Decision-Oriented Multimodal Financial Analytics: Dataset and Reinforcement Learning Approach

Jie Sun[1,2] Tengxiang Luo[2] Huamao Gu[2] Jing Hua[2] Rui Xie[3] Zuohua Ding[1]

[1] *School of Computer Science and Technology(School of Artificial Intelligence), Zhejiang Sci-Tech University, Hangzhou 310000, China*

[2] *School of Computer Science and Technology,Zhejiang Gongshang University,Zhejiang 310000, China*

[3] *Zhejiang Zheshiyou Comprehensive Energy Sales Co., Ltd, Zhejiang 310000, China*

Abstract

Large Language Models (LLMs) in the financial domain have predominantly focused on high-risk financial market prediction. However, there remains a severe scarcity of tools and corresponding high-quality datasets capable of generating in-depth financial analysis reports to assist human experts in investment decision making. To address this research gap, this paper makes two main contributions. First, we construct and release MultiModal Stock Analytics (MMSA), the first large-scale multimodal financial data set that deeply couples real stock charts with expert-level analytical texts. Second, we propose a novel reinforcement learning method named Stock-R1, which enhances the Group Relative Policy Optimization (GRPO) framework by introducing two core mechanisms. The first mechanism operationalizes the principles of Axiomatic Construction in a composite reward function to ascertain the validity and optimality of the analysis. The second mechanism introduces a Progressive Curriculum Reward strategy, which enables efficient curriculum learning by dynamically reshaping the incentive landscape. Extensive experiments on MMSA demonstrate that Stock-R1's performance significantly outperforms several State-of-the-Art (SOTA) models, including those with larger parameter counts. Furthermore, compared to the conventional Supervised Fine-Tuning (SFT) method, it achieves an improvement of up to 30% in the F1 score on key analytical tasks. This work provides a new benchmark dataset and an efficient training methodology, driving the paradigm shift of financial Large Vision-Language Models (LVLMs) from market prediction to decision support.

Keywords: LLM;Large vision-language models; Group relative policy optimization; Multi-modular stock analytics; Reinforcement learning

1. Introduction

Financial market analysis is fundamentally a multimodal task, relying on experts' integrated interpretation of visual charts and textual signals. However, traditional quantitative models [1–7]

†Corresponding author: Zuohua Ding (Email: zouhuading@hotmail.com)

and existing public datasets [5, 7, 8] are predominantly unimodal. Although some research has begun to leverage financial images [9, 10], the market still significantly lacks high-quality data sets that tightly link charts rich in technical indicators with their structured and in-depth textual interpretations. This focus on chart imagery is deliberate. While time series data provide raw numerical input, technical charts serve as a critical abstraction layer, converting numerical sequences into the visual patterns central to expert human analysis. Our research motivation is thus to emulate this human-centric visual reasoning process. Consequently, for the task of generating reports that replicate expert discourse, chart images are the natural and indispensable input modality, positioning our work to address the distinct challenge of deep visual chart interpretation.

While the latest Large Vision-Language Models (LVLMs) [11–20] provide the technical foundation to address this challenge, their fine-tuning paradigms within the financial domain remain bottlenecked. Mere Supervised Fine-Tuning (SFT) [16] is insufficient for cultivating deep cross-modal analytical capabilities, prompting our turn to Reinforcement Learning (RL). Within the RL paradigm, methods based on human preference [21–27] are ill-suited for the objectivity required in financial analysis due to their inherent subjectivity. Consequently, rule-based reinforcement learning [28–30] —which aligns models with objective, verifiable criteria—emerges as a more promising technical path. However, the complexity and hierarchical nature of financial analysis also reveal a fundamental limitation of existing rule-based RL paradigms when handling complex generation tasks: the lack of a sophisticated mechanism capable of precisely modeling multidimensional, structured logic and dynamically guiding the learning process.

To address the dual challenges of data and methodology, this paper first introduces MMSA, a novel multimodal stock analysis dataset containing 20k chart-text pairs. Each entry consists of a professional stock chart and an in-depth textual description generated by GPT-4o [31] and rigorously calibrated by 15 financial experts. To our knowledge, MMSA is the first dataset to systematically couple visual stock information with deep textual interpretations, aiming to provide a cornerstone for training multimodal models capable of complex, cross-modal analytical reasoning.

Building on this data foundation, we further propose Stock-R1, a rule-based reinforcement learning framework specifically designed for stock chart analysis. At its core, it utilizes the GRPO [28–30] algorithm, guided by an axiomatic composite reward function, which we designed. The framework also incorporates a Progressive Curriculum Reward strategy. This strategy mimics the concept of curriculum learning [32] by dynamically increasing the difficulty of evaluation during training. This ensures steady evolution of the model's capabilities and effectively prevents reward hacking.

We conducted comprehensive experiments on the MMSA dataset and the results fully validate the effectiveness of our approach. The main contributions of this paper can be summarized as follows.

1. We built and are releasing MMSA, the first multimodal financial dataset that systematically couples professional stock charts with in-depth textual interpretations. This dataset provides a critical foundation for rule-based model optimization and will be made publicly available upon the paper's acceptance.

2. We propose the innovative Stock-R1 fine-tuning framework. Its core technical contribution lies in two synergistic mechanisms specifically designed to overcome the unique challenges of financial analysis: (1) An Axiomatic Reward Construction that translates the complex, hierarchical logic of expert-level financial reports into deterministic, verifiable reward signals, transcending subjective human preferences. (2) A Progressive Curriculum Reward strategy that dynamically adjusts learning objectives, addressing the critical issues of reward sparsity and capability stagnation commonly encountered when generating complex financial reports that must simultaneously satisfy both structural regularity and multi-dimensional content correctness.
3. We validate a new, efficient fine-tuning paradigm that works without human preference labeling. This paradigm successfully extends deterministic rule-based reinforcement learning to complex multimodal financial scenarios, offering a viable solution for model alignment challenges in specialized domains like finance where data is scarce.

2. Related Work

2.1. Progress in the General Capabilities of Vision-Language Models

In recent years, the development of LVLMs has progressed at a remarkable pace [11–16]. Led by cutting-edge closed-source models like GPT-4o [31], LVLMs have achieved outstanding visual understanding and multimodal interaction capabilities by effectively integrating visual encoders with LLMs. This fusion of vision and text has significantly enhanced the models' ability to comprehend complex scenes, process chart-and-text inputs, and perform corresponding reasoning. As a result, it has driven the development of more advanced foundation models. In the open-source community, the capabilities of LVLMs have also made significant breakthroughs by aligning visual modules with state-of-the-art LLMs [33–35] and leveraging high-quality instruction data [15, 36] for end-to-end training. Among these, leading open-source models such as InternVL-2.5 [37] and Qwen2.5-VL [38] have demonstrated performance on challenging multidisciplinary, multimodal reasoning benchmarks like MMMU [39] that is gradually approaching or even surpassing closed-source models [40], showcasing their powerful potential in general-purpose question answering and reasoning.

2.2. The New Paradigm of Reinforcement Learning with Rule-Based Rewards

Despite the success of LVLMs in general visual understanding, how to efficiently fine-tune them to precisely execute complex tasks in specific domains remains a core challenge. In response, the research community has begun to explore new paths beyond traditional instruction fine-tuning, with direct optimization of model behavior through reinforcement learning emerging as a promising direction. A particularly prominent new paradigm is reinforcement learning with rule-based rewards, which shifts the optimization objective from mimicking human-annotated data to satisfying a series of objectively verifiable, programmatic criteria. Pioneering work in this direction, such as Deepseek-R1 [28], has demonstrated that even by skipping the supervised fine-tuning stage, high-performance models can be trained in tasks like mathematics [30] and code [29] generation solely through reinforcement learning algorithms like GRPO [28–30] and clear, rule-based rewards. This strategy provides models with a more consistent and unambiguous learning signal. Its data efficiency and significant effectiveness have inspired us in tackling the problem of scarce supervised data and rule-assessable task performance in stock chart analysis. This has laid the methodological foundation for the design of Stock-R1.

2.3. Financial Models

In recent years, the advancement of LLMs has catalyzed the creation of numerous finance-specific models [41–43]. While these models excel at processing vast amounts of financial text, they generally overlook critical visual information, such as the stock charts that human analysts heavily rely on. To bridge this gap, recent multimodal explorations in finance [9, 10] have begun to integrate visual data. However, their paradigm still largely focuses on directly predicting future trends by analyzing historical charts. However, given the inherent stochasticity of financial markets and the Efficient Market Hypothesis, the stability and reliability of prediction-centric methods are often difficult to guarantee.

We contend that a more robust and practical financial model should not aim to replace decision-makers by making uncertain predictions, but rather serve as an advanced decision support tool. Specifically, such a model should be capable of deeply integrating and understanding multimodal information like a human expert to generate a high-quality, logically coherent, and opinionated stock analysis report. Such reports can provide investors with a comprehensive basis for their decisions, assisting them in making their own judgments. However, the current research landscape lacks multimodal financial models centered on generating high-quality analytical reports, as well as the high-quality chart-text paired datasets required to train them.

3. MMSA:MultiModal Stock Analytics

The construction of the MMSA dataset follows a three-stage process, which includes data collection, automated text generation, and manual verification. The core of this process lies in a meticulously designed structured prompt engineering strategy, which guides a LVLM to generate professional texts that include deep analytical logic and are strictly aligned with the corresponding charts, thereby ensuring the high-quality standard of the dataset.

3.1. Data Collection and Generation

Data and Chart Construction Our dataset originates from the daily trading data of companies listed on the Shenzhen Stock Exchange, covering two distinct six-month periods (April 2023–September 2023 and October 2023–March 2024) across different market cycles to enhance the model's generalization capabilities. We further calculated four mainstream technical indicators (MACD, KDJ, RSI, and BOLL)[1] with their commonly used parameters. These indicators, along with price candlesticks and moving averages, are standardized and visualized to generate

[1] The four technical indicators are briefly described as follows: MACD (Moving Average Convergence Divergence) is a trend-following momentum indicator that shows the relationship between two moving averages of a price. KDJ (Stochastic Oscillator) is a momentum indicator commonly used to determine overbought and oversold conditions. Similarly, RSI (Relative Strength Index) is an indicator that measures momentum to identify overbought (typically >70) or oversold (typically <30) levels. BOLL (Bollinger Bands) is a volatility indicator used to define the relative high and low range of a price.

the charts required for our research.

Analytical Text Generation The generation of analytical text relies on a carefully designed sophisticated structured prompt engineering strategy to ensure professionalism and objectivity. This strategy is based on three core principles: role-based guidance, modular instructions, and quantitative signal definitions (detailed in Appendix A). Its highly standardized output provides an efficient validation scaffold for the subsequent manual verification stage, significantly reducing the review burden on experts.

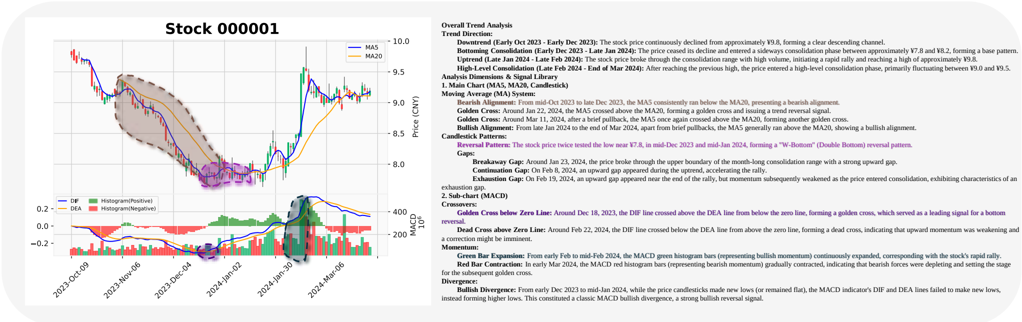

Figure 1. An illustrative sample from our MMSA dataset. The sample consists of a visual stock chart enriched with technical indicators (left), tightly coupled with a detailed and verifiable analytical text (right). In contrast to simpler chart-text pairs in existing works, this deep alignment between visual features and semantic analysis provides a robust foundation for training models to understand complex financial logic.

3.2. Two-Stage Human Quality Control

To guarantee the professionalism and reliability of the generated text in the MMSA dataset, we established a rigorous two-stage human review process, executed on a specially developed online annotation platform.

Cross-Screening and Revision by Graduate Students This stage was carried out by 10 graduate students with backgrounds in financial engineering or related fields. Each chart-text pair was randomly assigned to three annotators for an independent back-to-back review. The review was based on three core criteria: 1. Factual Consistency: Verifying point-by-point that all signals described in the report were in complete agreement with the visual information in the chart. 2. Logical and Professional Soundness: Assessing the coherence of the analytical logic and the accuracy of professional terminology. 3. Objectivity: Ensuring that the text is devoid of subjective assumptions or speculative market predictions.

We used a majority vote system ($\geq 2/3$ approval) to determine whether a sample advanced to the next round. To quantify the consistency of our review criteria, we calculated the Fleiss' Kappa coefficient, which yielded a value of 0.76, indicating substantial agreement in our annotation standards.

Final Review by Experts Samples that passed the initial screening were submitted to a panel of five senior financial analysts. Leveraging their extensive practical experience, the expert panel conducted a final review and refinement of the samples' analytical depth and the precision of key signal interpretations. Only samples that received unanimous approval from the expert panel were ultimately included in the MMSA dataset.

Through this stringent two-stage review process, approximately 28% of the initially generated chart-text pairs were discarded for failing to meet our quality standards, which thoroughly ensures the high professional caliber of the final data.

3.3. Dataset Analysis and Comparison

The final MMSA dataset we constructed contains 20,000 high-quality chart-text pairs. The sample distribution of its technical indicators is shown in Table 1. Figure 1 displays a typical MMSA sample, intuitively presenting the visual richness of our charts and the depth of our analytical texts.

Table 1. Distribution Statistics of the MMSA Dataset by Technical Indicator

Class	MACD	KDJ	RSI	BOLL
Num	4663	5258	5037	5042

As shown in Table 2, MMSA differs significantly from existing financial datasets in terms of modality, content depth, and core objective. Unlike unimodal datasets such as FinBen, MMSA introduces the crucial visual information of charts. When compared to other multimodal datasets like FinVis-GPT and FinTral, MMSA's advantages are even more pronounced:

Visual Richness MMSA's charts integrate multiple core technical indicators, rather than the single candlestick charts found in most datasets, providing a richer and more comprehensive visual foundation for in-depth technical analysis.

Textual Professionalism The text in MMSA consists of professional analytical reports produced through a process of programmatic generation and expert verification. These reports contain complete logical chains, achieving a qualitative leap from superficial description to deep analysis.

Core Objective MMSA focuses on training a model's decision support capabilities, rather than the market prediction tasks common to existing datasets, thereby filling a market gap for high-quality, analysis-oriented datasets.

4. Stock-R1

4.1. Challenges and Methodological Formulation

Although rule-based reinforcement learning has shown success in domains like mathematics and code generation, financial analysis presents a unique set of challenges that necessitate significant methodological innovation. Unlike tasks with a single, clear objective, generating a high-quality financial report requires the model to simultaneously satisfy multiple, and at times even orthogonal, constraints such as factual accuracy, logical coherence, and structural integrity.

Table 2. Feature Comparison of MMSA with Existing Related Datasets

Feature	Finben	FinVis-GPT	FinTral	MMSA(ours)
Data modal	Text only	Charts + text	Charts + text	Chart+Text (Main/Sub)
Main task	SA,NER,etc.	Chart Desc., Trend Pred.	Multiple Tasks, Trend Pred.	In-depth Analysis Report
Visual content	-	K-line Chart	K-line Chart	K-line Chart + Tech. Indicators
Textual content	Original News/ Fin. Reports	Simple Desc. + Pred. Labels	Diverse Text	Expert-Level Analysis
Data scale	5K	100K	180K	10K
Goal-oriented	Text Understanding	Market Prediction	Market Prediction	Decision Aid

This complexity gives rise to two fundamental challenges: 1) The complexity of Reward Definition: how to quantify these abstract, multi-dimensional quality standards into a computable single scalar reward signal. 2) The non-stationary learning process: how to design a reward mechanism that adapts to the model's continuously evolving capabilities during training, thereby avoiding reward sparsity in the early stages or capability stagnation in the later stages.

Therefore, the technical novelty of our work lies in the adaptive policy optimization framework we propose to address these challenges: Stock-R1. The construction of this framework is founded on three key design principles. First, we adopt the concept of Axiomatic Reward Construction to solve the reward definition problem. Second, we introduce the high-level principle of Curriculum Learning to handle the dynamic nature of the learning process. Third, we select GRPO as the core algorithm, as its focus on discriminating between superior and inferior outputs aligns perfectly with the nature of our task, which has a deterministic evaluation function. These design principles constitute the core design principles underpinning Stock-R1, and their specific implementations will be detailed in the subsequent sections.

4.2. The Stock-R1 Method

The specific designs of Stock-R1 are the instantiation of the aforementioned theoretical principles within the Group Relative Policy Optimization framework. We materialize this process through an axiomatic reward construction and a progressive curriculum reward strategy, as illustrated in Figure 2.

4.2.1. Preliminaries

Stock-R1 is built upon GRPO, an advanced critic-free reinforcement learning framework. The core idea of GRPO is to guide the model's optimization by directly comparing the relative quality of a set of candidate outputs. Its training process and optimization objective are summarized as

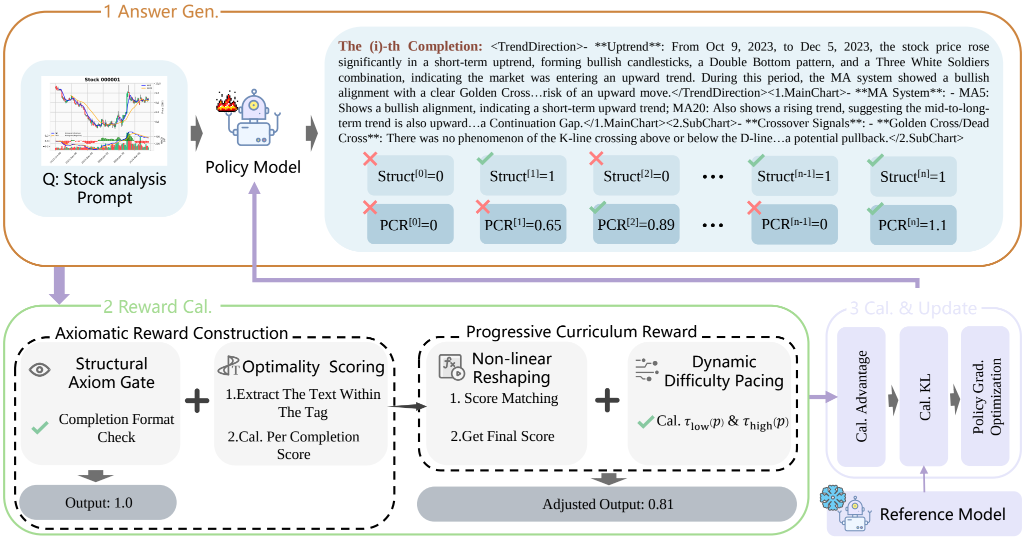

Figure 2. Overall framework of Stock-R1.An illustration of the forward pass in our training pipeline. Given a query (e.g., a stock chart), the policy model generates a group of candidate responses. These responses are then evaluated by our reward function, and the resulting feedback is used to update the policy model via the Group Relative Policy Optimization (GRPO) algorithm.

follows:

1. Batch Sampling and Evaluation: Given an input sample q,an older version of the current policy $\pi_{\theta old}$ is first used to generate a group containing Ncandidate outputs $\{o_1, o_2, ..., o_N\}$. Subsequently, a deterministic reward function evaluates each output to obtain a set of rewards $\{r_1, r_2, ..., r_N\}$.

2. Relative Advantage Calculation:GRPO calculates the relative advantage value A_i for each output o_i by applying Z-score normalization to the rewards within the group.

3. Policy Optimization:Finally, the policy model π_θ is updated by maximizing the following objective function $\mathcal{J}_{GRPO}(\theta)$:

$$\mathcal{J}_{GRPO}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\pi_\theta(o_i|q)}{\pi_{\theta_{old}}(o_i|q)}A_i - \beta\mathcal{KL}\left(\pi_\theta(o_i|q)|\pi_{ref}(o_i|q)\right)\right) \tag{1}$$

This objective function consists of two parts: the first is the policy optimization term, which, by multiplying the importance sampling ratio with the relative advantage value A_i, incentivizes the model to favor the generation of outputs that receive higher relative rewards within the group. The second part is a KL-divergence regularization term, controlled by the hyperparameter β, which ensures that the updated policy π_θ does not deviate too far from a fixed reference model π_{ref}, thereby guaranteeing training stability.

4.2.2. Axiomatic Reward Construction

We realize the concept of axiomatic construction into a composite reward function formulated as a product of two components: $R_{\text{total}}(o_i, s)$. This design clearly divides the evaluation process into two hierarchical levels: first, a determination of validity based on axioms, followed by an exploration of optimality based on an evaluation function. To this end, our composite reward function consists of two parts: a format reward R_{format} and a textual structural similarity reward R_{sim}.

$$R_{\text{total}}(o_i, s) = R_{\text{format}}(o_i) \times R_{\text{sim}}(o_i, s) \tag{2}$$

Here, o_i is the i-th analytical report generated by the model, and s is the corresponding ground truth, which is the expert-calibrated text from the MMSA dataset.

Enforcing the Structural Axiom We define the structured template that the report must follow as a core axiom. This axiom is enforced through a binary format reward, R_{format}. This function checks whether the output contains the three core label pairs: <Trend Direction>, <Main Chart>, and <Sub-chart>. Only when the output fully adheres to this axiom is it considered a valid solution eligible for the next stage of evaluation; otherwise, it is rejected outright. This constitutes the binary filtering of the solution space. It is a binary gating mechanism:

$$R_{\text{format}}(o_i) = \begin{cases} 1, & \text{if match} \\ 0, & \text{else} \end{cases} \tag{3}$$

Quantifying Optimality with the Content Evaluation Function For all valid solutions that have passed the axiom verification, we activate the content evaluation function R_{sim} to quantify their degree of optimality. The creation of this function presupposes the existence of a unified cross-modal representation space. The model must comprehend the input visual chart information within this space and generate the corresponding textual analysis. Our evaluation function, built on this premise, uses weighted cosine similarity to calculate the semantic consistency between the three core parts of the model's output and the corresponding parts of the ground truth. This essentially guides the policy toward the optimal solution most aligned with expert knowledge. It does so by providing a continuous value signal within the valid solution space filtered by the axiom.

$$R_{\text{sim}} = w_{\text{trend}} S_{\text{trend}} + w_{\text{main}} S_{\text{main}} + w_{\text{sub}} S_{\text{sub}} \tag{4}$$

4.2.3. Progressive Curriculum Reward

Through this strategy, we transform the abstract concept of curriculum learning into a concrete, executable algorithmic mechanism. The strategy dynamically and structurally reshapes the entire Incentive Landscape through two synergistic mechanisms, thereby providing the model with the most suitable optimization objectives and gradients at different capability stages.

Non-linear Reshaping of the Incentive Landscape We first address the problem of a uniform gradient in the original reward signal through a tiered reward shaping mechanism. A linear similarity score R_{sim} cannot distinguish the essential difference between "poor" and "poorer," or "good" and "excellent." To this end, we introduce a dynamic elimination threshold $\tau_{\text{low}}(p)$ and an incentive threshold $\tau_{\text{high}}(p)$, deconstructing the continuous reward space into three heterogeneous regions with distinctly different incentive gradients.

$$r_i = \begin{cases} \min\left(1.0 + (R_{\text{sim}} - \tau_{\text{high}}(p)) \cdot \lambda, r_{\max}\right), & \text{if } R_{\text{sim}} \geq \tau_{\text{high}}(p) \quad \text{(incentive)} \\ R_{\text{sim}}, & \text{if } \tau_{\text{low}}(p) \leq R_{\text{sim}} < \tau_{\text{high}}(p) \quad \text{(standard)} \\ 0, & \text{if } R_{\text{sim}} < \tau_{\text{low}}(p) \quad \text{(punish)} \end{cases} \tag{5}$$

In the punish region, the reward is strictly set to 0. This creates a strong negative incentive, theoretically establishing a clear quality baseline and efficiently pruning completely unacceptable, low-quality outputs from the exploration space. In the standard region, the reward remains consistent with the original similarity score R_{sim}, providing a baseline, linear positive feedback. When the output quality surpasses the incentive threshold, we introduce an excess reward mechanism. The reward signal here is designed to grow linearly, with the growth rate controlled by an excess reward coefficient λ. This creates a high-return zone in the incentive region, intended to provide a steep optimization gradient that strongly attracts the model to explore and converge towards excellent performance.

Dynamic Difficulty Pacing of the Learning Trajectory To set the curriculum in motion, we further designed a progressive threshold adaptation mechanism. This mechanism is a direct embodiment of the curriculum learning philosophy, ensuring that the structure of the incentive landscape is not static but matches the model's ever-increasing capabilities. Specifically, the elimination threshold $\tau_{\text{low}}(p)$ and the incentive threshold $\tau_{\text{high}}(p)$ increase monotonically with the normalized training progress p:

$$\tau_{\text{low}}(p) = \tau_{\text{low}_{\text{init}}} + (\tau_{\text{low}_{\text{final}}} - \tau_{\text{low}_{\text{init}}}) \cdot \tag{6}$$

$$\tau_{\text{high}}(p) = \tau_{\text{high}_{\text{init}}} + \left(\tau_{\text{high}_{\text{final}}} - \tau_{\text{high}_{\text{init}}}\right) \cdot \tag{7}$$

This rising tide lifts all boats approach to difficulty adjustment transforms the training process into a continuous, adaptive challenge. It theoretically ensures that the model receives meaningful and appropriately difficult learning tasks at any stage of its development. This guides it to transition smoothly from learning the basic format to pursuing content excellence, and effectively prevents reward hacking that can occur when the model is satisfied with early, lower standards, thereby promoting steady and continuous capability improvement.

5. Experiments

5.1. Experimental Setup

Dataset Split: All experiments were conducted on our constructed MMSA dataset. To ensure a fair and reliable evaluation, we randomly split the entire 20,000 samples according to an 8:1:1 ratio. This resulted in a training set of 16,000 samples for the model's parameter learning; a validation set of 2,000 samples for selecting the best model checkpoint and tuning key hyperparameters; and a test set of 2,000 samples. This test partition was completely held out from the training and model selection processes, used only once at the end for the final performance evaluation of all models.

Implementation Details: All our reinforcement learning-based experiments were completed within the Stock-R1 framework. To ensure fairness, we selected Qwen2.5-VL-3B as the base model for all comparative experiments. The model was trained for a total of 2 epochs using the AdamW optimizer with a learning rate set to 1e-6. All experiments were conducted on 4 NVIDIA GeForce RTX 4090 GPUs.

5.2. *Evaluation Metrics*

Given that financial chart-and-text analysis is an emerging task, there is a lack of established evaluation standards. To this end, we designed a comprehensive, multi-dimensional evaluation system. The validity of this system rests on a core premise: all our automated metrics are intended to quantify the consistency between the model's output and the ground-truth reports in the MMSA dataset. These ground-truth reports have, in turn, been rigorously validated by a panel of financial experts and are thus representative of a high-quality, professional standard of analysis. Therefore, achieving a higher score on our designed metrics can be considered strong evidence that a model's analytical capabilities are approaching an expert level.

This evaluation system is specifically centered around the following three core capabilities: Trend Analysis, Main Analysis, and Sub Analysis. It integrates temporal localization accuracy and the precision of multi-class technical signal recognition based on keyword matching. The specific definitions, matching rules, and calculation methods for all metrics are provided in Appendix B.

Table 3. Overall Performance Comparison of All Models on the MMSA Test Set

Model	Trend analysis	Main analysis	Sub analysis
	F1-Macro/Avg. IoU	F1-Score	Avg. F1-Macro
TBAC-VLR1-3B	0.2007/0.7592	0.3963	0.3762
InternVL2.5-4B	0.0549/0.6767	0.4546	0.4138
Qwen2VL-7B	0.0655/0.6893	0.3681	0.4340
Qwen2.5VL-7B	0.2093/0.6918	0.5235	0.4691
Qwen2.5VL-3B	0.0925/0.6991	0.3354	0.3061
+time-series-based SFT	0.1425/0.7018	0.2798	0.2554
+SFT	0.0670/0.6783	0.3326	0.3080
+Stock-R1	**0.2677**(↑0.1752)/**0.7580**(↑0.0589)	**0.6423**(↑0.3069)	**0.5144**(↑0.2083)

5.3. *Main Results and Analysis*

To comprehensively evaluate the effectiveness of the Stock-R1 method, we compared it against three categories of models on the MMSA test set: 1. our proposed Stock-R1; 2. Method Baselines, including SFT-only, a time-series-based SFT model, and Zero-shot versions of Qwen2.5-VL-3B, to precisely measure the effective gains of our reinforcement learning framework; and 3. leading SOTA models, covering several industry-leading models such as TBAC-VLR1-3B-preview, InternVL2_5-4B, and the larger-parameter Qwen2.5-VL-7B.

Our analysis begins with a fundamental validation of our choice of input modality by comparing the performance of the chart-based SFT model with the text-based Time-Series baseline. As presented in Table 3, the results reveal a nuanced but clear picture.

Notably, the Time-Series model achieves a competitive F1-score and IoU in Trend Analysis (0.1425 / 0.7018), demonstrating that a language model can effectively learn to infer general trends and temporal segments directly from ordered numerical sequences. However, a significant performance degradation is observed in tasks requiring the identification of complex visual patterns. The F1-scores for Main Analysis and Sub Analysis dropped to 0.2798 and 0.2554, respectively, underperforming the chart-based SFT model.

This divergence in performance is highly informative. It suggests that while sequential numerical data is sufficient for basic trend recognition, it is a less effective modality for identifying patterns whose definitions are inherently visual-spatial, such as chart formations (e.g., 'Double Top') and indicator divergences. The 2D chart representation provides a crucial visual abstraction that makes these complex patterns more tractable for the model to recognize. This empirically validates our hypothesis that for the task of generating a comprehensive, expert-level analysis report, the visual modality is a critical component for capturing the full spectrum of technical signals.

Table 4. Detailed Results on the Trend and Main Analysis Tasks

Model	Trend analysis				Main analysis		
	Prec.	Rec.	F1	Avg. IoU	Prec.	Rec.	F1
TBAC-VLR1-3B	0.2069	0.2127	0.2007	**0.7592**	0.7568	0.2684	0.3963
InternVL2. 5-4B	0.0599	0.0575	0.0549	0.6767	0.6513	0.3491	0.4546
Qwen2-VL-7B	0.0693	0.0696	0.0655	0.6893	0.5273	0.2828	0.3681
Qwen2. 5-VL-7B	0.2112	0.2242	0.2093	0.6918	0.5623	0.4896	0.5235
Qwen2.5VL-3B	0.0991	0.0988	0.0925	0.6991	0.6553	0.2254	0.3354
+SFT	0.0775	0.0687	0.0670	0.6783	0.6477	0.2238	0.3326
+Stock-R1	**0.2663**	**0.2801**	**0.2677** (↑0.1752)	0.7580	**0.8811**	**0.5053**	**0.6423** (↑0.3069)

Overall Performance Analysis Our proposed Stock-R1 demonstrated outstanding performance across all evaluation dimensions. Compared to the baseline, Stock-R1 achieved significant F1-score improvements of 0.17, 0.30, and 0.20 in trend, main, and sub analysis, respectively. This provides compelling evidence for the substantial superiority of our proposed reinforcement learning framework and PCR strategy. Furthermore, Stock-R1 not only performed best at the 3B parameter level but was also highly competitive against 7B-level SOTA models, surpassing Qwen2-VL-7B-Instruct on several key metrics and achieving performance comparable to the top-performing Qwen2.5-VL-7B. This fully demonstrates the excellent parameter efficiency and advanced nature of our method.

Trend Analysis Dimension Stock-R1 exhibited a comprehensive advantage. As shown in Table 4, it not only achieved the highest F1-score but also led in the average IoU metric, indicating that it reached the highest precision in both semantic trend judgment and temporal localization. In contrast, most general-purpose LVLMs, without specific optimization like Stock-R1, struggle to independently learn reliable chart trend analysis capabilities. Their main errors manifest as an inability to generate valid trend interval judgments or misclassifying the trend type after localization.

Table 5. Detailed Results on the Sub-chart Analysis Task, Broken Down by Indicator. The models are abbreviated as follows: TBAC (TBAC-VLR1-3B), IVL2 (InternVL2. 5-4B), QW2-7B (Qwen2-VL-7B), QW2.5-7B (Qwen2.5-VL-7B), QW2.5-3B (Qwen2.5VL-3B).

Indicator	Metric	TBAC	IVL2	QW2-7B	QW2.5-7B	QW2.5-3B	+SFT	+Stock-R1
KDJ	Prec.	0.8104	0.7414	0.7330	0.7483	0.7449	0.7455	**0.9746**
	Rec.	0.1814	0.3274	0.2297	**0.3519**	0.1527	0.1726	0.3186
	F1	0.2964	0.4542	0.3498	0.4787	0.2535	0.2802	**0.4803** (↑0.2268)
MACD	Prec.	0.8849	0.9064	0.8434	0.8606	0.8990	0.8750	**0.9648**
	Rec.	0.4428	0.4318	0.4921	0.4437	0.3945	0.3792	**0.6131**
	F1	0.5903	0.5850	0.6215	0.5855	0.5484	0.5291	**0.7497** (↑0.2013)
RSI	Prec.	**0.7643**	0.7060	0.7401	0.7184	0.6310	0.7101	0.6616
	Rec.	0.2425	0.2010	0.2772	0.2977	0.1315	0.1398	**0.3061**
	F1	0.3682	0.3129	0.4033	**0.4209**	0.2176	0.2336	0.4185 (↑0.2009)
BOLL	Prec.	0.6464	0.5471	0.5238	0.6163	0.6094	0.5574	**0.6517**
	Rec.	0.1547	0.2095	0.2757	0.2865	0.1231	0.1137	**0.2983**
	F1	0.2497	0.3030	0.3613	0.3912	0.2048	0.1889	**0.4093** (↑0.2045)
	Avg. F1	0.3762	0.4138	0.4340	0.4691	0.3061	0.3080	**0.5144** (↑0.2083)

Main Analysis Dimension Stock-R1 achieved the best overall performance while ensuring high reliability. As can be seen in Table 4, although its recall was slightly lower than that of Qwen2.5-VL-7B, indicating that the model adopts a more conservative approach in signal generation. its extremely high precision ensured the reliability of the output, ultimately resulting in the highest F1-score among all compared methods.

Sub Analysis Dimension Stock-R1 also showed a significant advantage and presented interesting trade-offs. As detailed in Table 5, its performance improved remarkably when processing MACD and BOLL. For the KDJ and RSI indicators, while its F1-scores were competitive with models like Qwen2.5-VL-7B, with each having its own strengths, the performance improvement over the baseline was still substantial, proving our method is equally effective for various oscillator-type indicators.

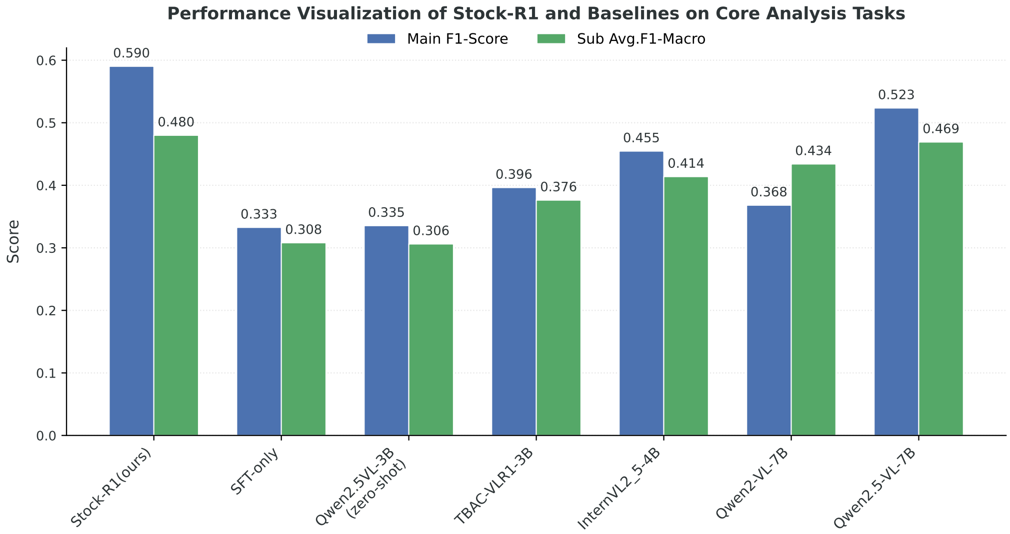

Figure 3. F1-Score Comparison of All Models on Main and Sub Analysis Tasks.

In summary, despite minor trade-offs on individual sub-metrics, the substantial improvements of Stock-R1 on key analytical tasks demonstrate its outstanding capability and immense potential as an advanced financial analysis support tool. A performance comparison of all models across the main dimensions is illustrated in Figure 3.

5.4. Ablation Study

To systematically validate the effectiveness of the core designs within our proposed Stock-R1 method—namely, the Progressive Curriculum Reward strategy and the fine-grained reward signals—we conducted a series of detailed ablation studies using the complete Stock-R1 model as the benchmark. For this purpose, we designed three core variants to quantitatively assess the contribution of each component: the first variant, w/o Progressive Curriculum Reward, removes the entire Progressive Curriculum Reward module and directly uses the original composite reward signal R_{total}; the second variant, w/o Progression, retains the reward/penalty region division of Progressive Curriculum Reward but removes its progressive mechanism by using fixed reward thresholds; the final variant, w/o Weighted Reward, calculates the similarity reward R_{sim} without applying weights to the different analytical dimensions. The performance comparison of all variant models against the full model and the SFT baseline is shown in Table 6.

Table 6. Ablation Study of the Core Components of Stock-R1

Model	Trend analysis	Main analysis	Sub analysis
	F1-Macro/Avg. IoU	F1-Score	Avg. F1-Macro
SFT-only	0.067/0.6783	0.3326	0.308
w/o Progressive Curriculum Reward	0.2454/0.7511	0.5902	0.4492
w/o Progression	0.2540/0.7535	0.6080	0.4750
w/o Weighted Reward	0.2615/0.7560	0.6250	0.4980
Stock-R1 (ours)	**0.2677/0.7580**	**0.6423**	**0.5144**

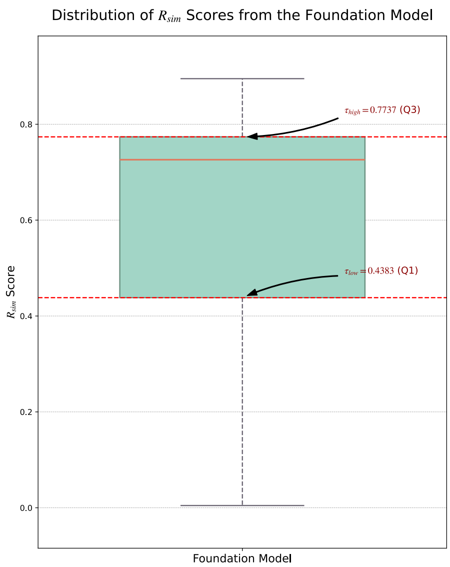

Figure 4. Distribution of similarity scores and justification for initial threshold selection.

Overall Effectiveness of the Progressive Curriculum Reward Strategy First, the w/o Progressive Curriculum Reward variant, which completely removes the PCR module, led to the most significant performance degradation. As shown in Table 6, its Sub-chart Analysis Avg. F1-Macro score dropped sharply from the full model's 0.5144 to 0.4492, which directly proves that the Progressive Curriculum Reward strategy is the cornerstone of our entire method framework.

Necessity of the Dynamic Difficulty Pacing Mechanism Second, the w/o Progression variant isolates the contribution of the progressive mechanism by employing fixed reward thresholds. Specifically, we set the thresholds $\tau_{\text{low}}(p)$ and $\tau_{\text{high}}(p)$ to the 25th and 75th percentiles of the base model's R_{sim} score distribution (i.e., 0.4383 and 0.7737, respectively), as illustrated in Figure 4. Although this variant's performance was superior to that of the w/o PCR model, it still showed a significant gap when compared to the complete model. This result precisely underscores the necessity of a dynamic difficulty curriculum for enabling continuous breakthroughs in the model's capability.

Value of Fine-grained Reward Signals Finally, the w/o Weighted Reward variant, which removed reward weighting, achieved performance closest to the full model. This result indicates that a fine-grained evaluation strategy that mimics human experts' emphasis on different analytical dimensions is an effective enhancement for achieving optimal model performance.

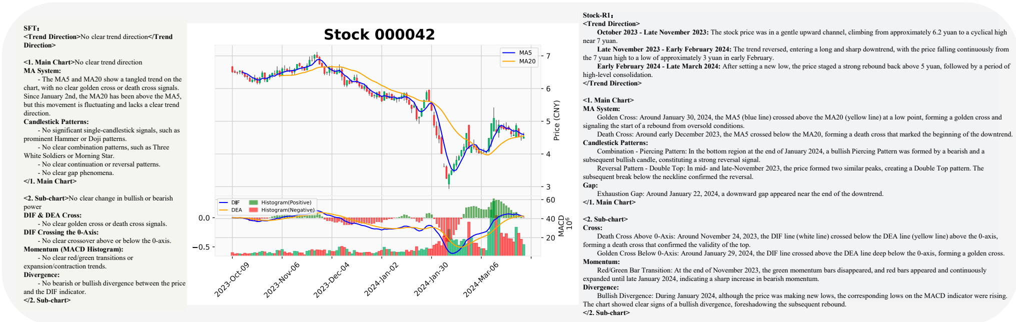

Figure 5. A case study comparison between the SFT-only model and our Stock-R1 model. The SFT model provides an evasive and ineffective analysis, whereas Stock-R1 delivers a detailed, multi-faceted, and expert-level report on the same chart, identifying specific patterns like the Double Top and Bullish Divergence.

5.5. *Qualitative Analysis and Case Study*

Beyond macro-level quantitative evaluations, micro-level qualitative analysis can more profoundly reveal the fundamental differences in models' analytical capabilities. Figure 5 presents a typical case selected from the test set to illustrate the superiority of Stock-R1 over the traditional SFT method. Although the baseline SFT model is syntactically coherent, it exhibits significant failures in semantic interpretation. Its output is replete with non-committal phrases such as "no clear trend direction," failing to extract any actionable, decision-relevant information from the visual data. This is, in essence, an ineffective analysis.

In stark contrast, the report generated by Stock-R1 demonstrates a complex and logically rigorous analytical process. It begins by accurately dividing the overall price movement into three macro-stages: a gentle rise, a sharp decline, and a subsequent rebound. More importantly, it showcases a profound ability to identify key turning points by synthesizing multidimensionalsignals. For the market top around November 2023, it accurately identified the bearish reversal signal, confirmed jointly by a Double Top pattern and an MACD death cross above the zero axis. Furthermore, during the bottom formation process in January 2024, it detected a confluence of bullish signals, which included not only a Piercing Pattern on the candlesticks and an MACD golden cross at a low level, but also the most critical phenomenon signaling a market reversal: an MACD Bullish Divergence.

The model's ability to identify a complex signal like "bullish divergence," which requires long-term associative analysis of both price and indicator trends, serves as compelling evidence for the effectiveness of our reinforcement learning framework. Our designed Axiomatic Reward and Progressive Curriculum Learning strategies have evidently pushed the model beyond superficial pattern matching, enabling it to develop a deeper and more holistic understanding of financial analysis logic. This case clearly illustrates that Stock-R1 has transitioned from a simple chart descriptor to a decision support tool with the capabilities of a junior analyst, validating its potential for application in complex financial scenarios.

6. Conclusion

In this paper, we addressed the lack of professional decision support tools in the financial domain. We pioneered a new path for automated financial analysis through two core contributions. First, we constructed MMSA, the first large-scale, high-quality multimodal stock analysis dataset. Second, we proposed a novel reinforcement learning method, Stock-R1. This method efficiently guides a model to learn complex analytical tasks by materializing axiomatic principles into a reward function and implementing a Progressive Curriculum Reward strategy to dynamically shape the incentive landscape. Experimental results demonstrate that Stock-R1 is significantly superior to various baselines and leading SOTA models in generating high-quality analytical reports. This not only validates the effectiveness of our method but also successfully transitions the research paradigm. Future work will extend this approach toward broader datasets, richer modalities, and more interactive analytical systems.

Author Contributions

Jie Sun: Conceptualized the research vision; designed the core multimodal architecture; developed key algorithm modules;

Tengxiang Luo: MMSA dataset creation; Methodology; Writing-original draft;

Huamao Gu: Optimized training pipelines; contributed to results analysis and visualization;

Jing Hua: Contributed to results/discussion sections; revised manuscript critically.

Rui Xie: Provide multimodal experimental data, as well as professional interpretation opinions on K-line chart interpretation.

Zuohua Ding: Provided critical intellectual guidance; edited manuscript structure; reviewed and approved final submission.

References

[1] Yejun Soun, Jaemin Yoo, Minyong Cho, Jihyeong Jeon, and U Kang. Accurate stock movement prediction with self-supervised learning from sparse noisy tweets. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 1691–1700. IEEE, 2022.

[2] Yumo Xu and Shay B Cohen. Stock movement prediction from tweets and historical prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1970–1979, 2018.

[3] Huizhe Wu, Wei Zhang, Weiwei Shen, and Jun Wang. Hybrid deep sequential modeling for social text-driven stock prediction. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pages 1627–1630, 2018.

[4] Julio Cesar Salinas Alvarado, Karin Verspoor, and Timothy Baldwin. Domain adaption of named entity recognition to support credit risk assessment. In *Proceedings of the australasian language technology association workshop 2015*, pages 84–90, 2015.

[5] Agam Shah, Ruchit Vithani, Abhinav Gullapalli, and Sudheer Chava. Finer: Financial named entity recognition dataset and weak-supervision model. *arXiv e-prints*, pages arXiv–2302, 2023.

[6] Soumya Sharma, Tapas Nayak, Arusarka Bose, Ajay Kumar Meena, Koustuv Dasgupta, Niloy Ganguly, and Pawan Goyal. Finred: A dataset for relation extraction in financial domain. In *Companion Proceedings of the Web Conference 2022*, pages 595–597, 2022.

[7] Dominique Mariko, Hanna Abi Akl, Estelle Labidurie, Stephane Durfort, Hugues De Mazancourt, and Mahmoud El-Haj. Financial document causality detection shared task (fincausal 2020). *arXiv preprint arXiv:2012.02505*, 2020.

[8] Qianqian Xie, Weiguang Han, Zhengyu Chen, Ruoyu Xiang, Xiao Zhang, Yueru He, Mengxi Xiao, Dong Li, Yongfu Dai, Duanyu Feng, et al. Finben: A holistic financial benchmark for large language models. *Advances in Neural Information Processing Systems*, 37:95716–95743, 2024.

[9] Gagan Bhatia, El Moatez Billah Nagoudi, Hasan Cavusoglu, and Muhammad Abdul-Mageed. Fintral: A family of gpt-4 level multimodal financial large language models. *arXiv preprint arXiv:2402.10986*, 2024.

[10] Ziao Wang, Yuhang Li, Junda Wu, Jaehyeon Soon, and Xiaofeng Zhang. Finvis-gpt: A multimodal large language model for financial chart analysis. *arXiv preprint arXiv:2308.01430*, 2023.

[11] Jinze Bai, Shuai Bai, Shusheng Yang, Shijie Wang, Sinan Tan, Peng Wang, Junyang Lin, Chang Zhou, and Jingren Zhou. Qwen-vl: A versatile vision-language model for understanding, localization, text reading, and beyond. arxiv 2023. *arXiv preprint arXiv:2308.12966*, 1(8), 2023.

[12] Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, et al. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24185–24198, 2024.

[13] Liangyu Chen, Sijia Chen, Siyuan Huang, Chen-Hsin Lee, Yixuan Wang, Jindong Chen, Xin-Yu Zhang, Zheyang Li, He-Yen Hsieh, Celina Han, Hong-Xin Chen, Siran Chen, Shiyu Hu, Xi Chen, Yu Kang, E Chen, Dahua Lin, C.-C. Jay Kuo, and Fengbo Zheng. Mimicit: Multi-modal In-Context Instruction Tuning. *arXiv preprint arXiv:2306.05425*, jun 2023.

[14] Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26296–26306, 2024.

[15] Peter Tong, Ellis Brown, Penghao Wu, Sanghyun Woo, Adithya Jairam Vedagiri IYER, Sai Charitha Akula, Shusheng Yang, Jihan Yang, Manoj Middepogu, Ziteng Wang, et al. Cambrian-1: A fully open, vision-centric exploration of multimodal llms. *Advances in Neural Information Processing Systems*, 37:87310–87356, 2024.

[16] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916, 2023.

[17] Jianfeng Dong, Xianke Chen, Minsong Zhang, Xun Yang, Shujie Chen, Xirong Li, and Xun Wang. Partially relevant video retrieval. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 246–257, 2022.

[18] Jianfeng Dong, Yabing Wang, Xianke Chen, Xiaoye Qu, Xirong Li, Yuan He, and Xun Wang. Reading-strategy inspired visual representation learning for text-to-video retrieval. *IEEE transactions on circuits and systems for video technology*, 32(8):5680–5694, 2022.

[19] Haiyang Mei, Yuanyuan Liu, Ziqi Wei, Dongsheng Zhou, Xiaopeng Wei, Qiang Zhang, and Xin Yang. Exploring dense context for salient object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3):1378–1389, 2021.

[20] Dan Guo, Kun Li, Bin Hu, Yan Zhang, and Meng Wang. Benchmarking micro-action recognition: Dataset, method, and application. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(7):6238–6252, 2024.

[21] Zixiang Chen, Yihe Deng, Huizhuo Yuan, Kaixuan Ji, and Quanquan Gu. Self-play fine-tuning converts weak language models to strong language models. *arXiv preprint arXiv:2401.01335*, 2024.

[22] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

[23] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36:53728–53741, 2023.

[24] Yuhao Dong, Zuyan Liu, Hai-Long Sun, Jingkang Yang, Winston Hu, Yongming Rao, and Ziwei Liu. Insight-v: Exploring long-chain visual reasoning with multimodal large language models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 9062–9072, 2025.

[25] Zhiqing Sun, Sheng Shen, Shengcao Cao, Haotian Liu, Chunyuan Li, Yikang Shen, Chuang Gan, Liang-Yan Gui, Yu-Xiong Wang, Yiming Yang, et al. Aligning large multimodal models with factually augmented rlhf. *arXiv preprint arXiv:2309.14525*, 2023.

[26] Tianyi Xiong, Xiyao Wang, Dong Guo, Qinghao Ye, Haoqi Fan, Quanquan Gu, Heng Huang, and Chunyuan Li. Llava-critic: Learning to evaluate multimodal models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 13618–13628, 2025.

[27] Yi-Fan Zhang, Tao Yu, Haochen Tian, Chaoyou Fu, Peiyan Li, Jianshu Zeng, Wulin Xie, Yang Shi, Huanyu Zhang, Junkang Wu, et al. Mm-rlhf: The next step forward in multimodal llm alignment. *arXiv preprint arXiv:2502.10391*, 2025.

[28] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

[29] Daya Guo, Qihao Zhu, Dejian Yang, Zhenda Xie, Kai Dong, Wentao Zhang, Guanting Chen, Xiao Bi, Yu Wu, YK Li, et al. Deepseek-coder: When the large language model meets programming–the rise of code intelligence. *arXiv preprint arXiv:2401.14196*, 2024.

[30] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Y Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

[31] Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024.

[32] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48, 2009.

[33] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[34] Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2024.

[35] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

[36] Bo Li, Yuanhan Zhang, Dong Guo, Renrui Zhang, Feng Li, Hao Zhang, Kaichen Zhang, Peiyuan Zhang, Yanwei Li, Ziwei Liu, et al. Llava-onevision: Easy visual task transfer. *arXiv preprint arXiv:2408.03326*, 2024.

[37] Zhe Chen, Weiyun Wang, Yue Cao, Yangzhou Liu, Zhangwei Gao, Erfei Cui, Jinguo Zhu, Shenglong Ye, Hao Tian, Zhaoyang Liu, et al. Expanding performance boundaries of open-source multimodal models with model, data, and test-time scaling. *arXiv preprint arXiv:2412.05271*, 2024.

[38] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025.

[39] Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, et al. Mmmu: A massive multi-discipline multimodal understanding and reasoning benchmark for expert agi. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9556–9567, 2024.

[40] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[41] Qianqian Xie, Weiguang Han, Xiao Zhang, Yanzhao Lai, Min Peng, Alejandro Lopez-Lira, and Jimin Huang. Pixiu: A large language model, instruction data and evaluation benchmark for finance. *arXiv preprint arXiv:2306.05443*, 2023.

[42] Shu Liu, Shangqing Zhao, Chenghao Jia, Xinlin Zhuang, Zhaoguang Long, Jie Zhou, Aimin Zhou, Man Lan, Qingquan Wu, and Chong Yang. Findabench: Benchmarking financial data analysis ability of large language models. *arXiv preprint arXiv:2401.02982*, 2024.

[43] Jialin Chen, Aosong Feng, Ziyu Zhao, Juan Garza, Gaukhar Nurbek, Cheng Qin, Ali Maatouk, Leandros Tassiulas, Yifeng Gao, and Rex Ying. Mtbench: A multimodal time series benchmark for temporal reasoning and question answering. *arXiv preprint arXiv:2503.16858*, 2025.

Author Biography

Jie Sun Ph.D. at Zhejiang Sci-Tech University, with research interests in computer vision and multimodal large models.

ORCID: 0000-0002-5196-7268
Email: sunjie@zjgsu.edu.cn

Tengxiang Luo is a master's student of School of Computer Science and Technology, Zhejiang Gongshang University, with research interests in Multimodal model.

ORCID: 0009-0006-0568-2874
Email: 23020100041@pop.zjgsu.edu.cn

Huamao Gu Professor and Ph.D. at Zhejiang Gongshang University, with research expertise in natural language processing and large language models.

Email: ghmsjq@zjgsu.edu.cn

Jing Hua Professor, Ph.D., and Ph.D. supervisor at Zhejiang Gongshang University, with research expertise in computer vision and large language models.

Email: jhua@zjgsu.edu.cn

Rui Xie Master's degree holder, engineer, General Manager of Zhejiang Zheshiyou Comprehensive Energy Sales Co., Ltd., engaged in research and development in the field of energy management.

Email: 13867111766@139.com

Zuohua Ding Professor, Ph.D., and Ph.D. supervisor at Zhejiang Sci-Tech University, with research expertise in software engineering, requirements modeling & analysis, and intelligent software systems & service robotics.

Email: zouhuading@hotmail.com