CAN LARGE MULTIMODAL MODELS THINK JUST ENOUGH? EVALUATING OUTPUT-SIDE ADAPTIVE REASONING IN LARGE MULTIMODAL MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Explicit reasoning can improve the performance of large multimodal models (LMMs) on complex visual understanding and reasoning tasks, but lengthy and unnecessary reasoning also incurs additional generation cost. Although recent output-side adaptive reasoning methods have shown potential for reducing unnecessary generation cost, their evaluation remains fragmented and inconsistent. In this work, we introduce **MARE-Bench**, a benchmark for systematically evaluating output-side adaptive reasoning methods for LMMs. MARE-Bench provides unified evaluation protocols across seven representative datasets spanning five evaluation dimensions: General Knowledge, Mathematical Reasoning, Logical Reasoning, Medical QA, and OCR. It covers seven representative adaptive reasoning methods and three LMM families. We analyze adaptive reasoning methods in terms of task performance, average output tokens, and reasoning behaviors, characterizing the performance–generation cost trade-off. Our experiments reveal several findings: (1) some methods outperform their base models, while others perform poorly; (2) FAST-GRPO achieves competitive overall performance across the evaluated setings in most cases; (3) the effects of adaptive reasoning vary across evaluation dimensions, with larger performance gains on Mathematical Reasoning and Logical Reasoning tasks, smaller gains in General Knowledge and Medical QA, and performance degradation on OCR; and (4) the performance–generation cost trade-off of adaptive reasoning methods can vary across base models. We believe that MARE-Bench will serve as a reliable foundation for future research on efficient adaptive reasoning in LMMs.

1 INTRODUCTION

In recent years, explicit reasoning, exemplified by Chain-of-Thought (CoT), has been widely adopted in large multimodal models (LMMs), substantially improving their performance on complex visual understanding and reasoning tasks (Wei et al., 2022; Kojima et al., 2022; Wang et al., 2023; Shao et al., 2024; Dong et al., 2025; Gao et al., 2025). However, lengthy and unnecessary reasoning incurs additional generation cost and substantially increases inference overhead (Wang et al., 2025a; Li et al., 2025). This raises a key question: **how can LMMs reason only when necessary, preserving their reasoning capabilities while reducing unnecessary generation cost?**

To address this question, recent studies have begun to explore output-side control of reasoning behavior in LMMs (Wang et al., 2025a; Yang et al., 2026; Chen et al., 2026; Zhang et al., 2026b). Among them, mode selection-based methods choose between explicit reasoning and direct answering (Wang et al., 2025a; Yang et al., 2026). Reasoning effort allocation-based methods adjust exploration effort or reasoning path length according to problem characteristics (Chen et al., 2026; Xiao & Gan, 2025). Reasoning expression optimization-based methods compress or restructure the reasoning process (Zhang et al., 2026b; Wang et al., 2026). Together, these approaches provide multiple ways for LMMs to adapt their reasoning behavior on demand. Related work also treats reasoning depth as an explicit inference decision through search over candidate traces (Yao et al., 2023a).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Despite these advances, output-side adaptive reasoning methods for LMMs still lack unified and systematic evaluation. First, most studies evaluate only a small number of adaptive reasoning methods on a limited set of datasets, making it difficult to comprehensively understand how different adaptive reasoning strategies perform across diverse task scenarios (Wang et al., 2025a; Xiao & Gan, 2025; Zhang et al., 2026b). Second, existing studies typically conduct experiments using their own evaluation pipelines, with inconsistent reasoning configurations, scoring metrics, and other settings, making reported results difficult to compare directly across methods (Wang et al., 2025a; Yang et al., 2026; Yuan et al., 2025). Third, existing evaluations mainly focus on task performance and output length, while paying less attention to whether reducing reasoning output induces different forms of performance degradation and whether additional reasoning effort can be translated into performance gains on complex reasoning tasks (Xiao & Gan, 2025; Zhang et al., 2026b; Chen et al., 2026).

To address these limitations, we introduce the **M**ultimodal **A**daptive **R**easoning **E**valuation **Bench**mark (**MARE-Bench**), a benchmark for systematically evaluating output-side adaptive reasoning methods for LMMs. As shown in Figure 1, MARE-Bench covers seven representative datasets across five evaluation dimensions: General Knowledge, Mathematical Reasoning, Logical Reasoning, Medical QA, and OCR. We evaluate seven representative output-side adaptive reasoning methods spanning three categories: Mode Selection-Based Methods, Reasoning Effort Allocation-Based Methods, and Reasoning Expression Optimization-Based Methods. MARE-Bench further covers three LMM families and adopts unified evaluation protocols, establishing a unified and systematic comparison framework across different methods. Beyond task performance, MARE-Bench uses average output tokens as a unified measure of output-side generation cost.

Building on these metrics, we further analyze adaptive reasoning methods from three perspectives: Reasoning Effort Adaptivity, Output Budget Sensitivity, and Error Attribution. These analyses provide a deeper characterization of reasoning behaviors and failure modes across methods.

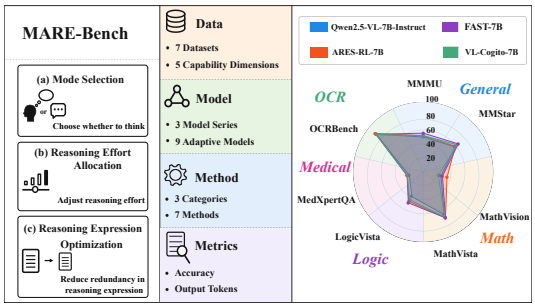

Figure 1: Overview of MARE-Bench and representative adaptive reasoning results across datasets.

Large-scale experiments reveal several key findings: (1) in terms of Average Accuracy (Avg. Acc.) and average output tokens, some methods outperform their base models, while others perform poorly; (2) FAST-GRPO achieves competitive overall performance across the evaluated setings in most cases; (3) the effects of adaptive reasoning vary across evaluation dimensions, with larger performance gains on Mathematical Reasoning and Logical Reasoning tasks, smaller gains in General Knowledge and Medical QA, and performance degradation on OCR; and (4) the performance–generation cost trade-off of adaptive reasoning methods can vary across base models.

Overall, our main contributions are as follows:

- We introduce **MARE-Bench**, a benchmark for systematically evaluating output-side adaptive reasoning methods for LMMs, covering diverse task scenarios, representative adaptive reasoning paradigms, and multiple LMM families under a unified evaluation protocol.
- We establish a multidimensional evaluation framework for output-side adaptive reasoning. Beyond analyzing task performance and average output tokens, MARE-Bench further examines differences in reasoning behavior through Reasoning Effort Adaptivity, Output Budget Sensitivity, and Error Attribution. We also conduct large-scale experiments on MARE-Bench and provide several findings that offer insights for future research on output-side adaptive reasoning methods for LMMs.

2 Related Work

Large Multimodal Models. Large multimodal models (LMMs) have made substantial progress in visual understanding and cross-modal reasoning tasks (Peng et al., 2024; Chen et al., 2024b; Lin

et al., 2024). Mainstream LMMs typically consist of a visual encoder, a large language model, and a cross-modal connector that aligns visual representations with the language space (Alayrac et al., 2022; Li et al., 2023; Liu et al., 2023; Dai et al., 2023; Zhang et al., 2025b). On the vision side, pretrained models such as ViT and CLIP are commonly used as visual backbones (Dosovitskiy et al., 2021; Radford et al., 2021), while decoder-only LLMs are widely adopted as language backbones in LMMs, as illustrated by LLaVA and Qwen-based models (Liu et al., 2023; Yang et al., 2024).

Building on this architecture, LMMs have continued to advance toward stronger visual perception, long-context modeling, and broader modality coverage. Qwen2.5-VL adopts native dynamic resolution and improves visual grounding, document parsing, and long-video understanding (Bai et al., 2025b); Qwen3-VL further strengthens long-context and spatiotemporal modeling capabilities (Bai et al., 2025a); and Qwen2.5-Omni extends multimodal input coverage to text, images, audio, and video (Xu et al., 2025a). Meanwhile, Mono-InternVL explores a more unified multimodal architecture through intrinsic visual pre-training (Luo et al., 2025). Recent work further extends LMMs to spatial understanding, visual tool use, and complex multi-step reasoning (Cai et al., 2026; Guo et al., 2026; Zhang et al., 2026a).

Reasoning Evaluation for Large Multimodal Models. As the explicit reasoning capabilities of LMMs continue to advance, evaluations of multimodal reasoning have expanded beyond final-answer correctness to examine reasoning processes and behaviors. Broader multimodal benchmarks extend coverage across visual, temporal, and general capability dimensions (Fu et al., 2025a; Li et al., 2024a;b; Fu et al., 2025b). MME-Reasoning covers three types of logical reasoning: inductive, deductive, and abductive reasoning (Yuan et al., 2026); MME-CoT evaluates multimodal Chain-of-Thought reasoning in terms of reasoning quality, robustness, and efficiency (Jiang et al., 2025); and MMMR further analyzes reasoning traces in terms of relevance, consistency, and structured errors, while examining reasoning failures (Tie et al., 2025). In addition, RH-Bench investigates the relationship between increased reasoning length, reduced use of visual information, and increased hallucination (Xu et al., 2025b). Other studies systematically compare the effects of explicit reasoning on perception tasks and complex reasoning tasks (Xu et al., 2025b; Jin et al., 2026).

Recent evaluations have also begun to focus on adaptive and efficient reasoning. AdaptMMBench examines mode selection between text-based reasoning and visual tool-augmented reasoning, together with key-step coverage, tool effectiveness, and computational efficiency (Zhang et al., 2026c). EffiReason-Bench, targeting large language models, compares efficient reasoning methods across different paradigms under a unified setting and introduces the Efficiency-Effectiveness Equilibrium Score (E^3-Score) to measure the trade-off between reasoning effectiveness and efficiency (Huang et al., 2025). Overall, existing multimodal evaluations primarily focus on reasoning processes, specific reasoning phenomena, or a single adaptive mechanism, while unified cross-paradigm comparisons of adaptive reasoning remain largely concentrated on large language models. A unified and systematic evaluation of output-side adaptive reasoning methods for LMMs therefore remains lacking. Tool-augmented reasoning further extends the reasoning process by incorporating external actions and tool use (Yao et al., 2023b; Schick et al., 2023).

3 MARE-BENCH

This section describes the methods, datasets, and metrics used in MARE-Bench.

3.1 OUTPUT-SIDE ADAPTIVE REASONING METHODS

Because reinforcement learning outcomes can be sensitive to implementation details, training choices, and hyperparameter configurations (Schulman et al., 2017; Henderson et al., 2018; Engstrom et al., 2020; Andrychowicz et al., 2021; Eimer et al., 2023), we avoid independently re-implementing the evaluated methods. Instead, we use their officially released implementations whenever available. Based on their primary control targets, we categorize these methods into three groups: Mode Selection-Based Methods, Reasoning Effort Allocation-Based Methods, and Reasoning Expression Optimization-Based Methods.

Mode Selection-Based Methods. Mode selection-based methods choose between direct answering and explicit reasoning based on the input. We consider three representative methods. R-4B develops both direct-answering and explicit-thinking capabilities through bi-mode training and learns mode selection based on task rewards (Yang et al., 2026). TON establishes think-or-not behav-

Category	Method	Trained model	Base model
Mode Selection-Based Methods	TON	TON-3B	Qwen2.5-VL-3B-Instruct
	R-4B	R-4B-RL	–
	Omni-AutoThink	Omni-AutoThink-7B	Qwen2.5-Omni-7B
Reasoning Effort Allocation-Based Methods	ARES	ARES-RL-7B	Qwen2.5-VL-7B-Instruct
	FAST-GRPO	FAST-3B	Qwen2.5-VL-3B-Instruct
		FAST-7B	Qwen2.5-VL-7B-Instruct
		FAST-8B-Preview	Qwen3-VL-8B-Instruct
	VL-Cogito	VL-Cogito-7B	Qwen2.5-VL-7B-Instruct
Reasoning Expression Optimization-Based Methods	SketchThinker-R1	SketchThinker-R1-3B	Qwen2.5-VL-3B-Instruct

Table 1: Overview of the evaluated output-side adaptive reasoning methods across three categories and their corresponding trained and base models.

ior by randomly removing parts of reasoning traces and then uses reinforcement learning to learn whether to engage in reasoning (Wang et al., 2025a). Omni-AutoThink combines adaptive supervised fine-tuning and reinforcement learning to dynamically switch thinking modes according to task complexity (Yang et al., 2025).

Reasoning Effort Allocation-Based Methods. Unlike mode selection-based methods, this category primarily regulates the amount of reasoning effort after reasoning is invoked. We include three representative methods. ARES uses high-entropy tokens to characterize critical exploration states and regulates exploration intensity through difficulty-aware optimization (Chen et al., 2026). FAST-GRPO combines question-difficulty estimation, adaptive length rewards, and difficulty-aware constraints to regulate reasoning length (Xiao & Gan, 2025). VL-Cogito adopts progressive curriculum reinforcement learning and dynamic length rewards, allowing reasoning effort to vary with task complexity (Yuan et al., 2025).

Reasoning Expression Optimization-Based Methods. These methods reduce unnecessary generation by changing how reasoning is expressed. We consider SketchThinker-R1 (Zhang et al., 2026b) as a representative method. It transforms conventional long-form reasoning into sketch-style reasoning that retains only key steps, and further reduces redundant generation through cold-start training for compact reasoning, conciseness evaluation, and reinforcement learning.

As shown in Table 1, in addition to the adaptive reasoning methods, MARE-Bench includes the corresponding base models as references, except for R-4B. Because different methods are built on different model families and parameter scales, these base models allow us to compare both the absolute performance of each method and its relative changes in task performance and generation cost with respect to the corresponding base model, thereby helping distinguish changes associated with adaptive reasoning mechanisms from differences attributable to the underlying base models.

3.2 DATASETS

To systematically evaluate output-side adaptive reasoning methods across different capability dimensions, we select seven public multimodal datasets spanning five evaluation dimensions: General Knowledge, Mathematical Reasoning, Logical Reasoning, Medical QA, and OCR. Specifically, (1) General Knowledge: we use the validation split of MMMU (Yue et al., 2024) and the val split of MMStar (Chen et al., 2024a). (2) Mathematical Reasoning: we use the test split of MathVision (Wang et al., 2024) and the testmini split of MathVista (Lu et al., 2024). (3) Logical Reasoning: we use the test split of LogicVista (Xiao et al., 2024). (4) Medical QA: we use the multimodal test split of MedXpertQA (Zuo et al., 2025). (5) OCR: we use the test split of OCRBench (Liu et al., 2024). These datasets evaluate multimodal model capabilities in general visual understanding, visual mathematical reasoning, logical relation analysis, medical knowledge and image understanding, and visual text recognition, respectively. Together, these datasets provide a unified basis for systematically comparing output-side adaptive reasoning methods across different capability dimensions.

3.3 Evaluation Metrics

We evaluate different output-side adaptive reasoning methods from two perspectives: task performance and output-side generation cost. Specifically, we use accuracy as the primary performance metric and average output tokens to measure generation cost.

(1) Accuracy. For the k-th dataset, let N_k denote the number of samples, and let $c_{k,i} \in \{0, 1\}$ indicate whether the model correctly answers the i-th sample. The accuracy on the k-th dataset and the equal-weighted average accuracy across the seven datasets are defined as

$$\mathrm{Acc}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} c_{k,i} \times 100\%, \qquad \mathrm{AvgAcc} = \frac{1}{K} \sum_{k=1}^{K} \mathrm{Acc}_k, \quad K = 7. \tag{1}$$

Response correctness follows each dataset's standard evaluation protocol. We refer to the equal-weighted average across datasets as Average Accuracy (Avg. Acc.).

(2) Average output tokens. Since this work focuses on output-side adaptive reasoning, we use the number of tokens actually generated by the model to measure generation cost. This metric captures the observable autoregressive generation cost on the output side rather than the model's internal computation. Let $t_{k,i}$ denote the number of output tokens generated by the model for the i-th sample in the k-th dataset. The average output tokens on the k-th dataset and the equal-weighted average across the seven datasets are defined as

$$\mathrm{Tok}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} t_{k,i}, \qquad \mathrm{AvgTok} = \frac{1}{K} \sum_{k=1}^{K} \mathrm{Tok}_k, \quad K = 7. \tag{2}$$

Output tokens include both the explicit reasoning process and the final answer generated by the model, and are counted based on the token IDs actually produced during generation.

A higher Avg. Acc. indicates better task performance, whereas a lower Avg. Tok indicates lower output-side generation cost. Together, these two metrics characterize the performance–generation cost trade-off of different methods.

4 Experiments

4.1 Settings

To ensure a fair comparison among different output-side adaptive reasoning methods, we adopt a unified evaluation pipeline. All experiments are conducted using the open-source evaluation toolkit VLMEvalKit (Duan et al., 2024), with vLLM (Kwon et al., 2023) used for model inference. Inputs are constructed following the default evaluation pipeline of VLMEvalKit, and each model is evaluated using its native chat template. For methods that require specific control instructions or mode indicators, we follow their official inference settings; aside from the method-specific control prompts, all other evaluation conditions are kept consistent.

Since TON trains separate models for counting, mathematical reasoning, and mobile navigation, we select the corresponding model according to the evaluation task. Specifically, TON-3B-Math is used for MMMU, MathVision, MathVista, and MedXpertQA. TON-3B-CLEVR is used for MMStar and LogicVista, which mainly involve visual-spatial relations and object counting. TON-3B-AITZ, which is designed for mobile interface understanding, is used for OCRBench. Notably, TON-3B-AITZ is used on OCRBench in a cross-domain transfer setting and is not specifically trained for OCR. In addition, R-4B uses Qwen3 as its language backbone together with a separately trained visual encoder and has no corresponding base model. We therefore report only its absolute results and exclude it from analyses of relative changes with respect to a base model. For the Qwen3-VL-8B series, FAST-8B-Preview is built on Qwen3-VL-8B-Instruct, while Qwen3-VL-8B-Thinking is additionally reported as a reference model.

Except in the Output Budget Sensitivity analysis, the maximum generation length is set to 8,192 tokens for all methods, and greedy decoding with temperature = 0 is used to avoid additional variance introduced by stochastic sampling.

Method	General				Math				Logic		Medical		OCR		Avg.	
	MMMU		MMStar		MathVision		MathVista		LogicVista		MedXpertQA		OCRBench			
	Acc.↑	#Tokens↓	Acc.↑	#Tokens↓	Acc.↑	#Tokens↓	Acc.↑	#Tokens↓	Acc.↑	#Tokens↓	Acc.↑	#Tokens↓	Acc.↑	#Tokens↓	Acc.↑	#Tokens↓
Qwen2.5-VL-3B																
Base	**47.11**	341.63	52.73	55.10	23.45	683.57	60.20	147.88	37.50	15.01	21.85	**9.88**	**83.00**	**7.79**	46.55	180.12
TON	43.22	**17.68**	53.47	**13.96**	16.48	**13.87**	61.30	**17.26**	33.71	**14.47**	22.55	17.61	57.20	32.71	41.13	**18.22**
FAST-GRPO	46.56	693.96	**56.53**	243.35	**24.84**	917.37	**79.20**	297.84	**40.62**	336.83	**22.60**	116.41	82.00	19.57	**50.34**	375.05
SketchThinker-R1	44.89	274.03	54.33	143.86	16.81	339.65	76.80	140.46	35.94	176.59	20.85	218.33	78.40	46.22	46.86	191.31
Qwen2.5-VL-7B																
Base	50.33	365.26	57.73	148.29	25.39	533.63	67.60	157.78	38.84	**3.46**	21.10	154.43	87.70	**7.81**	49.81	195.81
ARES	51.78	3843.66	58.67	2458.12	**34.84**	5912.19	73.10	2643.67	46.88	4615.45	**23.85**	3251.61	**88.00**	954.81	53.87	3382.79
FAST-GRPO	**55.33**	**222.97**	**64.07**	99.13	28.12	396.73	**73.30**	**138.84**	**49.55**	170.61	23.75	130.35	86.10	52.40	**54.32**	173.00
VL-Cogito	52.22	286.49	62.73	**79.83**	22.57	**294.95**	71.50	177.84	44.20	243.29	23.15	**2.42**	85.90	9.18	51.75	**156.29**
Qwen2.5-Omni-7B																
Base	**52.33**	**166.92**	**60.33**	**6.64**	**24.18**	**308.06**	**66.10**	**169.83**	38.84	**4.91**	21.55	**9.67**	**86.30**	**7.82**	**49.95**	**96.27**
Omni-AutoThink	**52.33**	575.06	53.80	1071.25	18.62	2692.72	59.30	1915.55	**39.51**	629.00	**23.20**	89.19	79.90	1610.86	46.67	1226.23
Qwen3-VL-8B																
Base	55.89	2175.69	63.73	837.04	24.97	**13.76**	68.40	607.60	41.52	**2.12**	26.25	1451.77	**87.90**	7.51	52.67	727.93
Base (Thinking)	55.89	3557.86	63.07	1453.49	**42.04**	5161.51	**75.00**	1251.23	54.91	2819.37	26.30	2094.28	84.00	245.08	57.32	2368.98
FAST-GRPO	**63.78**	**801.87**	**74.93**	**346.31**	38.88	583.57	71.80	**123.79**	**57.37**	481.20	**29.10**	**604.75**	87.10	**7.36**	**60.42**	**421.27**
R-4B																
R-4B-RL	60.00	2163.85	71.87	551.26	47.43	3054.87	77.40	972.63	58.48	1417.61	23.30	1204.91	83.90	67.36	60.34	1347.50

Table 2: Performance and generation cost comparison across different methods and benchmarks. Best results are in **bold**.

4.2 Main Results

Table 2 summarizes the results, while Figure 2 shows their distribution in terms of Avg. Acc. and average output tokens. We make the following observations:

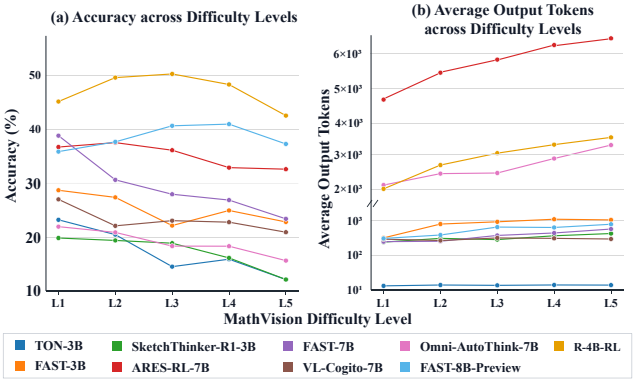

Figure 2: Accuracy–generation cost trade-off across seven datasets.

1. Considering both Avg. Acc. and average output tokens, some methods outperform their base models, while others perform poorly. On Qwen2.5-VL-7B, FAST-GRPO increases Avg. Acc. from 49.81% to 54.32% while reducing average output tokens from 195.81 to 173.00. VL-Cogito also increases Avg. Acc. to 51.75% while reducing average output tokens to 156.29. In contrast, on Qwen2.5-VL-3B, TON reduces average output tokens from 180.12 to 18.22, but Avg. Acc. decreases from 46.55% to 41.13%. On Qwen2.5-VL-7B, ARES increases Avg. Acc. from 49.81% to 53.87%, but average output tokens increase from 195.81 to 3,382.79. On Qwen2.5-Omni-7B, Omni-AutoThink decreases Avg. Acc. from 49.95% to 46.67% while increasing average output tokens from 96.27 to 1,226.23. These results indicate that some methods may not effectively realize the expected benefits of adaptive reasoning.

2. FAST-GRPO achieves competitive overall performance across the evaluated setings in most cases. FAST-GRPO achieves the highest Avg. Acc. among the evaluated methods on Qwen2.5-VL-3B, Qwen2.5-VL-7B, and Qwen3-VL-8B. On Qwen2.5-VL-7B, it increases Avg. Acc. from 49.81% to 54.32% while reducing average output tokens from 195.81 to 173.00. On Qwen3-VL-8B, it increases Avg. Acc. from 52.67% to 60.42% while reducing average output tokens from 727.93 to 421.27. These results show that FAST-GRPO consistently improves overall accuracy across different base models while reducing generation cost in the Qwen2.5-VL-7B and Qwen3-VL-8B settings.

3. The effects of adaptive reasoning vary across evaluation dimensions: Mathematical Reasoning and Logical Reasoning show larger performance gains, General Knowledge and Medical QA show smaller gains, while OCR exhibits performance degradation. Relative to the corresponding base models, accuracy on Mathematical Reasoning and Logical Reasoning improves by an average of 3.68 and 4.80 percentage points, respectively, across methods, with an average improve-

ment of 6.05 percentage points on MathVista. In contrast, the average improvements on General Knowledge and Medical QA are only 2.04 and 1.55 percentage points, respectively. On OCRBench, accuracy instead decreases by an average of 5.21 percentage points. We hypothesize that adaptive reasoning can allocate more reasoning effort to tasks requiring multi-step derivation and is therefore more beneficial for Mathematical Reasoning and Logical Reasoning. In contrast, OCR tasks primarily rely on direct visual text recognition, and unnecessary reinterpretation or modification of recognized text may lead to performance degradation.

4. The performance–generation cost trade-off of adaptive reasoning methods can vary across base models. To examine this effect, we compare FAST-GRPO across different base models. On Qwen2.5-VL-3B, Qwen2.5-VL-7B, and Qwen3-VL-8B, FAST-GRPO improves Avg. Acc. over the corresponding base models by 3.79, 4.50, and 7.76 percentage points, respectively. However, its effect on generation cost differs across base models. On Qwen2.5-VL-3B, FAST-GRPO increases average output tokens from 180.12 to 375.05, whereas on Qwen2.5-VL-7B and Qwen3-VL-8B, it reduces average output tokens from 195.81 to 173.00 and from 727.93 to 421.27, respectively. These results suggest that the performance–generation cost trade-off of adaptive reasoning methods can be influenced by the underlying base model.

4.3 Analysis and Discussion

4.3.1 Reasoning Effort Adaptivity

Does increasing reasoning effort help solve more difficult problems? We analyze the average output tokens and accuracy of different adaptive reasoning methods at each MathVision difficulty level. As shown in Figure 3, MathVision contains 3,040 visual mathematical problems from real mathematics competitions, covering 16 mathematical subjects, with the problems categorized into five difficulty levels from L1 to L5. As problem difficulty increases, most methods generate more output tokens on average. Among them, ARES-RL-7B, FAST-8B-Preview, and R-4B-RL maintain relatively stable accuracy across different difficulty levels. In contrast, TON-3B and VL-Cogito-7B show relatively small changes in average output tokens across difficulty levels, while their accuracy exhibits a clear overall decline. However, longer reasoning does not always lead to better performance on more difficult problems: the average output tokens of FAST-7B and Omni-AutoThink-7B increase substantially with difficulty, yet their accuracy continues to decline. These results show that greater output length is sometimes associated with smaller performance degradation as difficulty increases, but longer outputs do not consistently correspond to better performance. Thus, output length alone cannot explain performance on difficult problems.

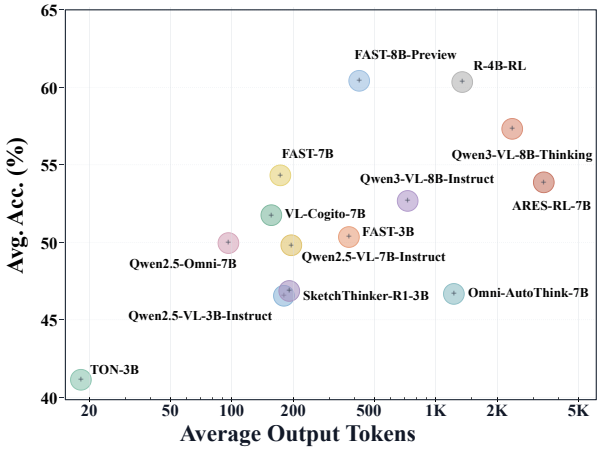

Figure 3: Average output tokens and accuracy across MathVision difficulty levels.

4.3.2 Output Budget Sensitivity

Does increasing the output budget improve answer accuracy? On MathVision, we analyze the accuracy and Truncation Rate of Qwen2.5-VL-7B-Instruct and three adaptive reasoning methods built on it under different maximum generation lengths. We further examine how generation repetition consumes the available output budget and may reduce the budget available for completing the reasoning process. As shown in Figures 4(a) and 4(b), different methods exhibit markedly different sensitivity to the output budget. For ARES-RL-7B, as the maximum generation length increases from 2K to 16K, the Truncation Rate steadily decreases from 88.68% to 23.03%, while accuracy increases from 9.34% to 44.01%, indicating that its performance is clearly constrained by the output budget. In contrast, FAST-7B and VL-Cogito-7B maintain consistently low Truncation Rates and

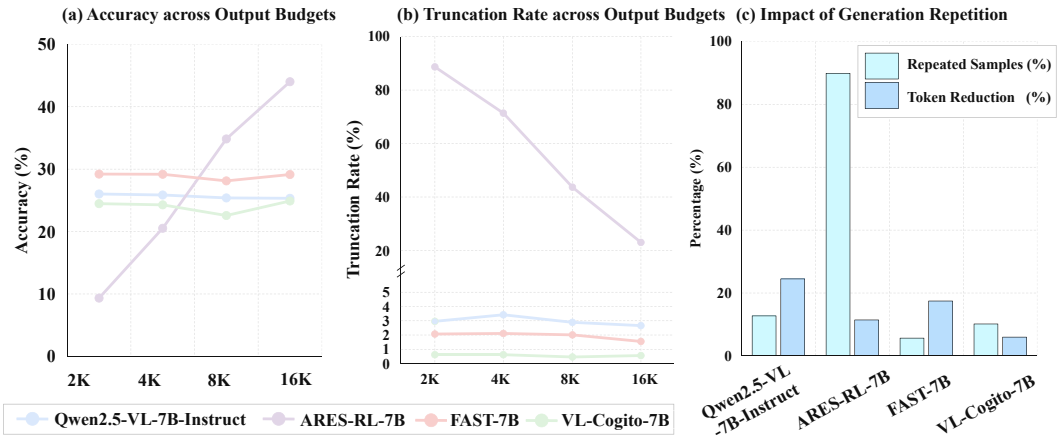

Figure 4: Output Budget Sensitivity and generation repetition on MathVision.

relatively stable accuracy across different budgets, and increasing the output budget does not yield clear gains.

We further analyze how generation repetition consumes the available output budget. Generation repetition is a form of reasoning-trace degradation in which the model repeatedly generates the same reasoning segments (Hiraoka & Inui, 2025; Cheng et al., 2026; Liu et al., 2026; Ha et al., 2026). Compared with their base models, different adaptive reasoning methods increase or reduce this behavior to varying degrees. Specifically, we remove exactly repeated text segments from the MathVision outputs of each method and recompute average output tokens. As shown in Figure 4(c), 89.87% of samples from ARES-RL-7B contain repetition, but its average output tokens decrease by only 11.44% after deduplication. In contrast, the proportions of samples containing repetition for FAST-7B and VL-Cogito-7B are 5.66% and 10.16%, respectively, while their average output tokens decrease by 17.46% and 5.98% after deduplication. These results indicate that larger values of average output tokens are not necessarily attributable to generation repetition; some methods inherently tend to generate longer reasoning processes.

Overall, increasing the output budget does not necessarily improve accuracy; it may yield clear gains only when reasoning is constrained by the output limit.

4.3.3 Error Attribution

To analyze the failure modes of different adaptive reasoning methods, we uniformly sample 50 cases from the MMMU results of Qwen2.5-VL-7B-Instruct and each of the three adaptive reasoning methods built on it, yielding 200 cases in total for manual error attribution. Following prior taxonomies of multimodal reasoning errors and reasoning-process failure modes (Zhang et al., 2025a; Wang et al., 2025b; Xiao & Gan, 2025; Ma et al., 2026), we categorize the errors into four types: (1) Visual Perception Failure, where the decisive error arises from incorrect recognition or understanding of visual information such as images or charts; (2) Reasoning Error Propagation, where logical, computational, or derivation errors occur during the reasoning process and subsequently affect later reasoning and the final answer; (3) Knowledge Conflict & Gap, where errors arise from missing knowledge, incorrect use of knowledge, or conflicts between existing knowledge and the information provided in the current problem; and (4) Unre-

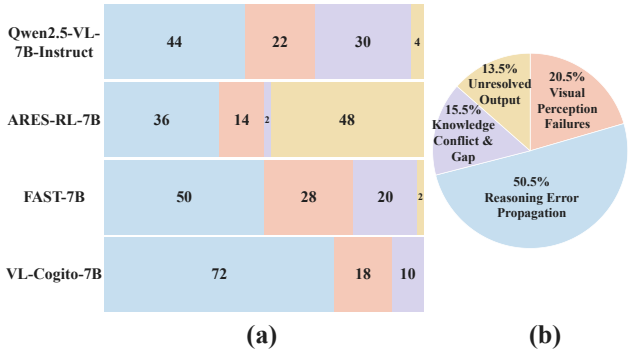

Figure 5: Error type distribution on MMMU.

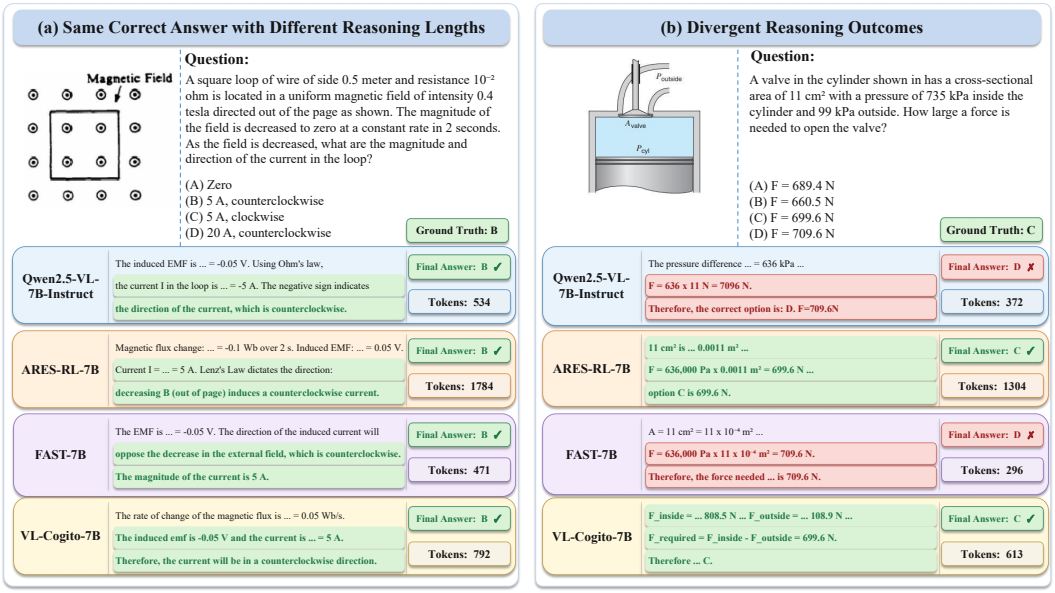

Figure 6: Case study of adaptive reasoning methods: (a) same correct answer with different reasoning lengths and (b) divergent reasoning outcomes.

solved Output, where the model fails to produce a clear and stable final answer due to truncation, persistent generation repetition, or similar issues.

As shown in Figure 5, Reasoning Error Propagation accounts for the majority of errors (50.5%), indicating that local errors introduced during reasoning often propagate to subsequent derivations and ultimately lead to incorrect answers. Visual Perception Failure (20.5%) and Knowledge Conflict & Gap (15.5%) also remain important causes of errors in adaptive reasoning. In contrast, 48% of the errors from ARES-RL-7B are categorized as Unresolved Output, consistent with its higher Truncation Rate and longer outputs. These findings suggest that future work on adaptive reasoning should prioritize the logical correctness of the reasoning process while further improving the reliability of visual perception and knowledge utilization.

4.4 Case Study

To further illustrate the reasoning behavior of different adaptive reasoning methods on individual samples, we present two representative cases in Figure 6. In the first case, all four models produce the correct answer, but their output lengths differ substantially. FAST-7B generates 471 output tokens, whereas ARES-RL-7B generates 1,784 output tokens, illustrating differences in reasoning effort across methods on the same problem. In contrast, the second case shows divergent outcomes resulting from different reasoning processes: ARES-RL-7B and VL-Cogito-7B arrive at the correct answer through valid calculations, whereas Qwen2.5-VL-7B-Instruct and FAST-7B make errors during calculation and produce incorrect answers. These cases illustrate differences in reasoning length and process across adaptive reasoning methods, complementing the preceding analyses.

5 Conclusion

In this work, we introduce MARE-Bench, a benchmark for systematically evaluating output-side adaptive reasoning methods for large multimodal models (LMMs). Under unified evaluation protocols, we systematically evaluate seven representative methods across three LMM families on seven representative datasets spanning five evaluation dimensions. We further analyze their reasoning behavior from the perspectives of Reasoning Effort Adaptivity, Output Budget Sensitivity, and Error Attribution. Our results show that different methods exhibit clear differences in performance and generation cost, and that their effectiveness is jointly influenced by task type and the base model. FAST-GRPO demonstrates consistent performance across base models. We hope MARE-Bench will advance research on efficient adaptive reasoning in LMMs.

AI USE STATEMENT

Large language models (LLMs) were used solely to assist with writing refinement and language polishing during the preparation of this manuscript. The authors remain fully responsible for all content of this paper. LLMs were not used for research ideation, methodology design, experimental implementation, data analysis, or interpretation of results.

ETHICS STATEMENT

This work evaluates existing large multimodal models on publicly available benchmark datasets. We do not collect or process any private user data, and this study does not involve human subjects or human experimentation. For Medical QA benchmarks, we only evaluate model outputs using publicly available datasets and do not provide medical advice or clinical recommendations. We are not aware of any additional ethical concerns arising from this work.

REFERENCES

Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, Roman Ring, Eliza Rutherford, Serkan Cabi, Tengda Han, Zhitao Gong, Sina Samangooei, Marianne Monteiro, Jacob L. Menick, Sebastian Borgeaud, Andy Brock, Aida Nematzadeh, Sahand Sharifzadeh, Mikolaj Binkowski, Ricardo Barreira, Oriol Vinyals, Andrew Zisserman, and Karen Simonyan. Flamingo: A visual language model for few-shot learning. In *Advances in Neural Information Processing Systems*, volume 35, pp. 23716–23736, 2022. doi: 10.52202/068431-1723.

Marcin Andrychowicz, Anton Raichuk, Piotr Stańczyk, Manu Orsini, Sertan Girgin, Raphaël Marinier, Léonard Hussenot, Matthieu Geist, Olivier Pietquin, Marcin Michalski, Sylvain Gelly, and Olivier Bachem. What matters for on-policy deep actor-critic methods? a large-scale study. In *International Conference on Learning Representations*, 2021.

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Jie Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Mingsheng Li, Mei Li, Kaixin Li, Zicheng Lin, Junyang Lin, Xuejing Liu, Jiawei Liu, Chenglong Liu, Yang Liu, Dayiheng Liu, Shixuan Liu, Dunjie Lu, Ruilin Luo, Chenxu Lv, Rui Men, Lingchen Meng, Xuancheng Ren, Xingzhang Ren, Sibo Song, Yuchong Sun, Jun Tang, Jianhong Tu, Jianqiang Wan, Peng Wang, Pengfei Wang, Qiuyue Wang, Yuxuan Wang, Tianbao Xie, Yiheng Xu, Haiyang Xu, Jin Xu, Zhibo Yang, Mingkun Yang, Jianxin Yang, An Yang, Bowen Yu, Fei Zhang, Hang Zhang, Xi Zhang, Bo Zheng, Humen Zhong, Jingren Zhou, Fan Zhou, Jing Zhou, Yuanzhi Zhu, and Ke Zhu. Qwen3-VL technical report. *arXiv preprint arXiv:2511.21631*, 2025a.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. Qwen2.5-VL technical report. *arXiv preprint arXiv:2502.13923*, 2025b.

Zhongang Cai, Ruisi Wang, Chenyang Gu, Fanyi Pu, Junxiang Xu, Yubo Wang, Wanqi Yin, Zhitao Yang, Chen Wei, Tongxi Zhou, Qingping Sun, Hui En Pang, Jiaqi Li, Oscar Qian, Zhiqian Lin, Xuanke Shi, Kewang Deng, Xiaoyang Han, Zukai Chen, Xiangyu Fan, Hanming Deng, Lewei Lu, Liang Pan, Bo Li, Ziwei Liu, Quan Wang, Dahua Lin, and Lei Yang. Scaling spatial intelligence with multimodal foundation models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7879–7890, 2026.

Lin Chen, Jinsong Li, Xiaoyi Dong, Pan Zhang, Yuhang Zang, Zehui Chen, Haodong Duan, Jiaqi Wang, Yu Qiao, Dahua Lin, and Feng Zhao. Are we on the right way for evaluating large vision-language models? In *Advances in Neural Information Processing Systems*, volume 37, pp. 27056–27087, 2024a. doi: 10.52202/079017-0850.

Shuang Chen, Hangyu Guo, Yimeng Ye, Shijue Huang, Wenbo Hu, Jiayu Chen, Manyuan Zhang, Haoxi Li, Song Guo, and Nanyun (Violet) Peng. ARES: Multimodal adaptive reasoning via

difficulty-aware token-level entropy shaping. In *International Conference on Learning Representations*, 2026.

Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, Bin Li, Ping Luo, Tong Lu, Yu Qiao, and Jifeng Dai. 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 (CVPR)*, pp. 24185–24198, 2024b. doi: 10.1109/CVPR52733.2024.02283.

Zhengxiang Cheng, Dongping Chen, Mingyang Fu, and Tianyi Zhou. Optimizing length compression in large reasoning models. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 3232–3250, 2026. doi: 10.18653/v1/2026.acl-long.146.

Wenliang Dai, Junnan Li, Dongxu Li, Anthony Tiong, Junqi Zhao, Weisheng Wang, Boyang Li, Pascale N Fung, and Steven Hoi. InstructBLIP: Towards general-purpose vision-language models with instruction tuning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 49250–49267, 2023. doi: 10.52202/075280-2142.

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 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9062–9072, 2025. doi: 10.1109/CVPR52734.2025.00847.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

Haodong Duan, Junming Yang, Yuxuan Qiao, Xinyu Fang, Lin Chen, Yuan Liu, Xiaoyi Dong, Yuhang Zang, Pan Zhang, Jiaqi Wang, Dahua Lin, and Kai Chen. VLMEvalKit: An open-source toolkit for evaluating large multi-modality models. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 11198–11201, 2024. doi: 10.1145/3664647.3685520.

Theresa Eimer, Marius Lindauer, and Roberta Raileanu. Hyperparameters in reinforcement learning and how to tune them. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 9104–9149, 2023.

Logan Engstrom, Andrew Ilyas, Shibani Santurkar, Dimitris Tsipras, Firdaus Janoos, Larry Rudolph, and Aleksander Madry. Implementation matters in deep RL: A case study on PPO and TRPO. In *International Conference on Learning Representations*, 2020.

Chaoyou Fu, Peixian Chen, Yunhang Shen, Yulei Qin, Mengdan Zhang, Xu Lin, Jinrui Yang, Xiawu Zheng, Ke Li, Xing Sun, Yunsheng Wu, Rongrong Ji, Caifeng Shan, and Ran He. MME: A comprehensive evaluation benchmark for multimodal large language models. In *Advances in Neural Information Processing Systems*, volume 38, pp. 162549–162567, 2025a. doi: 10.52202/085713-4899.

Chaoyou Fu, Yuhan Dai, Yongdong Luo, Lei Li, Shuhuai Ren, Renrui Zhang, Zihan Wang, Chenyu Zhou, Yunhang Shen, Mengdan Zhang, Peixian Chen, Yanwei Li, Shaohui Lin, Sirui Zhao, Ke Li, Tong Xu, Xiawu Zheng, Enhong Chen, Caifeng Shan, Ran He, and Xing Sun. Video-MME: The first-ever comprehensive evaluation benchmark of multi-modal LLMs in video analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24108–24118, 2025b. doi: 10.1109/CVPR52734.2025.02245.

Jun Gao, Yongqi Li, Ziqiang Cao, and Wenjie Li. Interleaved-modal chain-of-thought. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19520–19529, 2025. doi: 10.1109/CVPR52734.2025.01818.

Zirun Guo, Minjie Hong, Feng Zhang, Kai Jia, and Tao Jin. Thinking with programming vision: Towards a unified view for thinking with images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 33467–33476, 2026.

Rui Ha, Rui Pu, Chaozhuo Li, Li Sun, and Sen Su. From "aha moments" to controllable thinking: Toward meta-cognitive reasoning in LRMs via decoupled reasoning and control. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 6698–6710, 2026. doi: 10.18653/v1/2026.acl-long.304.

Peter Henderson, Riashat Islam, Philip Bachman, Joelle Pineau, Doina Precup, and David Meger. Deep reinforcement learning that matters. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, pp. 3207–3214, 2018. doi: 10.1609/aaai.v32i1.11694.

Tatsuya Hiraoka and Kentaro Inui. Repetition neurons: How do language models produce repetitions? In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pp. 483–495, 2025. doi: 10.18653/v1/2025.naacl-short.41.

Junquan Huang, Haotian Wu, Yubo Gao, Yibo Yan, Junyan Zhang, Yonghua Hei, Song Dai, Jie Zhang, Puay Siew Tan, and Xuming Hu. Effireason-bench: A unified benchmark for evaluating and advancing efficient reasoning in large language models. *arXiv preprint arXiv:2511.10201*, 2025.

Dongzhi Jiang, Renrui Zhang, Ziyu Guo, Yanwei Li, Yu Qi, Xinyan Chen, Liuhui Wang, Jianhan Jin, Claire Guo, Shen Yan, Bo Zhang, Chaoyou Fu, Peng Gao, and Hongsheng Li. MME-CoT: Benchmarking chain-of-thought in large multimodal models for reasoning quality, robustness, and efficiency. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 27793–27830, 2025.

Zhuoran Jin, Kejian Zhu, Hongbang Yuan, Yupu Hao, Pengfei Cao, Yubo Chen, Kang Liu, and Jun Zhao. Look light, think heavy: What multimodal chain-of-thought reasoning can and cannot do. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 8567–8608, 2026. doi: 10.18653/v1/2026.acl-long.387.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. In *Advances in Neural Information Processing Systems*, volume 35, pp. 22199–22213, 2022. doi: 10.52202/068431-1613.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with PagedAttention. In *Proceedings of the 29th Symposium on Operating Systems Principles*, pp. 611–626, 2023. doi: 10.1145/3600006.3613165.

Bohao Li, Yuying Ge, Yixiao Ge, Guangzhi Wang, Rui Wang, Ruimao Zhang, and Ying Shan. SEED-Bench: Benchmarking multimodal large language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13299–13308, 2024a. doi: 10.1109/CVPR52733.2024.01263.

Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 19730–19742, 2023.

Kunchang Li, Yali Wang, Yinan He, Yizhuo Li, Yi Wang, Yi Liu, Zun Wang, Jilan Xu, Guo Chen, Ping Lou, Limin Wang, and Yu Qiao. MVBench: A comprehensive multi-modal video understanding benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22195–22206, 2024b. doi: 10.1109/CVPR52733.2024.02095.

Ming Li, Jike Zhong, Shitian Zhao, Yuxiang Lai, Haoquan Zhang, Wang Bill Zhu, and Kaipeng Zhang. To think or not to think: A study of thinking in rule-based visual reinforcement fine-tuning. In *Advances in Neural Information Processing Systems*, volume 38, pp. 185484–185535, 2025. doi: 10.52202/085713-5578.

Ji Lin, Hongxu Yin, Wei Ping, Pavlo Molchanov, Mohammad Shoeybi, and Song Han. VILA: On pre-training for visual language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26689–26699, 2024. doi: 10.1109/CVPR52733.2024.02520.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 34892–34916, 2023. doi: 10.52202/075280-1516.

Yongjiang Liu, Haoxi Li, Xiaosong Ma, Jie Zhang, and Song Guo. Think how to think: Mitigating overthinking with autonomous difficulty cognition in large reasoning models. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 38105–38126, 2026. doi: 10.18653/v1/2026.acl-long.1766.

Yuliang Liu, Zhang Li, Mingxin Huang, Biao Yang, Wenwen Yu, Chunyuan Li, Xu-Cheng Yin, Cheng-Lin Liu, Lianwen Jin, and Xiang Bai. OCRBench: On the hidden mystery of OCR in large multimodal models. *Science China Information Sciences*, 67(12):220102, 2024. doi: 10.1007/s11432-024-4235-6.

Pan Lu, Hritik Bansal, Tony Xia, Jiacheng Liu, Chunyuan Li, Hannaneh Hajishirzi, Hao Cheng, Kai-Wei Chang, Michel Galley, and Jianfeng Gao. MathVista: Evaluating mathematical reasoning of foundation models in visual contexts. In *International Conference on Learning Representations*, 2024.

Gen Luo, Xue Yang, Wenhan Dou, Zhaokai Wang, Jiawen Liu, Jifeng Dai, Yu Qiao, and Xizhou Zhu. Mono-InternVL: Pushing the boundaries of monolithic multimodal large language models with endogenous visual pre-training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24960–24971, 2025. doi: 10.1109/CVPR52734.2025.02324.

Ji Ma, Wei Suo, Peng Wang, and Yanning Zhang. Understanding and mitigating hallucinations in multimodal chain-of-thought models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 40224–40234, 2026.

Zhiliang Peng, Wenhui Wang, Li Dong, Yaru Hao, Shaohan Huang, Shuming Ma, Qixiang Ye, and Furu Wei. Grounding multimodal large language models to the world. In *International Conference on Learning Representations*, 2024.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763, 2021.

Timo Schick, Jane Dwivedi-Yu, Roberto Dessi, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools. In *Advances in Neural Information Processing Systems*, volume 36, pp. 68539–68551, 2023. doi: 10.52202/075280-2997.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Hao Shao, Shengju Qian, Han Xiao, Guanglu Song, Zhuofan Zong, Letian Wang, Yu Liu, and Hongsheng Li. Visual CoT: Advancing multi-modal language models with a comprehensive dataset and benchmark for chain-of-thought reasoning. In *Advances in Neural Information Processing Systems*, volume 37, pp. 8612–8642, 2024. doi: 10.52202/079017-0275.

Guiyao Tie, Xueyang Zhou, Tianhe Gu, Ruihang Zhang, Chaoran Hu, Sizhe Zhang, Mengqu Sun, Yan Zhang, Pan Zhou, and Lichao Sun. MMMR: Benchmarking massive multi-modal reasoning tasks. In *Proceedings of the NeurIPS 2025 Datasets and Benchmarks Track*, 2025.

Jiaqi Wang, Kevin Qinghong Lin, James Cheng, and Mike Zheng Shou. Think or not? selective reasoning via reinforcement learning for vision-language models. In *Advances in Neural Information Processing Systems*, volume 38, pp. 123247–123285, 2025a. doi: 10.52202/085713-3712.

Ke Wang, Junting Pan, Weikang Shi, Zimu Lu, Houxing Ren, Aojun Zhou, Mingjie Zhan, and Hongsheng Li. Measuring multimodal mathematical reasoning with MATH-Vision dataset. In *Advances in Neural Information Processing Systems*, volume 37, pp. 95095–95169, 2024. doi: 10.52202/079017-3014.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *International Conference on Learning Representations*, 2023.

Yizhi Wang, Linan Yue, and Min-Ling Zhang. Bridging efficiency and transparency: Explainable CoT compression in multimodal large reasoning models. *arXiv preprint arXiv:2602.09485*, 2026.

Ziyue Wang, Yurui Dong, Fuwen Luo, Minyuan Ruan, Zhili Cheng, Chi Chen, Peng Li, and Yang Liu. How do multimodal large language models handle complex multimodal reasoning? placing them in an extensible escape game. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4807–4817, 2025b. doi: 10.1109/ICCV51701.2025.00457.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 24824–24837, 2022. doi: 10.52202/068431-1800.

Wenyi Xiao and Leilei Gan. Fast-Slow thinking GRPO for large vision-language model reasoning. In *Advances in Neural Information Processing Systems*, volume 38, pp. 190068–190098, 2025. doi: 10.52202/085713-5712.

Yijia Xiao, Edward Sun, Tianyu Liu, and Wei Wang. LogicVista: Multimodal llm logical reasoning benchmark in visual contexts. *arXiv preprint arXiv:2407.04973*, 2024.

Jin Xu, Zhifang Guo, Jinzheng He, Hangrui Hu, Ting He, Shuai Bai, Keqin Chen, Jialin Wang, Yang Fan, Kai Dang, Bin Zhang, Xiong Wang, Yunfei Chu, and Junyang Lin. Qwen2.5-Omni technical report. *arXiv preprint arXiv:2503.20215*, 2025a.

Zhongxing Xu, Chengzhi Liu, Qingyue Wei, Juncheng Wu, James Zou, Xin Wang, Yuyin Zhou, and Sheng Liu. More thinking, less seeing? assessing amplified hallucination in multimodal reasoning models. In *Advances in Neural Information Processing Systems*, volume 38, pp. 92357–92384, 2025b. doi: 10.52202/085713-2777.

An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, Guanting Dong, Haoran Wei, Huan Lin, Jialong Tang, Jialin Wang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Ma, Jianxin Yang, Jin Xu, Jingren Zhou, Jinze Bai, Jinzheng He, Junyang Lin, Kai Dang, Keming Lu, Keqin Chen, Kexin Yang, Mei Li, Mingfeng Xue, Na Ni, Pei Zhang, Peng Wang, Ru Peng, Rui Men, Ruize Gao, Runji Lin, Shijie Wang, Shuai Bai, Sinan Tan, Tianhang Zhu, Tianhao Li, Tianyu Liu, Wenbin Ge, Xiaodong Deng, Xiaohuan Zhou, Xingzhang Ren, Xinyu Zhang, Xipin Wei, Xuancheng Ren, Xuejing Liu, Yang Fan, Yang Yao, Yichang Zhang, Yu Wan, Yunfei Chu, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, Zhifang Guo, and Zhihao Fan. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*, 2024.

Dongchao Yang, Songxiang Liu, Disong Wang, Yuanyuan Wang, Guanglu Wan, and Helen Meng. Omni-AutoThink: Adaptive multimodal reasoning via reinforcement learning. *arXiv preprint arXiv:2512.03783*, 2025.

Qi Yang, Bolin Ni, Shiming Xiang, and Houwen Peng. R-4B: Incentivizing general-purpose auto-thinking in MLLMs via bi-mode annealing and reinforce learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7891–7900, 2026.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 11809–11822, 2023a. doi: 10.52202/075280-0517.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations*, 2023b.

Jiakang Yuan, Tianshuo Peng, Yilei Jiang, Yiting Lu, Renrui Zhang, Kaituo Feng, Chaoyou Fu, Tao Chen, Lei Bai, and Bo Zhang. MME-Reasoning: A broad-spectrum benchmark for evaluating logical reasoning in MLLMs. In *Proceedings of the International Conference on Machine Learning*, 2026.

Ruifeng Yuan, Chenghao Xiao, Sicong Leng, Jianyu Wang, Long Li, Weiwen Xu, Hou Pong Chan, Deli Zhao, Tingyang Xu, Zhongyu Wei, Hao Zhang, and Yu Rong. VL-Cogito: Progressive curriculum reinforcement learning for advanced multimodal reasoning. *arXiv preprint arXiv:2507.22607*, 2025.

Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, and Wenhu Chen. 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 (CVPR)*, pp. 9556–9567, 2024. doi: 10.1109/CVPR52733.2024.00913.

Chi Zhang, Haibo Qiu, Qiming Zhang, Yufei Xu, Zhixiong Zeng, Siqi Yang, Peng Shi, Lin Ma, and Jing Zhang. Perceptual-evidence anchored reinforced learning for multimodal reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 41111–41120, 2026a.

Di Zhang, Jingdi Lei, Junxian Li, Xunzhi Wang, Yujie Liu, Zonglin Yang, Jiatong Li, Weida Wang, Suorong Yang, Jianbo Wu, Peng Ye, Wanli Ouyang, and Dongzhan Zhou. Critic-V: VLM critics help catch VLM errors in multimodal reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9050–9061, 2025a. doi: 10.1109/CVPR52734.2025.00846.

Ruiyang Zhang, Dongzhan Zhou, and Zhedong Zheng. SketchThinker-R1: Towards efficient sketch-style reasoning in large multimodal models. In *International Conference on Learning Representations*, 2026b.

Xintong Zhang, Xiaowen Zhang, Jongrong Wu, Zhi Gao, Shilin Yan, Zhenxin Diao, Kunpeng Gao, Xuanyan Chen, Yuwei Wu, Yunde Jia, and Qing Li. AdaptMMBench: Benchmarking adaptive multimodal reasoning for mode selection and reasoning process. *arXiv preprint arXiv:2602.02676*, 2026c.

Zhi Zhang, Srishti Yadav, Fengze Han, and Ekaterina Shutova. Cross-modal information flow in multimodal large language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19781–19791, 2025b. doi: 10.1109/CVPR52734.2025.01842.

Yuxin Zuo, Shang Qu, Yifei Li, Zhang-Ren Chen, Xuekai Zhu, Ermo Hua, Kaiyan Zhang, Ning Ding, and Bowen Zhou. MedXpertQA: Benchmarking expert-level medical reasoning and understanding. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 80961–80990, 2025.