Auxiliary Clues Aware's Geometry Problem Solving

Qigong Lei[1,*] **Qingsong Wang**[1,*] **Wang Lin**[1] **Qianyi Yang**[1]
Chang Yao[1,2] **Jingyuan Chen**[1,†]
[1]*Zhejiang University*
[2]*Innovation and Management Center,*
School of Software Technology (Ningbo), Zhejiang University
{leiqigong,wqsong,linwanglw,qianyiy}@zju.edu.cn
{changy,jingyuanchen}@zju.edu.cn
[*]Equal contribution. [†]Corresponding authors.

Abstract

Geometry Problem Solving (GPS) serves as a rigorous test of the complex reasoning capabilities of Multimodal Large Language Models (MLLMs), particularly when solutions require auxiliary constructions absent from the input diagram. However, explicit diagramming methods suffer from unstable image generation quality and cascading errors, while existing implicit attention mechanisms primarily focus on visible elements, limiting their ability to support reasoning that requires additional geometric constructions. To address these challenges, we propose **AuxCA** (Auxiliary-Clues-Aware GRPO), a reinforcement learning framework that helps models learn transferable auxiliary-construction strategies that can be applied beyond the training domain without explicit diagram generation at inference time. Specifically, we introduce Auxiliary Clues Dependency (**ACD**) to quantify the influence of auxiliary visual clues on token generation and integrate it into GRPO through fine-grained advantage weighting. Furthermore, we incorporate an auxiliary incentive reward to encourage the effective use of auxiliary geometric clues in correct solutions. Extensive experiments demonstrate that AuxCA-8B outperforms the explicit-plotting baseline Bagel-Canvas and achieves competitive performance with leading proprietary models on auxiliary-dependent plane geometry problems. Improvements on analytic geometry, solid geometry, and external benchmarks further suggest that the learned strategies can transfer beyond the training domain. Code is available at `https://github.com/fukua23/AuxCA`.

1 Introduction

Geometry Problem Solving [19, 21, 6, 33, 12] (GPS) is a key touchstone for evaluating MLLMs' complex reasoning capability, particularly for tasks requiring auxiliary construction. Driven by this challenge, recent works have transitioned from focusing on *low-level visual perception* to *mimicking high-level human cognition*. Early works, such as FM-ViT [15] and GeoX [32], primarily focused on optimizing visual encoders to precisely capture fine-grained geometric features like point-line relationships, or bridging the semantic gap through cross-modal alignment [40]. Subsequently, approaches represented by Visual Sketchpad [10] and DeepEyes [38] introduced the Visual Chain-of-Thought (VCoT), enabling models to mimic human behavior by explicitly generating or editing auxiliary diagrams to assist the reasoning process. Recently, inspired by the human cognitive mechanism of "combining numbers and shapes", MathCanvas [23] proposed a two-stage training framework that further strengthens the model's VCoT capabilities in Mathematics. Trained under this framework, Bagel-Canvas generates intermediate diagrams through a diffusion-style rectified-flow process and achieved state-of-the-art performance on MathCanvas-Bench. By making geometric

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

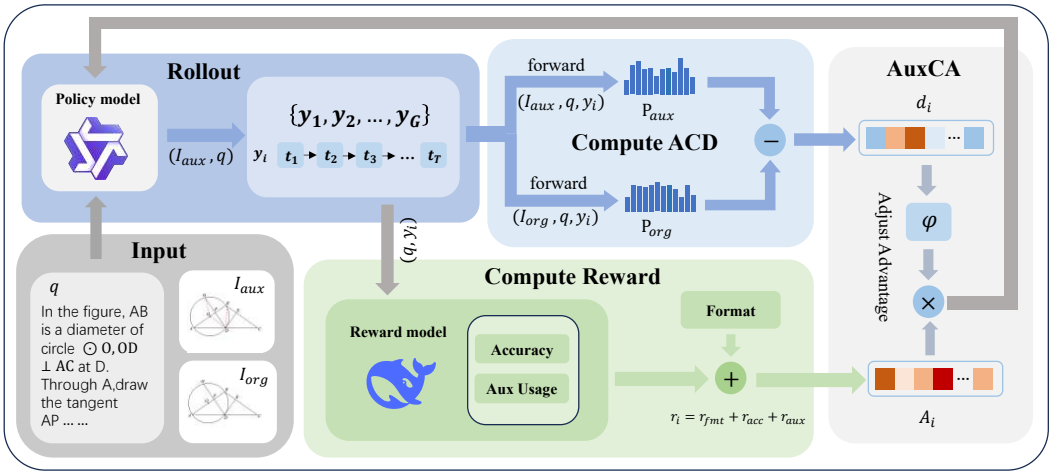


Figure 1: **The overall framework of AuxCA.** Following GRPO, the policy model first samples a group of outputs $\{y_1, \cdots, y_G\}$ conditioned on the auxiliary-enhanced input I_{aux}. We then calculate the token-level Auxiliary Clues Dependency d_i by measuring the probability shift between predictions conditioned on I_{aux} and those conditioned on I_{org} without auxiliary visual clues. The sampled sequences are then scored by the reward model to derive the sequence-level advantages A_i. Finally, the model parameters are optimized by adjusting the advantage A_i based on the dependency d_i via the function ψ, thereby helping the model learn transferable auxiliary-construction strategies.

constructions explicit, this visual reasoning paradigm improves the interpretability of the reasoning process and supports more accurate solutions to complex geometry problems.

Despite the significant potential demonstrated by VCoT, this explicit visual generation paradigm faces severe challenges, including reliance on **external tool invocation**, **high inference costs**, **unstable image generation quality**, and **error cascading**. To avoid the overhead of explicit diagramming, recent research, such as MINT-CoT [4], has begun to explore implicit visual enhancement paths, attempting to replace diagramming by improving the model's attention to specific visual patches. However, this approach suffers from a critical limitation in geometry problem-solving scenarios: the model can only attend to visual elements already present in the image. In contrast, the geometric solution process often requires introducing additional auxiliary clues (such as perpendicular lines or circumcircles) to derive the final answer. Furthermore, existing visual encoders lack prior perception of these "potential geometric structures", making it difficult for models to spontaneously focus on blank or implicit regions where auxiliary geometric structure should exist without guidance. In other words, relying solely on attention mechanisms is insufficient to enable the model to focus on the correct entry point for solving the problem from scratch.

To address these challenges, we propose **AuxCA (Auxiliary-Clues-Aware GRPO)**, a framework that helps models learn transferable auxiliary-construction strategies without generating auxiliary diagrams at inference time. Instead of relying on external plotting tools, AuxCA uses contrastive reasoning to identify **regions containing key auxiliary clues**. Specifically, we introduce **Auxiliary Clues Dependency (ACD)**, which measures the probability shift in generating the same answer when the model is conditioned on images with auxiliary clues versus the original images. In this way, ACD quantifies the dependency of each generated token on auxiliary visual information. Tokens whose probabilities increase significantly in the presence of auxiliary clues are regarded as closely associated with the additional geometric structures required for solving the problem. Based on ACD, AuxCA applies **fine-grained advantage weighting** to high-dependency tokens during GRPO-based reinforcement learning [22]. By emphasizing tokens that depend on auxiliary clues, this mechanism helps the model learn auxiliary-construction strategies that can generalize to unseen geometry problems. Moreover, we assign an additional reward when the model answers correctly while leveraging such geometric information, encouraging it to actively attend to latent auxiliary clues during reasoning. As a result, AuxCA achieves higher reasoning efficiency than explicit Visual CoT without requiring explicit diagramming.

The main contributions of this paper are as follows:

- We introduce Auxiliary Clues Dependency (ACD), a token-level measure of how auxiliary geometric clues change the probability of a generated token.
- Based on ACD, we propose AuxCA, a framework that reweights token-level advantages in GRPO. This helps the model learn auxiliary-construction strategies that can generalize to unseen geometry problems without relying on explicit plotting at inference time.
- We train AuxCA-8B and demonstrate substantial gains on auxiliary-dependent plane geometry, with improvements extending to analytic geometry, solid geometry, and external benchmarks.

2 Related Work

Geometry Problem Solving. Research in the GPS [26, 5, 39, 17, 16, 37] has witnessed significant progress in recent years, with its primary evolutionary trajectory shifting from enhancing low-level visual perception and optimizing cross-modal alignment toward unleashing models' active visual reasoning and generative capabilities. Addressing the limitations of traditional visual encoders in capturing high-frequency details of geometric figures, such as point-line relationships, FM-ViT [15] and GeoX [32] have significantly improved perception quality through frequency-domain modulation mechanisms and domain-specific masked autoencoder pre-training, respectively. Meanwhile, Math-PUMA [40] leverages the strong logical capabilities of plain text to supervise the visual modality, narrowing the semantic gap between modalities via progressive alignment. To mimic the human cognitive process of "thinking while drawing", Visual Sketchpad [10] and MathCanvas [23] introduced the Visual Chain-of-Thought (VCoT), enabling models to explicitly facilitate reasoning by generating or editing auxiliary lines and diagrams. Furthermore, MINT-CoT [4] overcomes the limitations of coarse-grained bounding boxes through an interleaved visual token selection mechanism, achieving pixel-level fine-grained interaction between image and text. Finally, inspired by DeepSeek-R1 [22], works such as DeepEyes [38], LMM-R1 [20], and VisTA [11] have incorporated reinforcement learning into multimodal reasoning. By utilizing outcome feedback to incentivize active visual exploration, self-correction, and autonomous tool usage, these approaches are propelling the field from mere perceptual understanding toward verifiable, complex logical reasoning.

Improvements to GRPO. To address the issues of restricted gradient updates and low sample efficiency in large-scale training, DAPO [35] proposed a Decoupled Clip mechanism to allow for larger policy updates on advantaged samples. Combined with a Dynamic Sampling strategy, this approach effectively accelerates training convergence. Meanwhile, to tackle the "difficulty bias", where simple samples dominate gradient updates during the later stages of training, UniVG-R1 [3] introduced a Difficulty-Aware Weight Adjustment strategy. By dynamically scaling gradients based on sample performance, it successfully incentivizes models to conquer more challenging samples in universal visual grounding tasks. Furthermore, in terms of enhancing the precision of complex reasoning and multimodal alignment, researchers have identified the drawback of GRPO applying uniform optimization pressure across all tokens. HICRA [28] reveals the hierarchical distinction between high-level strategic planning and low-level program execution, resolving the issue of learning signal dilution via hierarchy-aware credit assignment. VICRA [9] addresses the bottleneck of stagnating visual perception capabilities in MLLMs; by quantifying the visual dependency of tokens and reallocating optimization weights, it effectively corrects the optimization imbalance between textual and visual modalities. In addition, TR-GRPO [14] places greater emphasis on high-probability tokens, whereas TW-GRPO [8] improves GRPO by assigning higher rewards to high-entropy tokens.

3 Method

In this section, we present the details of our proposed method, AuxCA. In Section 3.1, we define the auxiliary clues dependency for tokens. Section 3.2 describes how we utilize this dependency to perform fine-grained optimization on GRPO, thereby enhancing the model's capability in solving geometry problems. Finally, Section 3.3 details the reward design tailored for GPS scenarios.

3.1 Auxiliary Clues Dependency

The geometry problem-solving task is typically formalized as follows: given the problem text q and the original image I_{org}, the goal is for the model M to generate the corresponding solution y, denoted as $y = M(I, q)$. However, when dealing with complex geometry problems, introducing an image containing auxiliary clues, denoted as I_{aux}, often plays a decisive role. It provides critical directional guidance and significantly improves solution accuracy. This improvement essentially stems from the model focusing more on specific tokens that carry important auxiliary information during generation. To capture this characteristic, we define Auxiliary Clues Dependency (ACD) to measure the sensitivity of each token to auxiliary geometric elements during the generation process.

We construct ACD by leveraging the probability shift of specific tokens generated by the model under different visual inputs (I_{org} versus I_{aux}). First, we input the image with auxiliary clues I_{aux} and the problem text q into the model to obtain the target solution sequence y_{aux} under auxiliary information enhancement, along with the probability of each token $\pi_\theta(y_{\text{aux},t} \mid I_{\text{aux}}, q, y_{\text{aux},<t})$. Subsequently, to deconstruct the contribution of auxiliary clues, we perform counterfactual inference. We measure the likelihood of the model generating the same token $y_{\text{aux},t}$ without auxiliary hints (*i.e.*, using the original image I_{org}). To do this, while keeping the solution sequence y_{aux} unchanged, we replace the input image with I_{org} and perform forward propagation again to calculate the corresponding conditional probability $\pi_\theta(y_{\text{aux},t} \mid I_{\text{org}}, q, y_{\text{aux},<t})$. Finally, we quantify the dependency of the token generated at step t on auxiliary clues using Eq. 1:

$$d_t = \pi_\theta(y_{\text{aux},t} \mid I_{\text{aux}}, q, y_{\text{aux},<t}) - \pi_\theta(y_{\text{aux},t} \mid I_{\text{org}}, q, y_{\text{aux},<t}), \tag{1}$$

where $d_t \in [-1, 1]$. A value closer to 1 indicates that the generation of this token relies heavily on the extra visual clues provided by the auxiliary image I_{aux}. Conversely, a smaller d_t suggests that the generation of the token does not depend on auxiliary clues. In the next section, we will detail how to utilize ACD to optimize the problem-solving capability of the model.

3.2 Auxiliary-Clues-Aware GRPO

The success of DeepSeek-Math [22] validates the efficiency of GRPO. Formally, for a given input, GRPO samples a group of candidate answers $\{y_1, y_2, \cdots, y_G\}$ from the policy model, where G denotes the group size. Subsequently, each generated answer is evaluated by a rule-based validator or a reward model to obtain rewards $\{r_1, r_2, \cdots, r_G\}$. The advantage value A_i for each answer is then computed by normalizing the rewards against the group reward mean. The training objective of GRPO is to maximize the expectation of these relative advantages, as formulated below:

$$\begin{aligned}\mathcal{J}_{\text{GRPO}}(\theta) = \mathbb{E}_{(I,q)\sim P(Q),\{y_i\}\sim\pi_{\theta_{\text{old}}}} \Bigg[\quad \frac{1}{G}\sum_{i=1}^{G}\frac{1}{|y_i|}\sum_{t=1}^{|y_i|}\Bigg\{ \\ \min\left(\frac{\pi_\theta(y_{i,t}|\cdot)}{\pi_{\theta_{\text{old}}}(y_{i,t}|\cdot)}A_{i,t}, \quad \text{clip}\left(\frac{\pi_\theta(y_{i,t}|\cdot)}{\pi_{\theta_{\text{old}}}(y_{i,t}|\cdot)}, 1-\epsilon, 1+\epsilon\right)A_{i,t}\right) - \beta\,\mathbb{D}_{\text{KL}}[\pi_\theta \parallel \pi_{\text{ref}}]\Bigg\}\Bigg],\end{aligned} \tag{2}$$

where ε and β are hyperparameters. $A_{i,t}$ represents the relative advantage, calculated as:

$$A_{i,t} = \frac{r_i - \text{mean}(\{r_1, r_2, \cdots, r_G\})}{\text{std}(\{r_1, r_2, \cdots, r_G\})}, \tag{3}$$

As seen in Eq. 2, the original GRPO implicitly assumes that all tokens within the same answer sequence contribute equally, thus assigning them identical advantage values. However, in geometry reasoning tasks, the importance of tokens is distributed unevenly. Tokens that carry not only semantic information but also critical clues about auxiliary geometric constructions often play a decisive role in the solution path.

To address this limitation, we propose a Token-level Fine-grained Advantage Weighting strategy, leveraging the ACD. The core idea is to dynamically adjust the advantage value of a token based on its dependency on auxiliary clues during generation. Specifically, if a token relies heavily on auxiliary information (indicated by a high ACD), we amplify its advantage value. Emphasizing these key tokens helps the model learn auxiliary-construction strategies that can generalize to unseen geometry problems without requiring explicit diagram generation at inference time. Conversely, for tokens with low dependency, the advantage is suppressed. This mechanism aims to guide the model to utilize auxiliary clues more effectively.

The specific implementation process is illustrated in Figure 1. First, we sample a group of responses from the policy model M given the input $(q, I_{\text{aux}}, I_{\text{org}})$, following the original GRPO formulation. Subsequently, we calculate the auxiliary clues dependency $\{\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_G\}$ for each generated sequence. Next, we introduce a mapping function ψ to transform the dependency obtained in the previous step into a relative advantage weighting coefficient $\psi_{i,t}$ for each token. Finally, we adjust the original advantage values token-by-token using this coefficient. The corrected advantage value is:

$$A_{i,t}^{\text{AuxCA}} = \psi(d_{i,t}) \cdot A_{i,t}, \tag{4}$$

where ψ is defined as:

$$\psi(d_{i,t}) = 1 + \left(d_{i,t} - \frac{1}{|y_i|} \sum_{t=1}^{|y_i|} d_{i,t} \right). \tag{5}$$

Through mathematical derivation (refer to the Appendix C for details), we demonstrate that this weighting strategy possesses the property of Total Advantage Conservation. That is, for the same answer sequence, the cumulative advantage value remains unchanged before and after reweighting:

$$\sum_{t=1}^{|y_i|} A_{i,t} = \sum_{t=1}^{|y_i|} \psi(d_{i,t}) \cdot A_{i,t}. \tag{6}$$

This equation indicates that the function of ψ is not to scale the global advantage signal of the entire answer. Instead, it performs a fine-grained redistribution of advantage within the sequence while keeping the total amount constant. Through comparison with another exponential function (see Appendix D), we find that this property can effectively prevent training instability caused by advantage drift, which is the core motivation for selecting this form as the mapping function.

3.3 Reward Design

Due to the unique nature of geometry proof problems, traditional rule-based matching reward functions are often inapplicable. On one hand, questions typically contain multiple sub-problems; on the other hand, the correct solution path is often not unique. To address this, we design a multi-dimensional reward mechanism that covers correctness, auxiliary clues incentives, and format specifications.

First, given the complexity and openness of the solution space, we abandon simple rule matching in favor of using an LLM as a judge. Following the standards of MathCanvas-Bench, we score each sub-problem independently. If the model answers all sub-problems correctly, it receives the full reward; if only a subset is correct, it receives partial points based on weights. This dense feedback signal enables the model to obtain positive reinforcement from partially correct reasoning, even if it has not fully solved the entire problem. Detailed prompts are provided in the Appendix A.1.

Second, the ACD-based advantage weighting strategy in Section 3.2 helps the model learn transferable auxiliary-construction strategies, but it does not explicitly reward the effective use of these clues in a correct solution. We need to encourage the model to leverage these geometric clues for deduction, thereby achieving high accuracy without relying on explicit diagramming. To this end, we introduce a sparse reward term r_{aux}. Specifically, we employ an LLM to examine the reasoning text generated by the model. This reward is triggered only when the model satisfies two conditions simultaneously: answering all sub-problems correctly and explicitly referencing the geometric relationships provided

Table 1: **Main results on MathCanvas-Bench.** We report the Complete and Weighted accuracy scores across three subsets (Aux-PlaGeo, AnaGeo, and SolidGeo) and the overall dataset. "✓" in the "Think" column indicates models that utilize chain-of-thought or reasoning tokens. The best closed-source and open-source highest accuracy of MLLMs are marked in red and blue, respectively.

Model	Size	Think	Aux-PlaGeo		AnaGeo		SolidGeo		overall	
			Complete	Weighted	Complete	Weighted	Complete	Weighted	Complete	Weighted
Closed Source										
Gemini-2.5-Pro	-	✓	40.2	54.5	47.4	58.6	35.4	48.8	48.3	59.4
Gemini-2.5-Flash	-	✗	29.4	39.8	44.9	55.7	31.2	40.6	38.9	48.8
GPT-4.1	-	✗	7.7	18.7	20.9	30.3	8.9	23.8	18.2	28.9
GPT-4o	-	✗	7.1	17.1	10.5	18.2	5.8	18.1	9.8	18.8
GPT-5	-	✓	32.4	43.8	47.8	56.2	26.1	35.4	42.7	50.9
Claude-Sonnet-4	-	✓	24.1	31.8	29.4	37.2	22.8	32.4	24.3	36.4
Seed-1.6-Thinking	-	✓	39.8	53.3	42.3	57.6	36.1	45.9	44.5	55.3
Qwen3-VL-Plus	-	✓	34.6	44.5	41.6	54.3	34.7	43.2	41.2	52.1
Open Source										
Qwen-2.5-VL-Instruct	7B	✗	7.2	16.9	9.7	16.4	6.0	19.1	9.2	17.9
Qwen-3-VL-Instruct	8B	✗	13.1	23.0	24.4	34.9	11.3	25.6	19.6	29.6
Qwen-3-VL-Thinking	8B	✓	18.4	24.6	28.1	33.4	14.6	27.4	26.9	33.3
Qwen-3-VL-Thinking	32B	✓	28.3	35.1	43.3	49.2	23.9	36.3	36.9	44.4
Gemma-3-it	27B	✗	11.3	22.5	16.8	25.8	7.6	21.4	14.9	25.4
InternVL3.5	8B	✓	14.5	19.3	22.8	27.8	11.7	22.8	20.9	26.9
InternVL3.5-A3B	30B	✓	19.8	30.4	36.1	43.8	16.1	33.0	28.8	39.3
Keye-VL-1.5	8B	✓	18.2	27.1	15.8	31.4	13.4	27.9	22.2	31.2
Bagel-Canvas	14B	✗	27.2	39.9	16.0	27.2	18.0	35.3	21.9	34.4
AuxCA	8B	✓	31.7	44.6	31.2	41.0	22.8	40.4	30.3	42.2
Δ Over Base Model			+13.3	+20.0	+3.1	+7.6	+8.2	+13.0	+3.4	+8.9

by the auxiliary clues. This design reinforces the value of auxiliary clues at a macro behavioral level. Experiments (see Appendix E) indicate that model performance is optimal when $r_{\text{aux}} = 0.2$.

Finally, to enforce output structure, we introduce r_{fmt} to guide the model to adhere to predefined output specifications (such as using `<think>` tags). Consequently, the final reward function is defined as:

$$r = r_{\text{fmt}} + r_{\text{acc}} + r_{\text{aux}}. \tag{7}$$

4 Experiments

Dataset This study focuses on solving plane geometry problems that are **Auxiliary-dependent**. To this end, we filtered and constructed a dedicated training set containing 70k samples from the plane geometry subset of MathCanvas-Instruct [23]. All these samples require the introduction of additional auxiliary geometric structures to complete the reasoning process (detailed selection are provided in the Appendix A.2 and Appendix B).

Benchmark During evaluation, we extracted 635 data points requiring the addition of extra auxiliary elements to the original image from MathCanvas-Bench to serve as the core test set (`Aux-PlaGeo`). Furthermore, to comprehensively measure model performance, we not only conducted extensive evaluations on two other related subsets Analytical geometry (`AnaGeo`), Solid geometry (`SolidGeo`) and the full set (`overall`) of this benchmark (see Table 1), but also performed generalization analysis on mainstream benchmarks such as MathVista [18], MathVerse [36], and MathVision [29, 30] (see Table 2). Performance is measured via Complete Accuracy, reflecting a strict all-or-nothing criterion for complete solutions, and Weighted Accuracy, which captures partial progress through fine-grained exponential weights.

Baseline To provide a comprehensive evaluation of AuxCA, we benchmarked it against leading state-of-the-art proprietary models, such as the Gemini-2.5 [7] series (Gemini-2.5-Pro, Gemini-2.5-Flash), the ChatGPT series (ChatGPT-4o [13], ChatGPT-4.1 [13], ChatGPT-5 [24]), Claude-Sonnet-4[1], Seed-1.6-Thinking[2], and Qwen3-VL-Plus [1]. Additionally, we included a diverse range of mainstream open-source models with parameters spanning from 7B to 32B. This selection specifically covers

[1] `https://www.anthropic.com/claude/sonnet`
[2] `https://seed.bytedance.com/en/seed1_6`

the Qwen series (Qwen-2.5-VL-Instruct [2], Qwen-3-VL-Instruct/Thinking [1]), the InternVL series (InternVL3.5-8B/30B) [31], Gemma-3-it [25], Keye-VL-1.5 [34], and Bagel-Canvas [23].

Implementation Details The base model used in the experiments is Qwen3-VL-8B-Thinking, while the reward model is DeepSeek-V3.2. All experiments were implemented using the `trl` [27] framework; specific hyperparameter settings and additional implementation details are provided in Appendix F. To ensure training stability, we performed Supervised Fine-Tuning (SFT) on the model for 2 epochs using the aforementioned 70k dataset. Subsequently, we randomly sampled 15k samples from this dataset for AuxCA reinforcement learning. Through this training, the model improves its sensitivity to auxiliary geometric elements and learns transferable auxiliary-construction strategies that can be applied beyond the training domain. We name the resulting model **AuxCA-8B**.

4.1 Main Experiments

Table 1 presents the performance of our model, AuxCA-8B, across various subsets and the full MathCanvas-Bench. The experimental results provide strong validation for the effectiveness of the reinforcement learning strategy proposed in this paper.

Learning Transferable Auxiliary-Construction Strategies As shown in Table 1, AuxCA consistently outperforms other models with comparable parameter sizes across all benchmarks, as well as larger models such as Bagel-Canvas, InternVL3.5-30B-A3B, and Gemma-3-27b-it. On the `Aux-PlaGeo` dataset, AuxCA achieves a substantial improvement of +20.0 in the Weighted Score (from 24.6 to 44.6). Beyond its strong performance in the training domain of plane geometry, AuxCA also demonstrates generalization to analytic geometry (`AnaGeo`) and solid geometry (`SolidGeo`). Specifically, on `SolidGeo`, AuxCA improves the Weighted Score by +13.0 over the base model, reaching 40.4. This result surpasses that of Qwen-3-VL-Thinking 32B (36.3), a model with four times more parameters. These performance gains suggest that incorporating the Auxiliary Clues Dependency (ACD) mechanism and a fine-grained advantage reweighting strategy helps AuxCA learn transferable auxiliary-construction strategies that can be applied beyond the training domain.

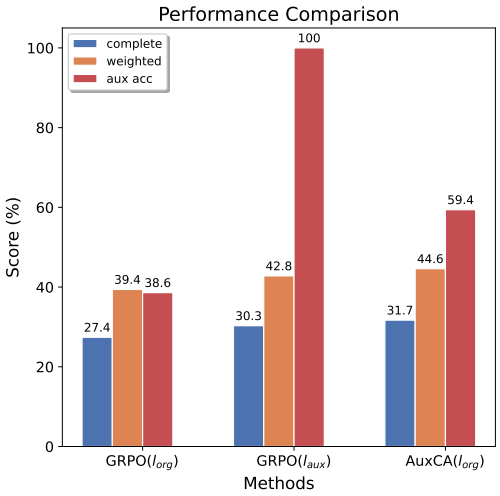


Figure 2: **Analysis of auxiliary construction and utilization.** *GRPO(I_{aux})* is evaluated with images containing correct auxiliary clues at test time, while all other methods use the original problem images. *aux acc* denotes the fraction of samples for which the auxiliary clues are correctly constructed.

To further assess the model's ability to propose and use auxiliary constructions, we evaluate the GRPO-trained model with an image augmented with correct auxiliary clues, denoted as I_{aux}, in place of the original input image I_{org}. We use this setting as a strong baseline. We also measure the proportion of samples in which AuxCA and GRPO correctly construct auxiliary clues when reasoning from the original image, referred to as auxiliary construction accuracy (aux acc). The results are shown in Figure 2. After AuxCA-based reinforcement learning, the model achieves a 20.8% higher auxiliary construction accuracy than GRPO, suggesting that AuxCA helps the model learn more effective auxiliary-construction strategies. Moreover, even when provided with correct auxiliary clues, GRPO still performs slightly worse than AuxCA, which uses only the original image. Together, these results suggest that AuxCA training improves the model's ability to propose useful auxiliary constructions and reason with the resulting geometric relationships.

Superiority over Explicit Plotting To compare auxiliary-construction strategies expressed through textual reasoning with explicit plotting, we evaluate AuxCA-8B against Bagel-Canvas (14B with 7B active parameters), a strong explicit-plotting baseline. Bagel-Canvas generates auxiliary diagrams to assist problem-solving, whereas AuxCA expresses auxiliary constructions through chain-of-thought reasoning without generating additional images. Despite having fewer total parameters (8B vs. 14B), AuxCA outperforms Bagel-Canvas on both Complete and Weighted Scores across all test subsets

Table 2: **Generalization capabilities on external benchmarks.**

(a) MathVista

Model	GPS
Base	68.7
AuxCA	**71.6**
Improve	+2.9

(b) MathVerse

Model	Text Dom	Text Lite
Base	69.3	65.1
AuxCA	**76.5**	**67.0**
Improve	+7.2	+1.9

(c) MathVision

Model	Angle	Area	Length
Base	22.5	21.2	23.2
AuxCA	**31.0**	**25.8**	**28.9**
Improve	+8.5	+4.6	+5.7

of MathCanvas-Bench. Notably, on the Weighted Score of the full set (`overall`), AuxCA achieves 42.2, compared with 34.4 for Bagel-Canvas. Under the same hardware and inference settings, AuxCA also reduces average end-to-end latency by 65.4%, despite generating more text tokens; details are provided in Appendix I. These results suggest that learning auxiliary-construction strategies through textual reasoning offers an effective alternative to explicit plotting, achieving higher solution accuracy in these evaluations while avoiding the need for additional image generation.

Comparison with Leading Closed-Source Models On the `Aux-PlaGeo` subset, which specifically evaluates problems requiring auxiliary constructions, AuxCA-8B achieves performance competitive with leading closed-source models. Its Weighted Score (44.6) exceeds those of GPT-4o (17.1), Claude-Sonnet-4 (31.8), and GPT-5 (43.8). These results suggest that domain-specific training data and targeted reinforcement learning can help smaller models learn effective auxiliary-construction strategies and achieve competitive performance on challenging geometry reasoning tasks. Together with the gains on `AnaGeo` and `SolidGeo`, they provide evidence that these strategies can transfer beyond the plane geometry training domain.

4.2 Ablation Study

To validate the independent contributions of each core component in our framework, we conduct ablation studies as shown in Table 3. First, we examine the effect of removing the Auxiliary Clues Dependency (ACD) mechanism (w/o ACD). Without this module, the algorithm degenerates into the standard GRPO with uniform advantage assignment, resulting in a drop of 5.2 in the Weighted Score (from 44.6 to 39.4) and a decrease of 4.3 in the Complete Score (from 31.7 to 27.4). These results support the effectiveness of our fine-grained reweighting strategy in improving geometric reasoning. Complementary analysis in Appendix J shows that 97 of the 100 highest-scoring tokens are manually identified as auxiliary-clue or construction-action tokens, providing further evidence that ACD captures content relevant to auxiliary construction. Next, we evaluate the impact of removing the auxiliary reward term (w/o r_{aux}). The complete score drops from 31.7 to 25.8, and the weighted score decreases from 44.6 to 38.1, indicating that an explicit reward signal is crucial for guiding the model to actively utilize auxiliary information during reasoning. Finally, we compare AuxCA with the SFT-only baseline without subsequent reinforcement learning, which achieves lower scores on both metrics, further supporting the effectiveness of the AuxCA reinforcement learning stage.

Table 3: **Ablation study on the core components of AuxCA.**

Model	Aux-PlaGeo	
	Complete	Weighted
AuxCA	**31.7**	**44.6**
- w/o ACD(GRPO)	27.4	39.4
- w/o r_{aux}	25.8	38.1
- SFT only	22.1	33.9
Base model	18.4	24.6

We attribute this to the fact that SFT and GRPO are trained directly with I_{aux} as input, in which the correct auxiliary clues are already provided in the image. As a result, these models primarily learn to reason using the given auxiliary information. In contrast, AuxCA leverages both I_{org} and I_{aux} to adjust token-level advantages. This not only strengthens the model's ability to reason effectively with correct auxiliary information, but also enhances its capability to construct auxiliary clues.

4.3 Generalization Analysis

To further evaluate the generalization capability of AuxCA, we conduct zero-shot evaluations on MathVista, MathVerse, and MathVision, as shown in Table 2. The results demonstrate that AuxCA consistently outperforms the base model across all three benchmarks, indicating that the learned auxiliary-construction strategies can transfer to unseen problems. Specifically, on the Text Dominant

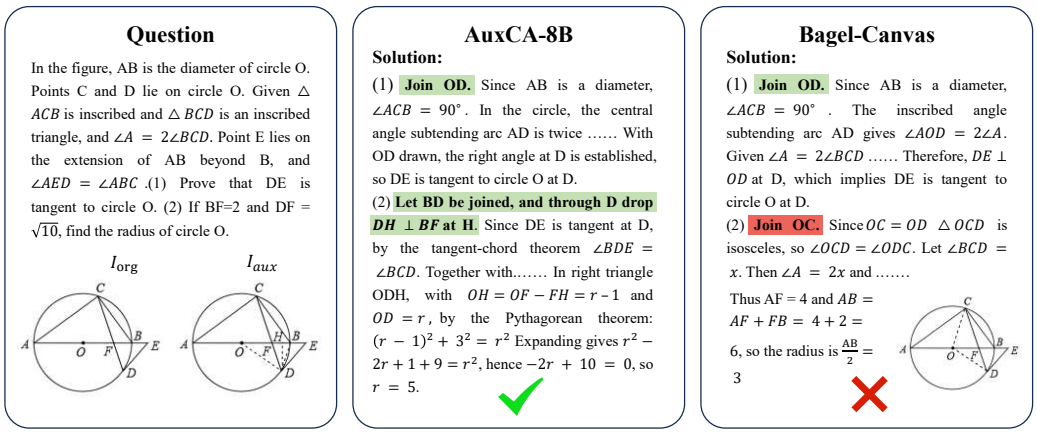


Figure 3: **Performance comparison with Bagel-Canvas on a complex geometric circle problem.**

subset of MathVerse, AuxCA achieves a 7.2-point improvement, substantially exceeding the gain observed on the Text Lite subset. This suggests that the auxiliary-construction strategies learned during training may also facilitate reasoning in settings with limited visual information, where geometric relationships must be inferred from incomplete visual cues. Furthermore, a fine-grained analysis of MathVision reveals notable improvements in tasks that often benefit from auxiliary constructions, including Angle (+8.5), Length (+5.7), and Area (+4.6). These findings suggest that AuxCA can generalize learned construction strategies, such as drawing altitudes and extending line segments, to diverse geometric reasoning tasks, potentially improving reasoning robustness and accuracy beyond the training domain. We also evaluate AuxCA-8B on problems that do not require additional auxiliary constructions, as shown in Appendix G.

4.4 Comparison with Generic Credit Assignment Methods

We conducted additional experiments comparing ACD with four generic credit assignment methods, as shown in Table 4. We observe that VICRA, which amplifies the advantages of image-related tokens, and HICRA, which amplifies the advantages of planning-related reasoning tokens, both achieve slight improvements over SFT, but still fall far short of AuxCA. In contrast, TR-GRPO, which assigns larger weights to higher-probability tokens, and TW-GRPO, which places greater emphasis on high-entropy tokens, both show performance degradation relative to SFT. This indicates that our mechanism, which reallocates token-level advantages based on the difference in output token probabilities with and without auxiliary clues, is better suited to geometry problem solving than these more generic credit assignment methods.

Table 4: **Comparison with Generic Credit Assignment Methods.**

Method	Aux-PlaGeo	
	Complete	Weighted
AuxCA	**31.7**	**44.6**
VICRA	24.8	37.2
HICRA	24.3	36.1
TR-GRPO	18.9	30.2
Tw-GRPO	17.6	28.7

4.5 Case Study

To illustrate AuxCA's superiority in handling complex auxiliary constructions, Figure 3 presents a representative circle geometry problem. In the foundational first part, both AuxCA and the explicit construction model, Bagel-Canvas, accurately identified the key clue (connecting OD) to complete the proof, confirming that both possess basic geometric perception. The performance gap widens in the challenging second part, which requires two non-intuitive auxiliary lines (connecting BD and dropping $DH \perp BF$) to form a right triangle. Despite generating no image, AuxCA precisely described this step in its reasoning text "Let BD be joined, and through D drop $DH \perp BF$" and correctly derived $r = 5$. This example illustrates how AuxCA can propose useful auxiliary constructions through textual reasoning without generating additional diagrams. In contrast, Bagel-Canvas fell into a visual trap by explicitly drawing an ineffective line (OC). This misleading visual feedback caused "Error

Cascading" and a calculation failure. Ultimately, this demonstrates that AuxCA avoids the pitfalls of explicit plotting, exhibiting more robust geometric structuring capabilities than explicit VCoT in complex, non-visual scenarios. In addition, we also analyze the failure cases of AuxCA-8B, with detailed results provided in the Appendix H.

5 Conclusion

We present AuxCA, a reinforcement learning framework that helps models learn transferable auxiliary-construction strategies for geometric reasoning without explicit plotting at inference time. By combining Auxiliary Clues Dependency (ACD) for fine-grained advantage weighting with an auxiliary incentive reward, AuxCA improves problem-solving performance. Extensive experiments demonstrate its effectiveness and generalization beyond the plane geometry training domain.

Acknowledgments

This work was supported by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang (Grant No. 2025C02022), the Ningbo "Yongjiang Talent Program" Youth Innovation Project (Grant No. 2024A-156-G), and the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 62507040).

References

[1] 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, 2025.

[2] 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, 2025.

[3] Sule Bai, Mingxing Li, Yong Liu, Jing Tang, Haoji Zhang, Lei Sun, Xiangxiang Chu, and Yansong Tang. Univg-r1: Reasoning guided universal visual grounding with reinforcement learning. *arXiv preprint arXiv:2505.14231*, 2025.

[4] Xinyan Chen, Renrui Zhang, Dongzhi Jiang, Aojun Zhou, Shilin Yan, Weifeng Lin, and Hongsheng Li. Mint-cot: Enabling interleaved visual tokens in mathematical chain-of-thought reasoning. *arXiv preprint arXiv:2506.05331*, 2025.

[5] Seunghyuk Cho, Zhenyue Qin, Yang Liu, Youngbin Choi, Seungbeom Lee, and Dongwoo Kim. Geodano: Geometric vlm with domain agnostic vision encoder. *arXiv preprint arXiv:2502.11360*, 2025.

[6] Seunghyuk Cho, Zhenyue Qin, Yang Liu, Youngbin Choi, Seungbeom Lee, and Dongwoo Kim. Plane geometry problem solving with multi-modal reasoning: A survey. In *Findings of the Association for Computational Linguistics: EACL 2026*, pages 110–131, 2026.

[7] Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

[8] Jisheng Dang, Jingze Wu, Teng Wang, Xuanhui Lin, Nannan Zhu, Hongbo Chen, Wei-Shi Zheng, Meng Wang, and Tat-Seng Chua. Reinforcing video reasoning with focused thinking, 2025.

[9] Jiale Hong, Dongjie Yang, Xuecai Hu, Yong Wang, Xiangxiang Chu, and hai zhao. Don't lose sight: Visually-grounded credit assignment for multimodal reasoning, 2026.

[10] Yushi Hu, Weijia Shi, Xingyu Fu, Dan Roth, Mari Ostendorf, Luke Zettlemoyer, Noah A Smith, and Ranjay Krishna. Visual sketchpad: Sketching as a visual chain of thought for multimodal language models. *Advances in Neural Information Processing Systems*, 37:139348–139379, 2024.

[11] Zeyi Huang, Yuyang Ji, Anirudh Sundara Rajan, Zefan Cai, Wen Xiao, Haohan Wang, Junjie Hu, and Yong Jae Lee. Visualtoolagent (vista): A reinforcement learning framework for visual tool selection. *arXiv preprint arXiv:2505.20289*, 2025.

[12] Zihan Huang, Tao Wu, Wang Lin, Shengyu Zhang, Jingyuan Chen, and Fei Wu. Autogeo: Automating geometric image dataset creation for enhanced geometry understanding. *IEEE Transactions on Multimedia*, 2025.

[13] 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.

[14] Tue Le, Nghi D. Q. Bui, Linh Ngo Van, and Trung Le. Token-regulated group relative policy optimization for stable reinforcement learning in large language models, 2025.

[15] Wang Lin, QingSong Wang, Yueying Feng, Shulei Wang, Tao Jin, Zhou Zhao, Fei Wu, Chang Yao, and Jingyuan Chen. Non-natural image understanding with advancing frequency-based vision encoders. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 29756–29766, 2025.

[16] Yong Lin, Shange Tang, Bohan Lyu, Jiayun Wu, Hongzhou Lin, Kaiyu Yang, Jia Li, Mengzhou Xia, Danqi Chen, Sanjeev Arora, et al. Goedel-prover: A frontier model for open-source automated theorem proving. *arXiv preprint arXiv:2502.07640*, 2025.

[17] Yong Lin, Shange Tang, Bohan Lyu, Ziran Yang, Jui-Hui Chung, Haoyu Zhao, Lai Jiang, Yihan Geng, Jiawei Ge, Jingruo Sun, et al. Goedel-prover-v2: Scaling formal theorem proving with scaffolded data synthesis and self-correction. *arXiv preprint arXiv:2508.03613*, 2025.

[18] 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 (ICLR)*, 2024.

[19] Jianzhe Ma, Wenxuan Wang, and Qin Jin. A survey of deep learning for geometry problem solving. *arXiv preprint arXiv:2507.11936*, 2025.

[20] Yingzhe Peng, Gongrui Zhang, Miaosen Zhang, Zhiyuan You, Jie Liu, Qipeng Zhu, Kai Yang, Xingzhong Xu, Xin Geng, and Xu Yang. Lmm-r1: Empowering 3b lmms with strong reasoning abilities through two-stage rule-based rl. *arXiv preprint arXiv:2503.07536*, 2025.

[21] Bowen Ping, Minnan Luo, Zhuohang Dang, Chenxi Wang, and Chengyou Jia. Autogps: Automated geometry problem solving via multimodal formalization and deductive reasoning. *arXiv preprint arXiv:2505.23381*, 2025.

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

[23] Weikang Shi, Aldrich Yu, Rongyao Fang, Houxing Ren, Ke Wang, Aojun Zhou, Changyao Tian, Xinyu Fu, Yuxuan Hu, Zimu Lu, et al. Mathcanvas: Intrinsic visual chain-of-thought for multimodal mathematical reasoning. *arXiv preprint arXiv:2510.14958*, 2025.

[24] Aaditya Singh, Adam Fry, Adam Perelman, Adam Tart, Adi Ganesh, Ahmed El-Kishky, Aidan McLaughlin, Aiden Low, AJ Ostrow, Akhila Ananthram, et al. Openai gpt-5 system card. *arXiv preprint arXiv:2601.03267*, 2025.

[25] Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, et al. Gemma 3 technical report. *arXiv preprint arXiv:2503.19786*, 2025.

[26] Trieu H Trinh, Yuhuai Wu, Quoc V Le, He He, and Thang Luong. Solving olympiad geometry without human demonstrations. *Nature*, 625(7995):476–482, 2024.

[27] Leandro von Werra, Younes Belkada, Lewis Tunstall, Edward Beeching, Tristan Thrush, Nathan Lambert, Shengyi Huang, Kashif Rasul, and Quentin Gallouédec. TRL: Transformers Reinforcement Learning, 2020.

[28] Haozhe Wang, Qixin Xu, Che Liu, Junhong Wu, Fangzhen Lin, and Wenhu Chen. Emergent hierarchical reasoning in llms through reinforcement learning. *arXiv preprint:2509.03646*, 2025.

[29] 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 *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024.

[30] Ke Wang, Junting Pan, Linda Wei, Aojun Zhou, Weikang Shi, Zimu Lu, Han Xiao, Yunqiao Yang, Houxing Ren, Mingjie Zhan, and Hongsheng Li. Mathcoder-VL: Bridging vision and code for enhanced multimodal mathematical reasoning. In *The 63rd Annual Meeting of the Association for Computational Linguistics*, 2025.

[31] Weiyun Wang, Zhangwei Gao, Lixin Gu, Hengjun Pu, Long Cui, Xingguang Wei, Zhaoyang Liu, Linglin Jing, Shenglong Ye, Jie Shao, et al. Internvl3. 5: Advancing open-source multimodal models in versatility, reasoning, and efficiency. *arXiv preprint arXiv:2508.18265*, 2025.

[32] Renqiu Xia, Mingsheng Li, Hancheng Ye, Wenjie Wu, Hongbin Zhou, Jiakang Yuan, Tianshuo Peng, Xinyu Cai, Xiangchao Yan, Bin Wang, et al. Geox: Geometric problem solving through unified formalized vision-language pre-training. *arXiv preprint arXiv:2412.11863*, 2024.

[33] Liangyu Xu, Yingxiu Zhao, Jingyun Wang, Yingyao Wang, Bu Pi, Chen Wang, Mingliang Zhang, Jihao Gu, Xiang Li, Xiaoyong Zhu, et al. Geosense: Evaluating identification and application of geometric principles in multimodal reasoning. *arXiv preprint arXiv:2504.12597*, 2025.

[34] Biao Yang, Bin Wen, Boyang Ding, Changyi Liu, Chenglong Chu, Chengru Song, Chongling Rao, Chuan Yi, Da Li, Dunju Zang, et al. Kwai keye-vl 1.5 technical report. *arXiv preprint arXiv:2509.01563*, 2025.

[35] Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, Lingjun Liu, et al. Dapo: An open-source llm reinforcement learning system at scale. *arXiv preprint arXiv:2503.14476*, 2025.

[36] Renrui Zhang, Dongzhi Jiang, Yichi Zhang, Haokun Lin, Ziyu Guo, Pengshuo Qiu, Aojun Zhou, Pan Lu, Kai-Wei Chang, Peng Gao, et al. Mathverse: Does your multi-modal llm truly see the diagrams in visual math problems? *arXiv preprint arXiv:2403.14624*, 2024.

[37] Renrui Zhang, Xinyu Wei, Dongzhi Jiang, Ziyu Guo, Shicheng Li, Yichi Zhang, Chengzhuo Tong, Jiaming Liu, Aojun Zhou, Bin Wei, et al. Mavis: Mathematical visual instruction tuning with an automatic data engine. *arXiv preprint arXiv:2407.08739*, 2024.

[38] Ziwei Zheng, Michael Yang, Jack Hong, Chenxiao Zhao, Guohai Xu, Le Yang, Chao Shen, and Xing Yu. Deepeyes: Incentivizing" thinking with images" via reinforcement learning. *arXiv preprint arXiv:2505.14362*, 2025.

[39] Yichi Zhou, Jianqiu Zhao, Yongxin Zhang, Bohan Wang, Siran Wang, Luoxin Chen, Jiahui Wang, Haowei Chen, Allan Jie, Xinbo Zhang, et al. Solving formal math problems by decomposition and iterative reflection. *arXiv preprint arXiv:2507.15225*, 2025.

[40] Wenwen Zhuang, Xin Huang, Xiantao Zhang, and Jin Zeng. Math-puma: Progressive upward multimodal alignment to enhance mathematical reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 26183–26191, 2025.

A Prompts

A.1 Prompts used to reward model during training

You are an expert mathematics teacher and a precise data evaluator.
Your task is to analyze a given math problem, compare a predicted solution against a ground truth answer, and determine if the prediction is correct.
Additionally, you need to assess whether the predicted solution relies on visual cues or supplementary elements present in the image but not explicitly specified in the problem text.

INPUT FORMAT:
You will be provided with a JSON string containing the following fields:
- 'question_text': The full text of the mathematical problem.
- 'gt_answers': A LIST of the final answers' strings (e.g., ["5 cm", "10 cm"]). Each element is the correct, final answer for a specific sub-question. This is the gold standard and no additional extraction is required.
- 'prediction_solution': The full solution text from the model, from which you must extract the final answer(s).

TASK & OUTPUT REQUIREMENTS:
Your output must be a single, valid JSON object.
The process involves three main steps: **Answer Parsing**, **Correctness Judgment**, and **Auxiliary Element Detection**.

Step 1: Answer Parsing and Extraction
Your task is to create a list of predicted answers (named 'pred_answers') based on the given 'prediction_solution'. This list must match the provided list of correct answers ('gt_answers') exactly.
Parsing and Extraction Rules:
- Your primary task is to **extract** the final answer(s) from the 'prediction_solution' text to create the 'pred_answers' list.
- **IMPORTANT:** The answers to different sub-questions may appear in different places within the 'prediction_solution', not necessarily grouped together at the end. You must treat this as a **matching task**.
- For each part of the 'gt_answers' list, you must scan the **entire** 'prediction_solution' to find the corresponding predicted answer. Look for explicit labels (e.g., "(1)", "Part A"), final conclusions, boxed answers (e.g., \\boxed{{...}}), or statements that directly answer a part of the original question.
- The final 'pred_answers' list **must have the exact same length as the 'gt_answers' list**.
- For each sub-question, if you cannot find a corresponding answer in the 'prediction_solution', you **must** use 'null' as a placeholder in that position.
This rule is critical and applies in all cases where an answer is missing, including when the 'prediction_solution' appears incomplete or is truncated before all sub-questions are addressed.
CRITICAL RULE: The final 'gt_answers' and 'pred_answers' lists **must** be of equal length. The number of parts in 'gt_answers' dictates the required length.

Step 2: Correctness Judgment
Your second task is to compare the 'pred_answers' list against the 'gt_answers' list, element by element.
Judgment Rules:
1. **Numerical Equivalence:** Treat numbers as correct if they are mathematically equivalent (e.g., '5', '5.0', '10/2'). Allow for minor floating-point rounding differences.
2. **Textual Equivalence:** For text answers, judge based on semantic meaning, not exact matching. Ignore case, whitespace, and phrasing differences (e.g., "CB is parallel to PD" is equivalent to "Line CB || Line PD").
3. **Generate Correctness List:** Create a boolean list named 'correctness'. The i-th element is 'true' if the i-th predicted answer is correct, 'false' otherwise. This list **must** have the same length as the answer lists.

Step 3: Detection of Pre-drawn Auxiliary Elements
The input image for this problem contains pre-drawn auxiliary elements (i.e., additional visual elements not explicitly mentioned in the problem description, such as points, lines, circles, etc.), intended to assist in solving the problem.
Analyze the reasoning process of 'prediction_solution' to determine whether it utilizes any pre-drawn auxiliary elements visible in the image.
Since the image cannot be viewed, **infer from the text**:
- Does the solution reference additional visual elements not explicitly mentioned in the problem description, such as points, lines, circles, etc.?
- Does the solution rely on specific geometric relationships (e.g., similar triangles or the Pythagorean theorem) that require constructions implied by the image but not explicitly stated in 'question_text'?
- **Evaluation Criteria:** Set 'using_auxiliary' to 'true' if the solution explicitly references these auxiliary elements or their derived geometric properties not mentioned in the problem text. Otherwise, set it to 'false'.

Final JSON Output Structure:
Your entire response must be a single, valid JSON object matching the schema below. Do not include any text outside of this JSON object.
```json
{{
  "using_auxiliary": true/false,
  "correctness": [
    true/false,
    ...
  ]
}}
```
INPUT DATA:
f{input_data}
Process this mathematical problem according to all requirements above. Ensure the output is valid JSON.

A.2 Prompts used to data selection

I will present you with two geometric images. The first is the [Original Problem Image], and the second is the [Auxiliary Image].
Please observe these two images carefully.
Task: Determine whether the [Auxiliary Image] is derived merely by adding auxiliary lines (such as dashed lines, red lines, extension lines, etc.) or other letter/character markings to the [Original Problem Image], while the shape and position of the main geometric subject remain unchanged.
If yes, please answer only: YES
If the [Auxiliary Image] is a completely different image, or differs significantly from the [Original Problem Image], please answer: NO
After your reasoning, please start a new line and **output only the conclusion string (<|begin_of_box|>YES<|end_of_box|> or <|begin_of_box|>NO<|end_of_box|>)**. That is:
If it is an auxiliary line relationship, please output: <|begin_of_box|>YES<|end_of_box|>
If not, please output: <|begin_of_box|>NO<|end_of_box|>"

B Automated Filtering Pipeline

To construct a dedicated dataset for auxiliary-dependent plane geometry, we performed a rigorous filtering process on the Plane Geometry subset of *MathCanvas-Instruct*. Our objective was to isolate problems where the solution strictly necessitates the addition of auxiliary geometric structures (e.g., auxiliary lines, extended segments) rather than generating entirely new figures or solving via pure calculation.

Given the scale of the dataset, manual verification was infeasible. Therefore, we developed an automated pipeline leveraging the advanced visual reasoning capabilities of **GLM-4.5V**. For each data sample, we fed the model a pair of images:

- **Original Image (I_{org}):** The initial geometric figure provided in the problem statement.
- **Auxiliary Image (I_{aux}):** The figure parsed from the solution steps, containing potential auxiliary constructions.

The VLM was tasked with verifying whether I_{aux} is visually derived from I_{org} solely by adding auxiliary markers without altering the position or shape of the main geometric subjects.

We employed a strict prompt showing in Appendix A.2 template to ensure high precision in data selection. Only samples where the model returned the token `YES` were included in our final **Aux-PlaGeo** training set.

C Proof of Total Advantage Conservation

In this section, we provide the detailed mathematical derivation for the *Total Advantage Conservation* property presented in Eq. (6) of the main paper. This property ensures that our reweighting mechanism redistributes the advantage within a sequence based on auxiliary dependency without altering the cumulative advantage of the entire sequence, thereby maintaining training stability.

C.1 Definitions and Preliminaries

Recall the definition of the token-level weighting function $\psi(d_{i,t})$ from Eq. (5):

$$\psi(d_{i,t}) = 1 + \left(d_{i,t} - \frac{1}{|y_i|} \sum_{k=1}^{|y_i|} d_{i,k} \right), \tag{8}$$

where y_i represents the sampled sequence, $|y_i|$ denotes its length, and $d_{i,t}$ is the auxiliary dependency for the t-th token.

In the GRPO framework, the advantage $A_{i,t}$ is typically derived from the sequence-level reward. Therefore, for a given sequence y_i, the advantage value is constant across all tokens:

$$A_{i,t} = A_i, \quad \forall t \in \{1, \ldots, |y_i|\}. \tag{9}$$

C.2 Derivation

We aim to prove that the total advantage remains unchanged after reweighting:

$$\sum_{t=1}^{|y_i|} A_{i,t} = \sum_{t=1}^{|y_i|} \psi(d_{i,t}) \cdot A_{i,t}. \tag{10}$$

Proof. Starting from the right-hand side (RHS) of the equation:

$$
\begin{aligned}
\text{RHS} &= \sum_{t=1}^{|y_i|} \psi(d_{i,t}) \cdot A_{i,t} \\
&= \sum_{t=1}^{|y_i|} \left[1 + \left(d_{i,t} - \frac{1}{|y_i|} \sum_{k=1}^{|y_i|} d_{i,k} \right) \right] \cdot A_i \quad \text{(Substitute } \psi \text{ and } A_{i,t} = A_i) \\
&= A_i \cdot \sum_{t=1}^{|y_i|} \left(1 + d_{i,t} - \bar{d}_i \right) \quad \text{(Let } \bar{d}_i = \frac{1}{|y_i|} \sum_{k=1}^{|y_i|} d_{i,k}) \\
&= A_i \cdot \left(\sum_{t=1}^{|y_i|} 1 + \sum_{t=1}^{|y_i|} d_{i,t} - \sum_{t=1}^{|y_i|} \bar{d}_i \right) \quad \text{(Linearity of summation)} \\
&= A_i \cdot \left(|y_i| + |y_i| \cdot \bar{d}_i - |y_i| \cdot \bar{d}_i \right) \quad \text{(Note that } \sum d_{i,t} = |y_i| \bar{d}_i) \\
&= A_i \cdot |y_i| \\
&= \sum_{t=1}^{|y_i|} A_i \\
&= \sum_{t=1}^{|y_i|} A_{i,t} = \text{LHS}.
\end{aligned}
$$

□

C.3 Remark

The derivation above confirms that the term $\left(d_{i,t} - \bar{d}_i \right)$ acts as a zero-sum adjustment term. This guarantees that while tokens with higher auxiliary dependency ($d_{i,t} > \bar{d}_i$) receive amplified advantages, and those with lower dependency receive attenuated advantages, the total policy gradient magnitude for the sequence remains consistent with the original GRPO baseline.

D Ablation Study on the Mapping Function ψ

To determine the optimal strategy for converting the auxiliary dependency into advantage weights, we investigated two distinct mapping functions during the exploration phase on the **Qwen2.5-VL-3B-Instruct** model. Let $d_{i,t}$ denote the token-level auxiliary dependency and $\bar{d}_i$ be the sequence mean.

D.1 Candidate Functions

We compared a linear centering approach against a non-linear exponential approach:

- **Linear Mapping (ψ_1):** This function applies a centered linear adjustment:

$$
\psi_1(d_{i,t}) = 1 + \left(d_{i,t} - \frac{1}{|y_i|} \sum_{k=1}^{|y_i|} d_{i,k} \right). \tag{11}
$$

- **Exponential Mapping (ψ_2):** This function applies a Softmax-like non-linear scaling:

$$
\psi_2(d_{i,t}) = \exp \left(d_{i,t} - \frac{1}{|y_i|} \sum_{k=1}^{|y_i|} d_{i,k} \right). \tag{12}
$$

Table 5: **Comparison of different mapping functions on Qwen2.5-VL-3B-Instruct.** The Linear Mapping (ψ_1) outperforms the Exponential Mapping (ψ_2).

Function	Formulation	Complete Score
ψ_1 (Linear)	$1 + (d_{i,t} - \bar{d}_i)$	**7.15%**
ψ_2 (Exponential)	$\exp(d_{i,t} - \bar{d}_i)$	6.30%

D.2 Results and Analysis

The comparative results are presented in Table 5.

Analysis. The experimental evidence favors the Linear Mapping (ψ_1), which yields a performance gain of **+0.85%** compared to ψ_2.

From a theoretical perspective, this result aligns with the *Total Advantage Conservation* property (proved in Sec. C). The linear function ψ_1 ensures that the sum of advantages remains invariant after re-weighting ($\sum \psi_1 \cdot A = \sum A$), acting purely as an internal redistribution mechanism. Conversely, the exponential function ψ_2 distorts the total advantage magnitude of the sequence. This distortion likely introduces variance and instability into the gradient estimation, leading to suboptimal convergence.

Based on these findings, we selected ψ_1 as the standard mapping function for training the larger **Qwen3-VL-Thinking** model.

E Selection of Auxiliary Incentive Reward r_{aux}

To determine the optimal magnitude for the auxiliary incentive reward r_{aux}, we conducted a pilot study using the **Qwen2.5-VL-3B-Instruct** model. The coefficient r_{aux} plays a pivotal role in balancing the model's focus between primary reasoning and auxiliary structure utilization. We evaluated the model's Complete Score across a range of values from 0 to 0.5 on the overall MathCanvas-Bench. The results are summarized in Table 6.

Table 6: **Sensitivity analysis of r_{aux} on Qwen2.5-VL-3B-Instruct.** We swept the auxiliary reward weight from 0 to 0.5. The best performance is highlighted in **bold**.

r_{aux} **Value**	0	0.1	**0.2**	0.3	0.4	0.5
Complete Score	7.15%	7.36%	**7.50%**	7.15%	7.28%	6.93%

Analysis. As illustrated in Table 6, the performance exhibits an inverted U-shaped trend. Initially, increasing r_{aux} yields performance gains, peaking at **7.50%** when $r_{\text{aux}} = 0.2$. This indicates that a moderate auxiliary signal effectively encourages the model to leverage geometric clues. However, performance degrades as r_{aux} increases further (> 0.2), likely because an excessive weight on the auxiliary objective interferes with the optimization of the primary reasoning reward (r_{acc}).

Decision. Based on this pilot exploration, we selected $r_{\text{aux}} = 0.2$ as the optimal hyperparameter. We applied this setting to the training of the main **Qwen-3-VL-Thinking** model to support the learning of transferable auxiliary-construction strategies.

F Implementation Details

F.1 details on the LLM judge

We employ DeepSeek-V3.2 as our LLM judge (prompt in Appendix A.1). At the early stage of our study, to verify its reliability, we compared it with human evaluation and other judge models on 100 randomly sampled instances from Aux-PlaGeo, repeating five times, with results provided in Table 7. All models achieved over 98% consistency with human evaluation owing to reference answers. Considering efficiency and cost, we selected DeepSeek-V3.2 as the judge model.

Table 7: Comparison of Different Judges

Judge	1	2	3	4	5	Avg
DeepSeek-V3.2	99	98	99	100	98	98.8%
GLM-4.6	98	99	98	99	97	98.2%
GPT-5	99	99	100	98	99	**99.0%**

Table 8: Impact of training sample size

Size	Complete	Weighted
5k	24.1	34.7
10k	28.4	40.3
12k	29.9	42.4
14k	31.2	43.9
15k	**31.7**	**44.6**

F.2 Impact of training sample size

We experimented with various RL dataset sizes after one training epoch. As shown in Table 8, performance rises steadily with larger datasets but shows diminishing marginal gains: expanding from 14k to 15k only improves complete accuracy by 0.5%. The 15k setting achieves the best overall performance at acceptable cost, so we use it as our main configuration.

F.3 Hyperparameter Settings

We provide the detailed hyperparameters used during the training phase in Table 9. The experiments were conducted using a computational node equipped with 8 NVIDIA A100 GPUs.

We utilized the Hugging Face `trl` library to implement the training pipeline. For reproducibility, please note that **any hyperparameters not explicitly listed in Table 9 follow the default settings of the `trl` library**.

Table 9: **Hyperparameters used for training AuxCA.** The model is trained using full fine-tuning (without PEFT) on 8 $\times$ A100 GPUs.

Parameter	Value
Learning Rate (LR)	2e-6
LR Scheduler	Cosine Decay
Warmup Ratio	0.1
Use PEFT	False
Rollout	8
Per-Device Train Batch Size	2
Max Completion Length	4096
Max Prompt Length	1024
Gradient Accumulation Steps	4
Num Train Epochs	1
Computing Infrastructure	8 $\times$ A100

G Performance on problems that do not require auxiliary construction

We conducted an additional experiment using the manually extracted subset of geometry problems in MathVista testmini that require no auxiliary construction, and compared AuxCA with the base

Table 10: Performance on the MathVista

Model	GPS	GPS w/o aux
AuxCA	**71.6**	**73.7**
base model	68.7	72.8

Table 11: Failure Analysis

Error Type	%
Correct clue but wrong reasoning	**70.6**
Construct wrong auxiliary clue	12.5
Incomplete solution	8.3
Missing auxiliary construction	3.3
Other	5.3

model. As shown in Table 10, AuxCA outperforms the base model by only 0.9% on the w/o aux subset, verifying that AuxCA training preserves the model's original reasoning ability while bringing a minor gain. In contrast, it yields a 2.9% improvement on the overall set, demonstrating significant performance boosts on problems requiring auxiliary construction as well as strong zero-shot reasoning ability on other benchmarks.

H Analysis on failure cases

H.1 Failure Analysis

We have added a failure case analysis. Using five error categories defined in Table 11, we employed DeepSeek-V3.2 to classify incorrect model outputs based on the problem and ground truth. As shown in Table 11, 70.6% of errors stem from "correct clue but wrong reasoning", meaning AuxCA helps the model construct auxiliary elements effectively, yet its reasoning with such clues is still insufficient. We will include a detailed failure case study in the paper and investigate how to strengthen this reasoning ability in future work.

I Inference Cost Comparison with Explicit Plotting

We compare AuxCA-8B and Bagel-Canvas on all 635 `Aux-PlaGeo` problems under the same hardware and inference settings. As shown in Table 12, AuxCA generates 1,723 text tokens per problem on average, compared with 683 for Bagel-Canvas. Nevertheless, its average end-to-end latency is 17.38 seconds, compared with 50.24 seconds for Bagel-Canvas, representing a 65.4% reduction. Text token counts do not capture the work required to generate images, whereas end-to-end latency includes the complete inference process. Together with the accuracy gains reported in Table 1, these results show that AuxCA achieves higher accuracy and lower latency on this subset despite producing longer textual responses.

J Analysis of ACD Scores

To examine whether ACD identifies auxiliary-construction content, we randomly sample 200 `Aux-PlaGeo` test problems and compute ACD for every generated token.

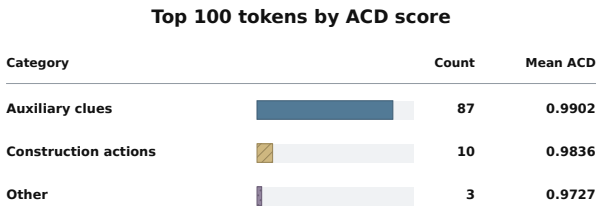


Figure 4: **Top 100 tokens by ACD.**

As shown in Figure 4, 87 of the 100 highest-scoring tokens are manually identified as auxiliary-clue tokens, such as "D" and "DE", with a mean ACD of 0.9902. Another 10 are construction-action tokens,

Table 12: Inference cost on 635 `Aux-PlaGeo` problems under the same hardware and inference settings. Token counts include text output only; latency covers the complete inference process.

Model	Avg. text tokens	Avg. latency (s)
AuxCA-8B	1,723	**17.38**
Bagel-Canvas	683	50.24

Table 13: **Distribution of token-level ACD scores.** Percentages are computed over all generated tokens from the 200 sampled `Aux-PlaGeo` problems.

ACD score	Percentage
$d_t > 0$	34.2%
$d_t = 0$	45.1%
$d_t < 0$	20.7%

including "connect", "extend", and "join", with a mean ACD of 0.9836; only 3 belong to other categories. The prominence of both geometric elements and construction actions suggests that ACD captures tokens relevant to auxiliary construction. We further examine low-scoring tokens and the overall score distribution in Appendix K. These analyses reveal a clear correspondence between ACD scores and auxiliary-construction content identified by human annotators, demonstrating the effectiveness of ACD.

K Further Analysis of ACD Scores

We further analyze the ACD scores of tokens generated on the 200 sampled `Aux-PlaGeo` problems. The 100 lowest-scoring tokens have a mean ACD of -0.4631 and are mainly function words, such as "the", "a", and "and". Unlike the auxiliary-clue and construction-action tokens concentrated at the high end of the ranking, these tokens generally do not express geometric construction steps.

Table 13 reports the distribution of scores over all generated tokens. Positive scores account for 34.2% of tokens, negative scores for 20.7%, and recorded zero scores for 45.1%. The zero-score group consists mainly of punctuation, numbers, and mathematical operators, for which no probability shift between the auxiliary-enhanced and original images was recorded. Under the centered weighting function $\psi(d_{i,t}) = 1 + (d_{i,t} - \bar{d}_i)$, tokens with higher ACD scores receive larger relative weights within a response; a zero ACD score does not itself imply zero training weight. Together with the manually annotated high-scoring tokens in Figure 4, this distribution supports the intended role of ACD in distinguishing auxiliary-construction content from more generic tokens.

L Limitations.

Despite the effectiveness of AuxCA in encouraging auxiliary clue construction, our failure analysis indicates that this ability does not always translate into correct reasoning. As shown in Table 11, 70.6% of the errors fall into the "correct clue but wrong reasoning" category, suggesting that the model can often construct useful auxiliary elements but still struggles to reliably exploit them in multi-step geometric reasoning. This highlights reasoning over constructed clues as a key bottleneck of the current framework. In addition, our error categorization is based on automatic classification by DeepSeek-V3.2, which may introduce annotation noise for ambiguous or partially correct solutions. Future work will investigate stronger reasoning mechanisms and more rigorous verification strategies to better leverage correctly constructed auxiliary clues.

M Social Impact.

The proposed method may have positive social impacts by improving AI-assisted education and visual reasoning. In particular, AuxCA can help models better identify and utilize auxiliary geometric structures, which may support intelligent tutoring systems, provide more interpretable problem-solving feedback, and help students understand complex geometric reasoning processes.

At the same time, the method also poses potential risks. More capable reasoning models may be misused for academic dishonesty if students rely on them to directly obtain solutions. Moreover, models may still generate plausible but incorrect reasoning, which could mislead users if not properly verified. Therefore, AuxCA-based systems should be deployed as assistive tools with appropriate safeguards, transparent reasoning outputs, and human oversight, especially in educational or high-stakes scenarios.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes], [No], or [N/A].
- [N/A] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for [N/A]).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will also be asked to include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While [Yes] is generally preferable to [No], it is perfectly acceptable to answer [No] provided a proper justification is given (e.g., error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering [No] or [N/A] is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Main claims are clearly stated in Abstract and Introduction, consistent with theoretical/experimental results and scope in the paper.

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: See Appendix H.

Guidelines:

- The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: All theorems include clear assumptions; complete proofs are provided in the main text and Appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Full reproducibility details are disclosed in main text and Appendix, enabling replication of core experimental results.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.

- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Code and data are publicly accessible with detailed reproduction instructions in the main text and Appendix.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).

- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Complete training/test details are provided in Section 4 and Appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: The reported evaluation results in this paper is averaged through multiple runs of experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Compute resources are detailed for all experiments in Appendix F.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.

- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The entire research follows the NeurIPS Code of Ethics, with anonymity and ethical norms fully respected.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Positive and negative societal impacts are analyzed in the Appendix H, with potential risks noted.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

Justification: This paper does not involve the release of models or datasets with high misuse risk.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All data used in this study are obtained from publicly available datasets. We have ensured that all licenses have been properly respected.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A]

Justification: The paper does not release new assets.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: This work does not involve crowdsourcing or research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: This work does not involve crowdsourcing or research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: This paper uses LLM as a reward model, with its evaluation and consistency with humans reported in the Appendix F.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.