한 시각적 질의응답 및 이미지 task에 강점을 보인다[12]. LLaVA-1.5는 CLIP 기반의 시각 인코더와 LLaMA를 통합하여 구성되었으며, 7B 파라미터 규모의 경량 모델로도 우수한 성능을 보이는 것이 특징이다[12,28]. 본 연구에서는 그 대표성과 접근성으로 인해 비교 대상으로 선정하였다.

### 1.3 Chart-MLLM

Chart-MLLM은 CQA에 특화된 MLLMs으로, 차트의 시각적 요소를 인식하고 데이터를 해석하는 능력 향상을 목표로 설계되었다[29]. 본 모델은 LLaVA-1.5를 백본으로 구축되었으며, 차트 관련 태스크에 특화된 학습을 위해 PlotQA, Chart2Table 등 도메인 전용 데이터셋을 활용한 추가 파인튜닝이 수행되었다[14,29]. 본 연구에서는 일반적인 MLLMs와의 비교를 통해 도메인 특화 학습이 차트 이해 성능에 미치는 영향을 분석하고자 해당 모델을 실험에 포함하였다.

## 2. Generation of Chart Images

본 연구에서는 제안한 프롬프트 기법이 노이즈가 포함된 차트 이미지 및 특정 시각적 패턴을 효과적으로 식별할 수 있는지를 평가하기 위해, Fig. 3에 제시된 바와 같이 데이터 셋 생성 파이프라인을 설계하였다.

Table 1. Distribution of Labeled Chart Patterns

Pattern Type	Number of Instances
Cup & handle	9
Head & Shoulders	10
Others	10

구체적으로 차트 이미지 수집을 위해 investing.com에서 제공하는 나스닥 100 지수 (NASDAQ 100) 구성 종목의 시계열 데이터를 수집하여 임의로 약 500개의 샘플링 데이터를 생성하였다. 이후 tradingview 플랫폼을 활용해 해당 데이터를 시각화함으로써 차트 이미지 데이터 셋을 구성하였다.

모델 평가를 위해 각 차트 이미지에 포함된 두 가지 시각적 패턴에 대한 라벨링 절차를 수행하였다. 우선, 수집된 이미지에 대해 GPT-4o를 활용한 자동 선분류를 수행했으며 이후 전문가의 검토를 거쳐 최종 라벨링을 확정하였다[31]. 라벨링된 데이터 통계는 Table 1에 요약되어 있다.

## V. Experiments

본 절에서는 앞선 사례 연구에서 확인된 MLLMs의 한계를 극복하고, 시각적 노이즈 제거 및 추론 보조 전략이 모델 성능에 미치는 영향을 실증적으로 분석하고자 두 가지 연구 질문(Research Questions)을 설정하였다.

*RQ1: MLLMs이 노이즈가 포함된 차트 이미지를 효과적으로 해석하기 위해서는 어떤 접근이 필요한가?*

사례 연구에서 확인된 바와 같이 MLLMs은 복잡한 실제 차트 이미지에 대해 일관된 추론을 수행하지 못하였다. 이를 해결하기 위해 본 연구는 SAM (Segment Anything Model)을 활용하여 차트 선만을 추출하고, 이를 기반으로 구성된 노이즈 제거 차트 이미지를 모델 입력으로 제공

Table 2. Summary of experimental configurations per reasoning strategy

Reasoning Strategy	In-Context Learning	Chain-of-Thought	Self-consistency
Batch size	4		
Model max length	512		
Top-k	50	50	40
Number of shots	2	-	0
Reasoning steps	-	5	-
Number of samples	-	-	5

Table 3. Prompt Strategy Combinations for Each Experiment

Experiment	ICL	CoT	SC
Experiments 01	⊙		
Experiments 02		⊙	
Experiments 03			⊙
Experiments 04	⊙	⊙	
Experiments 05	⊙		⊙
Experiments 06		⊙	⊙
Experiments 07	⊙	⊙	⊙

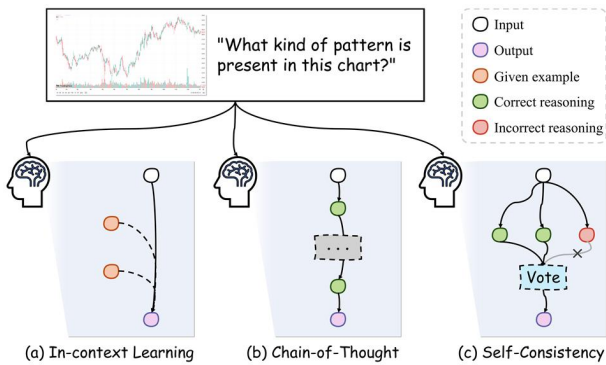


Fig. 4. Three prompting strategies designed to enhance MLLMs' ability to recognize finance patterns in chart images: (a) **In-Context Learning** provides prior examples to guide pattern inference, (b) **Chain-of-Thought** decomposes the task into simpler reasoning steps before final decision-making, and (c) **Self-Consistency** aggregates multiple reasoning paths to produce a more reliable prediction.

하는 전략을 제안하였다[32]. 이러한 전처리 과정을 통해 모델은 보조 지표, 배경 격자, 주석 등 불필요한 시각 정보를 제거한 노이즈의 영향을 최소화한 상태에서 추론이 가능해진다.

*RQ2: MLLMs이 차트 내 금융 패턴을 인식하기 위해서는 어떤 프롬프트 전략이 효과적인가?*

RQ2는 사람이 차트를 해석하고 패턴을 식별하는 인지 과정을 기반으로 프롬프트 설계 방안을 제안한다.

첫째, Fig. 4(a)의 ICL 방식은 사람이 새로운 차트를 분석할 때, 주어진 예시와 비교하여 패턴을 유추하는 인지 과정에 착안한 것으로 모델에게 사전 예시와 유사한 패턴을 제공함으로써 유사성 기반의 추론을 유도한다[15].

둘째, Fig. 4(b)의 CoT 방식은 복잡한 패턴을 인식하기에 앞서, 차트의 구조적 이해를 돕는 일련의 기초 질문을 통해 모델의 추론 과정을 단계적으로 유도하는 전략이다 [16]. 이를 통해 단편적인 판단이 아닌 논리적인 해석 기반의 패턴 인식을 도모한다.

셋째, Fig. 4(c)의 SC 전략은 하나의 응답에 의존하지 않고 다양한 사고 경로를 탐색하여 다수의 후보 응답을 생성한 뒤, 가장 빈도 높은 응답을 최종결과로 선택함으로써 신중하고 일관된 판단을 가능하게 한다[17].

본 연구는 세 가지 전략을 개별적으로 적용하는 데 그치지 않고, 전략 간 시너지 효과를 검증하기 위해 이들을 복합적으로 조합한 실험을 수행하였다. 이러한 실험 설계를 통해 단일 전략 대비 복합 전략의 효과성 및 전략 간 상호작용이 실제 성능에 미치는 영향을 더욱 정교하게 분석할 수 있으며, 이는 실질적인 프롬프트 설계 가이드라인 도출에 기여할 수 있다. 각 실험 전략별 하이퍼파라미터 설정 및 실험 조합에 관한 세부 사항은 Table 2와 3에 요약되어 있다. 본 연구에서 사용된 모든 모델은 파라미터를 동결한 상태로 평가하였으며, 실험은 단일 NVIDIA RTX A6000 GPU에서 수행되었다. 또한 각 프롬프트 예시는 Figure 5에 제시하였다.

## VI. Results

Table 4은 SAM의 사용 여부에 따라 다양한 프롬프팅 전략이 MLLMs의 패턴 분류 정확도에 미치는 영향을 비교한 결과를 보여준다[32]. SAM을 적용한 실험 조건에서 결과를 종합하면 Qwen이 평균 정확도 43.02%로 가장 우수한 성능을 기록하였다. LLaVA 37.37%와 Chart-LLaVA 36.94%가 그 뒤를 따랐다[30,32]. Qwen은 SC 단독 전략

Table 4. Prompting Strategy Performance with and without SAM across MLLMs.

	Model	Exp 01	Exp 02	Exp 03	Exp 04	Exp 05	Exp 06	Exp 07	Mean
Qwen	w/ SAM	37.93	27.59	<b>66.67</b>	44.83	41.38	37.93	44.83	43.02
	w/o SAM	34.48	31.03	34.48	<b>44.83</b>	41.38	34.48	41.38	37.29
LLaVA	w/ SAM	34.48	31.03	37.93	37.48	<b>51.72</b>	31.03	37.93	37.37
	w/o SAM	34.48	31.03	37.93	34.48	<b>51.72</b>	31.03	37.93	36.94
Chart-LLaVA	w/ SAM	34.48	31.03	<b>44.83</b>	37.93	34.48	34.48	41.38	36.94
	w/o SAM	<b>34.48</b>	31.03	31.03	27.59	31.03	31.03	24.14	30.05

인 실험 03에서 66.67%의 정확도를 달성하며 모든 실험 중 최고 성능을 기록하였다[17,29]. 또한 Chart-LLaVA도 동일 실험에서 44.83%로 가장 우수한 성능을 보였으며 LLaVA 역시 37.93%로 평균을 상회하며 세 모델 모두 SC 전략의 효과가 뚜렷하게 나타났다[17,28,30]. 이는 다중 추론 경로를 활용한 응답 집계 방식이 시각적 패턴 해석에 효과적임을 시사한다.

프롬프트 전략 조합에 대한 실험으로 Table 4 실험 04-07 를 진행하였고, 전략 간 상호작용 특성이 다르게 나타났다. Qwen은 ICL이 포함된 대부분의 조합 전략에서 정확도 40% 이상을 안정적으로 유지하며 복합 전략 구성에서도 견고한 성능을 보였다[15,29]. LLaVA는 ICL과 SC를 결합한 전략에서 가장 우수한 성능을 보였으며, 실험 02, 06 과 같이 대체로 CoT가 포함된 조합에서는 오히려 성능이 감소하는 경향을 보였다[15,16,17,28]. 이는 복합 프롬프트가 항상 시너지 효과를 발휘하는 것이 아니라 특정 전략 간 interference가 발생할 수 있음을 보여준다. Chart-LLaVA는 조합 전략 전반에서 큰 성능 변화가 관찰되지 않았으며 모든 프롬프트 전략을 결합한 실험 07에서 두 번째로 우수한 성능을 기록하며 상대적으로 프롬프트 구성에 대한 민감도가 높지 않은 것으로 해석된다[30].

Table 4에서 확인할 수 있듯이, SAM을 활용한 시각적 전처리는 전반적으로 모델의 성능 향상에 긍정적인 영향을 주는 것으로 나타났다[32]. Qwen은 SC 단독 전략이 적용된 실험 03에서, SAM 기반 차트 크롭 전처리를 적용했을 때 정확도 66.67%로 모든 실험 중 가장 우수한 성능을 보였다[17,29,32]. 이는 SAM을 사용하지 않은 실험 대비 약 32%p 이상 향상된 결과로, Qwen이 시각적 노이즈가 제거된 상태에서의 추론 기반 프롬프트에 민감하게 반응함을 보여준다[29,32]. 반면 LLaVA는 크롭 적용 후 약 1.4%의 성능 향상을 보이며 전처리의 효과가 제한적으로 나타났다[28]. 이는 해당 모델이 시각적 전처리에 의한 성능 영향보다는 프롬프트 구성 방식에 더 민감하게 반응하는 구조임을 나타낸다. Chart-LLaVA의 경우 크롭 도입 후 평균 정확도가 6.9% 증가하며 가장 안정적인 성능 개

선을 보였다[30]. 이는 모델 구조에 따라 시각적 노이즈 제거가 추론 정확도에 미치는 영향이 다르게 나타날 수 있음을 시사하며, MLLMs 기반 금융 차트 분석에서 시각적 전처리가 단순한 보조 수단이 아닌 모델별 특성과 전략 간 상호작용을 고려하여 신중하게 설계되어야 할 핵심 요소를 나타낸다.

전반적인 프롬프트 전략과 크롭에 대한 실험 결과를 종합해보면, 세 모델은 서로 다른 민감도와 반응을 보였다. Qwen은 전체적으로 가장 높은 정확도를 기록했으며 SC 단독 전략과 ICL 기반 전략 조합에서 강점을 보였다 [15,17,29]. LLaVA는 프롬프트 전략이나 전처리 변화에 따른 반응은 안정적인 편이었다[28]. Chart-LLaVA는 전반적으로 가장 낮은 평균 정확도를 보였지만 SAM 전처리에 따른 성능 향상이 가장 두드러져 시각적 노이즈 제거가 모델 성능에 결정적인 영향을 미치는 구조임을 보여준다 [30,32]. 이러한 차이는 MLLMs 간의 시각 정보 처리 방식, 프롬프트 해석 전략의 차이에서 기인할 수 있고 향후 모델별 설계 방향 및 전처리 전략의 선택에 있어 중요한 기준이 될 수 있다.

## VII. Conclusion

본 연구는 MLLMs가 복잡한 금융 차트 환경에서 추가 패턴을 인식하고 해석할 수 있는지를 실증적으로 분석하였다. 실험을 분석해 본 결과, MLLMs는 단순한 플롯 형태의 차트에 대해서는 비교적 양호한 시각 추론 성능을 보였지만 실제 환경에서 생성된 노이즈가 포함된 실제 추가 차트에서는 성능 저하가 뚜렷하게 나타났다.

이를 해결하고자 본 연구에서는 세 가지 프롬프트 전략을 기반으로 다양하게 조합하여 MLLMs에 적용하고, 그 효과를 실험적으로 검증하였다. SC 전략은 세 모델 모두에서 가장 큰 성능 향상을 이끌었으며, Qwen의 경우 해당 전략이 적용된 실험 03에서 66.67%의 정확도로 전체 실험 중 최고 성능을 보였다[17,29]. 이 결과는 다중 추론 경

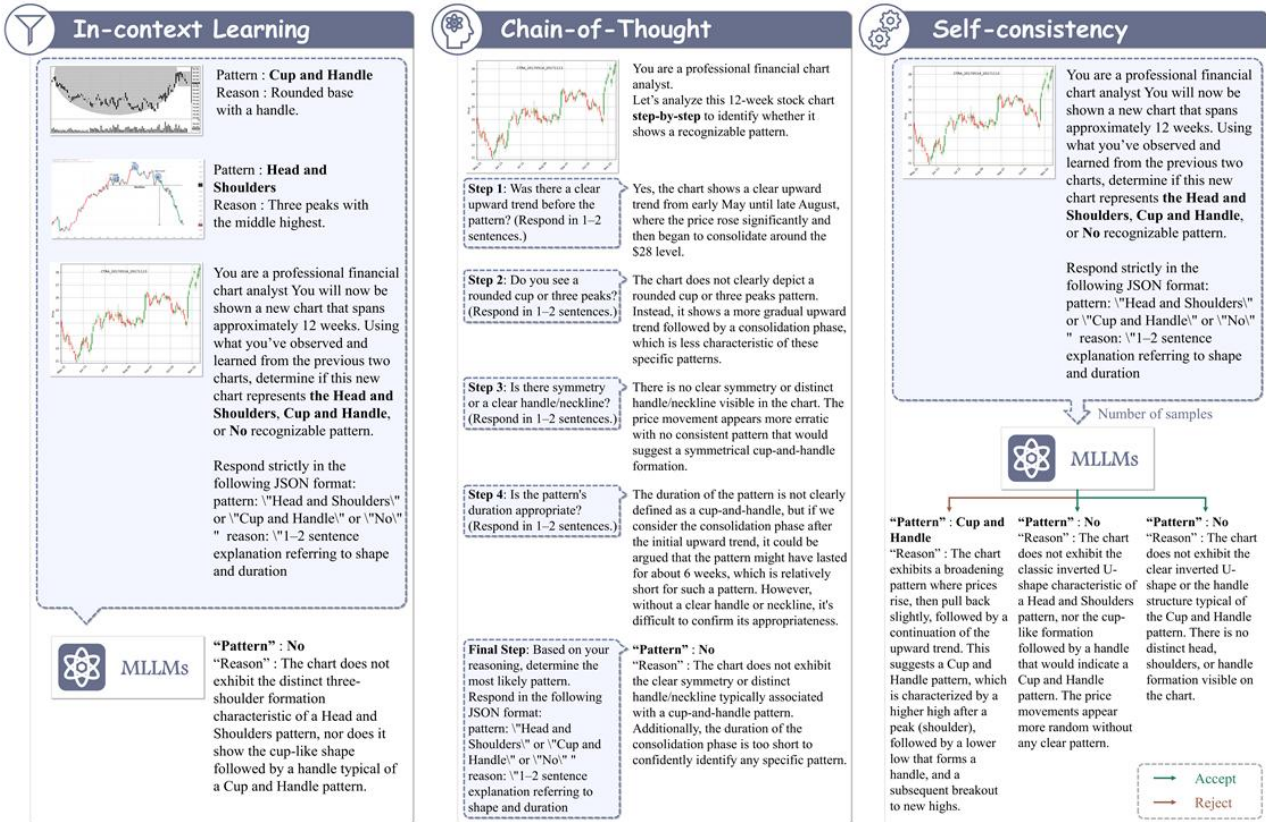


Fig. 5. Illustration of three prompting strategies—In-context Learning (ICL), Chain-of-Thought (CoT), and Self-consistency (SC)—applied to a stock chart pattern recognition task. The examples shown for each strategy are based on the responses generated by the multimodal language model Qwen2.5-VL-7B-Instruct.

로 기반의 응답 집계 방식이 시각적 패턴 해석에 효과적임을 시사한다.

또한 SAM 기반 차트 크롭은 모델별로 상이한 효과를 보였다[32]. Chart-LLaVA와 같은 차트 특화 모델에서 평균 정확도가 약 7% 상승하며 시각적 노이즈 제거가 MLLMs의 추론 성능에 유의미한 영향을 미칠 수 있음을 실증적으로 보여준다[30].

본 연구는 단순한 모델 선택을 넘어 도메인 특화된 프롬프트 전략과 시각 정보 전처리의 정밀한 설계가 MLLMs 기반 금융 차트 분석의 성능을 결정짓는 핵심 요소임을 강조한다. 본 실험을 통해 도출된 결과는 향후 MLLMs를 활용한 자동화된 투자 판단 시스템 구축에 있어 실질적인 기초 자료로 활용될 수 있을 것이다.

다만 본 연구는 다음과 같은 한계점을 가진다.

첫째, 실험에 활용된 데이터셋의 규모가 제한적이기에 모델 간 또는 프롬프트 전략 간 성능 차이를 통계적으로 일반화하기에는 제약이 있다. 그럼에도 불구하고, 본 연구는 제한된 라벨링 데이터를 바탕으로 MLLMs의 시각 추론 능력을 체계적으로 비교·분석한 초기적 시도로서, 프롬프트 전략의 조합과 시각 전처리 기법이 성능에 미치는 영향

을 정량적으로 분석했다는 점에서 기존 연구와 차별화된다.

둘째, 본 연구에서 주로 다룬 Head & Shoulders 및 Cup & Handle 패턴은 널리 활용되는 대표적인 시세 패턴이지만, 해당 패턴이 향후 수익률을 유의하게 예측하는 지에 대해서는 학술적으로 증명되지 않았다. 이에 따라 본 연구는 패턴의 실용성을 단정하기보다는 MLLM이 해당 패턴을 인식하고 분류할 수 있는지에 대한 탐색적 가능성을 중심으로 분석을 진행하였다.

향후 연구에서는 더욱 다양한 기술적 패턴을 포함하여 라벨링 규모를 확장한 실험 설정, 유의성 검증, 실제 투자 성과 평가 등을 통해 MLLM의 패턴 인식 능력을 정교하게 평가할 수 있는 기반을 마련할 것이다. 이를 바탕으로 fine-tuning된 사전 학습 모델을 개발하고, 단순한 패턴 분류를 넘어 실시간 시장 흐름 해석 및 투자 판단으로 이어지는 end-to-end 시스템 구축을 궁극적인 목표로 설정한다. 본 연구는 그 가능성을 입증하기 위한 초기적 시도로서, 멀티모달 금융 분석 시스템 개발을 위한 실증적 기반을 제공한다.

## ACKNOWLEDGEMENT

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2024S1A5C3A03046579)

## REFERENCES

- [1] Karadaş, Furkan, Bahaeddin Eravcı, and Ahmet Murat Özbayoğlu, "Multimodal Stock Price Prediction," arXiv, arXiv:2502.05186, Feb. 2025. <https://doi.org/10.5220/0013174500003890>
- [2] Li, Yinheng, et al., "Large language models in finance: A survey," Proceedings of the Fourth ACM International Conference on AI in Finance, Sep. 2023. <https://doi.org/10.48550/arXiv.2311.10723>
- [3] Lo, Andrew W., Harry Mamaysky, and Jiang Wang, "Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation," The Journal of Finance, vol. 55, no. 4, pp. 1705-1765, Aug. 2000. <https://doi.org/10.1111/0022-1082.00265>
- [4] Deng, Q., Y. Luo, and J. Ge, "Dual threshold based unsupervised face image clustering," Proceedings of the 2nd International Conference on Industrial Mechatronics and Automation, pp. 436-439, Wuhan, China, May 2010. <https://doi.org/10.1109/ICINDMA.2010.5538145>
- [5] Osler, Carol L., and P. H. Chang, "Head and shoulders: Not just a flaky pattern," FRB of New York Staff Report, no. 4, Aug. 1995. <https://doi.org/10.2139/ssrn.993938>
- [6] Murphy, John J., Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications, Penguin, 1999.
- [7] Leigh, William, et al., "Stock market trading rule discovery using technical charting heuristics," Expert Systems with Applications, vol. 23, no. 2, pp. 155-159, Aug. 2002. [https://doi.org/10.1016/S0957-4174\(02\)00034-9](https://doi.org/10.1016/S0957-4174(02)00034-9)
- [8] Islam, Mohammed Saidul, et al., "Are large vision language models up to the challenge of chart comprehension and reasoning?," arXiv, arXiv:2406.00257, Jun. 2024. <https://doi.org/10.48550/arXiv.2406.00257>
- [9] Masry, Ahmed, et al., "ChartInstruct: Instruction Tuning for Chart Comprehension and Reasoning," arXiv, arXiv:2403.09028, Aug. 2024. <https://doi.org/10.48550/arXiv.2403.09028>
- [10] Tsai, Yao-Hung Hubert, et al., "Multimodal transformer for unaligned multimodal language sequences," Proceedings of the Association for Computational Linguistics (ACL), vol. 2019, Florence, Italy, Jul. 2019. <https://doi.org/10.48550/arXiv.1906.00295>
- [11] Bai, Jinze, et al., "Qwen-vl: A frontier large vision-language model with versatile abilities," arXiv, arXiv:2308.12966, Aug. 2023. <https://doi.org/10.48550/arXiv.2308.12966>
- [12] Liu, Haotian, et al., "Visual instruction tuning," Advances in Neural Information Processing Systems, vol. 36, pp. 34892-34916, Dec. 2023. <https://doi.org/10.48550/arXiv.2304.08485>
- [13] Zeng, Xingchen, et al., "Advancing multimodal large language models in chart question answering with visualization-referenced instruction tuning," IEEE Transactions on Visualization and Computer Graphics, Jan. 2025. <https://doi.org/10.48550/arXiv.2407.20174>
- [14] Ford, James, et al., "Charting the Future: Using Chart Question-Answering for Scalable Evaluation of LLM-Driven Data Visualizations," arXiv, arXiv:2409.18764, Sep. 2024. <https://doi.org/10.48550/arXiv.2409.18764>
- [15] Brown, Tom, et al., "Language models are few-shot learners," arXiv, arXiv:2005.14165, May. 2020. <https://doi.org/10.48550/arXiv.2005.14165>
- [16] Wei, Jason, et al., "Chain-of-thought prompting elicits reasoning in large language models," arXiv, arXiv:2201.11903, Nov. 2022. <https://doi.org/10.48550/arXiv.2201.11903>
- [17] Wang, Xuezhi, et al., "Self-consistency improves chain of thought reasoning in language models," Proceedings of the 11th International Conference on Learning Representations (ICLR), pp. 1-17, May. 2023. <https://doi.org/10.48550/arXiv.2203.11171>
- [18] Song, Hyun-Jung, and Suk-Jun Lee, "A study on the optimal trading frequency pattern and forecasting timing in real time stock trading using deep learning: focused on KOSDAQ," The Journal of Information Systems, vol. 27, no. 3, pp. 123-140, Jul. 2018. <https://doi.org/10.5859/KAIS.2018.27.3.123>
- [19] Kim, Sun Woong, "Technical trading rules for bitcoin futures," Journal of Convergence for Information Technology, vol. 11, no. 5, pp. 94-103, May. 2021. <https://doi.org/10.22156/CS4SMB.2021.11.05.094>
- [20] Cho, Kyu Cheol, and Sung Hee Lee, "System Development of the Stock Price Prediction," Proceedings of the Korean Society of Computer Information Conference, pp. 161-162, Seoul, Korea, Jan. 2021. UCI(KEPA): I410-ECN-0101-2021-004-001463767
- [21] Jeong, Cheon Su, "Domain-specialized LLM: Financial fine-tuning and utilization method using Mistral 7B," Journal of Intelligence and Information Systems, vol. 30, no. 1, pp. 93-120, Mar. 2024. <https://doi.org/10.13088/jiis.2024.30.1.093>
- [22] Che, Wanliu, et al., "Predicting financial distress using multimodal data: An attentive and regularized deep learning method," Information Processing & Management, vol. 61, no. 4, article 103703, Jul. 2024. <https://doi.org/10.1016/j.ipm.2024.103703>
- [23] Masry, Ahmed, et al., "Unichart: A universal vision-language pretrained model for chart comprehension and reasoning," arXiv, arXiv:2305.14761, 2023. <https://doi.org/10.48550/arXiv.2305.14761>

14761

- [24] Han, Yucheng, et al., "ChartLLaMA: A multimodal LLM for chart understanding and generation," arXiv, arXiv:2311.16483, Nov. 2023. <https://doi.org/10.48550/arXiv.2311.16483>
- [25] Kojima, Takeshi, et al., "Large language models are zero-shot reasoners," *Advances in Neural Information Processing Systems*, vol. 35, pp. 22199–22213, Dec. 2022. [doi.org/10.48550/arXiv.2205.11916](https://doi.org/10.48550/arXiv.2205.11916)
- [26] Zhang, Zhuosheng, et al., "Multimodal chain-of-thought reasoning in language models," arXiv, arXiv:2302.00923, Feb. 2023. <https://doi.org/10.48550/arXiv.2302.00923>
- [27] Liang, Zijiang, et al., "A Survey of Multimodal Large Language Models," *Proceedings of the 3rd International Conference on Computer, Artificial Intelligence and Control Engineering*, Beijing, China, Jan. 2024. <https://doi.org/10.48550/arXiv.2306.13549>
- [28] Liu, Haotian, et al., "Improved baselines with visual instruction tuning," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, Jun. 2024. <https://doi.org/10.48550/arXiv.2310.03744>
- [29] Wang, Peng, et al., "Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution," arXiv, arXiv:2409.12191, Sep. 2024. <https://doi.org/10.48550/arXiv.2409.12191>
- [30] Masry, Ahmed, et al., "ChartQA: A benchmark for question answering about charts with visual and logical reasoning," arXiv, arXiv:2203.10244, Mar. 2022. <https://doi.org/10.48550/arXiv.2203.10244>
- [31] Hurst, Aaron, et al., "GPT-4o System Card," arXiv, arXiv:2410.21276, Oct. 2024. <https://doi.org/10.48550/arXiv.2410.21276>
- [32] Kirillov, Alexander, et al., "Segment anything," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, Oct. 2023. <https://doi.org/10.48550/arXiv.2304.02643>

## Authors



Eun Hong Park received the bachelor's degree in the Department of Business Administration from Soongsil University in 2021. She is a master student at the Yonsei University Graduate School of Information

since 2024. Her current research interests include deep learning and financial engineering.



Ji Hoon Park received the bachelor's degree in Industrial and Management Engineering from Myongji University in 2024. He is a master's student at the Yonsei University Graduate School of Information since 2024.

His current research interests include Multi-Modal Learning, and Generative Models.



Ha Young Kim is an Associate Professor at Graduate School of Information, Yonsei University, Korea. She received her Ph.D. degree at department of Mathematics, Purdue University, USA.

From 2011 to 2016, she was a research staff member in Samsung Advanced Institute of Technology (SAIT) of Samsung Electronics, Korea, working on various recognition systems with deep learning. Her primary research areas are deep learning and computational Finance. She has published in leading journals, including *Information Fusion*, *Applied Soft Computing*, *Expert Systems with Applications*, *Stochastic and Dynamics*, *Computers in Biology and Medicine*, *PLoS ONE*, *automation in construction*, *Journal of Computing in Civil Engineering* and *Annals of Finance*. She is the inventor of 8 patents and 13 patent applications.