MMD-Thinker: Adaptive Multi-Dimensional Thinking for Multimodal Misinformation Detection

Anonymous ACL submission

Abstract

Multimodal misinformation in both textual and visual forms, has flooded on social media, and has become a pressing issue in the era of AI-generated content (AIGC). Current studies in multimodal misinformation detection (MMD) typically utilize general-purpose Multimodal Large Language Models (MLLMs) to achieve impressive performance, but they encounter the limitation: Inflexible Thinking, where a single thinking mode (e.g., fast thinking or slow thinking) with a uniform reasoning style prevents the models from keeping pace with the fast-growing misinformation on the web. To address this, we propose MMD-Thinker, an adaptive multi-dimensional thinking framework for multimodal misinformation detection. First, we design three thinking modes with multi-dimensional reasoning including image, text, cross-modal, and attribution. Second, we employ instruction fine-tuning to clone all the thinking modes in misinformation analysis. Third, we introduce reinforcement learning with a mixed advantage function to build adaptive thinking capabilities in various scenarios. To support training, we conduct MMD-Instruct, a synthesis dataset comprising 8K image-text pairs, each annotated with reasoning and classification. Extensive experiments demonstrate that MMD-Thinker achieves state-of-the-art performance, while maintaining flexible inference and token usage.[1]

1 Introduction

The rapid development of AI-generated content (AIGC) technology in recent years has revolutionized the creation of multimodal content in the form of image-text pairs, while significantly lowering the barrier for producing multimodal misinformation (Wang et al., 2025a; Yan et al., 2025). As a result, a large amount of AI-generated multimodal misinformation has flooded on distinct social media platforms, such as Twitter and Facebook, resulting in unprecedented threats to societal stability and economy (Ma et al., 2024; Zhang et al., 2024). Although professional fact-checkers work on verifying the veracity of multimodal content, they fail to keep pace with the fast-growing and intricate multimodal misinformation. Therefore, developing automatic algorithms has become an urgent need to effectively detect multimodal misinformation in the era of AIGC.

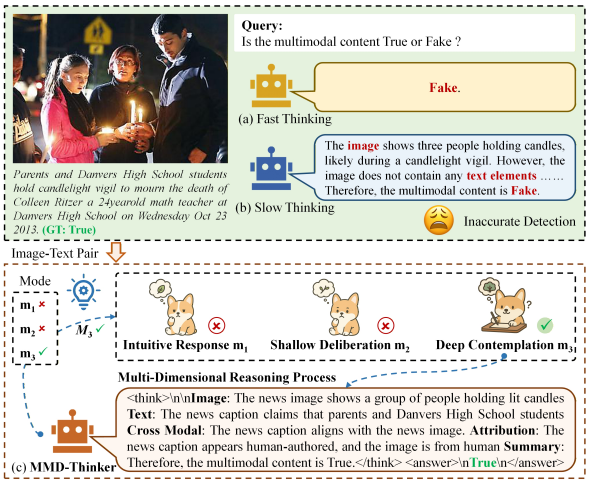

Figure 1: Illustrations of various MLLMs-assisted detection models in multimodal misinformation detection.

Meanwhile, Multimodal Large Language Models (MLLMs), such as Qwen2.5-VL (Bai et al., 2025), have made promising progress in different downstream tasks, such as visual question answering (Fang et al., 2025; Shah et al., 2025), mathematical reasoning (Liu et al., 2025a; Yu et al., 2025) and code generation (Guo et al., 2025b; Luo et al., 2025). The remarkable multimodal perception and reasoning capabilities of MLLMs make them an effective solution for addressing the task of multimodal misinformation detection (MMD).

Faced with this task, current MLLMs-assisted MMD studies (Shalabi et al., 2024; Cheng et al.,

[1] Code and data will be publicly available.

2024; Shao et al., 2024; Liu et al., 2025c) use task-specific instruction fine-tuning to mimic fast thinking for veracity classification, as illustrated in Figure 1(a). Such detection models provide quick decisions without deliberate thinking in a small token budget. However, this seriously depends on heuristics from vast data patterns (Zhang et al., 2025), yielding intuitive responses. In contrast, detection models (Qi et al., 2024; Yan et al., 2025) with slow thinking employ thorough reasoning processes for decision-making, as depicted in Figure 1(b). However, this focuses on the shallow semantic alignment between images and texts. In a nutshell, a single thinking mode with a uniform reasoning paradigm undermines the flexibility of detection, leading to suboptimal judgment. Therefore, it is significant to empower detection models with tailor-designed thinking modes that contain appropriate reasoning processes in coping with multimodal misinformation.

To address the aforementioned challenges, we propose MMD-Thinker, the first adaptive multi-dimensional thinking framework for the task of multimodal misinformation detection. Specifically, MMD-Thinker consists of three core components: (1) *Multi-Dimensional Thinking Mode Design* aligns with human cognitive behavior along both depth and breadth. In depth, it includes a spectrum from intuitive response and shallow deliberation to deep contemplation. In breadth, it covers image, text, cross-modal and attribution. (2) *Thinking Mode Learning* guides baseline models to learn all the thinking modes in misinformation analysis through task-specific instruction fine-tuning. (3) *Adaptive Thinking Mode Policy Optimization* initiates reinforcement learning (RL) with a mixed advantage function in both sample and mode for dynamic thinking. Furthermore, we perform a data synthesis pipeline that constructs a multimodal misinformation detection dataset, MMD-Instruct, with supervision in both reasoning and classification. MMD-Instruct comprises 7K automatically constructed training samples, and 1K manually verified test samples. Extensive experiments conducted on various MMD datasets demonstrate that MMD-Thinker achieves state-of-the-art (SOTA) results across multiple evaluation metrics, underscoring its effectiveness and generalization in MMD. Simultaneously, our framework makes the model flexible inference and token usage.

In summary, we focus on the models with adaptive multi-dimensional thinking in the MMD task, as shown in Figure 1(c). Therefore, we introduce MMD-Instruct, a dataset comprising both reasoning and classification annotations. Building upon this dataset, we propose MMD-Thinker, a framework that enhances misinformation analysis through instruction fine-tuning and reinforcement learning. Our main contributions are three-fold:

- We propose MMD-Thinker, the first adaptive multi-dimensional thinking framework for multimodal misinformation detection. MMD-Thinker includes thinking design, learning and optimization, and thus establishing both foundational reasoning and adaptive thinking capabilities in misinformation analysis.
- We introduce MMD-Instruct, the multimodal misinformation detection dataset of 8K samples with supervision in reasoning and classification. The dataset enables effective training for adaptive multi-dimensional thinking in detection models.
- Extensive experiments conducted on different benchmark datasets demonstrate that MMD-Thinker achieves state-of-the-art performance compared to MLLMs-assisted detection methods across various experimental settings, including zero-shot, instruction fine-tuning, and reinforcement learning.

2 Related Work

2.1 MLLMs for Multimodal Misinformation Detection

With the rapid advancement of general-purpose Multimodal Large Language Models (MLLMs) (Chen et al., 2023a; Li et al., 2023; Liu et al., 2023; Yao et al., 2024; Bai et al., 2025), the increasing attention of researchers has focused on leveraging the inherent capabilities of MLLMs across various domain applications (Zhang et al., 2023; Cheng et al., 2024; Lei et al., 2024).

Against the backdrop of the abuse of AIGC technologies, existing approaches in the realm of multimodal misinformation detection focus on AI-generated image-text pairs, and can be roughly categorized into two orientations according to parameter updating, i.e., zero-shot and instruction fine-tuning. In the setting of zero-shot, some recent work designs standard prompting with the direct-answer template to provide veracity classification without any reasoning chains (Wang

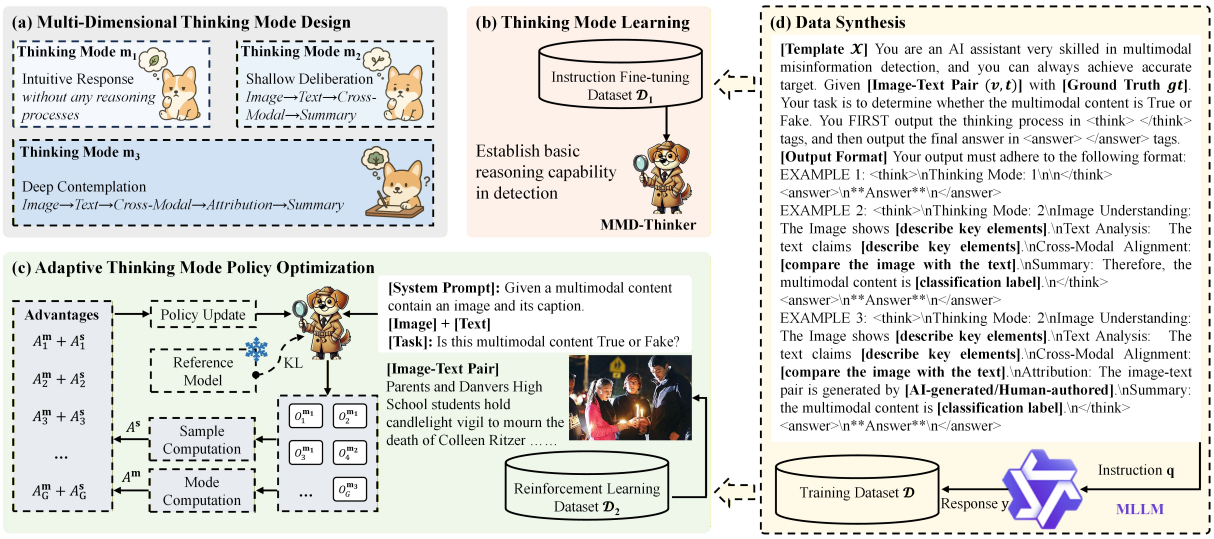

Figure 2: The MMD-Thinker framework consists of the following core components: (1) **Multi-Dimensional Thinking Mode Design** refers to three thinking modes with multi-dimensional reasoning processes, which aligns with human cognitive behavior in misinformation detection. (2) **Thinking Mode Learning** makes the model establish foundational reasoning capabilities in multimodal analysis. (3) **Adaptive Thinking Mode Policy Optimization** encourages the model for thinking mode switching. Additionally, (4) **Data Synthesis** is used to support training for adaptive multi-dimensional thinking in detection models.

et al., 2024b; Liu et al., 2025b). However, even advanced close-source models like GTP-4V (jia, 2025) still fail short in achieving satisfactory results, which is attributed to the fact that intuitive response is not always an effective way for any scenarios. On the other hand, researchers have begun to fine-tune general-purpose MLLMs by constructing instruction-following datasets with classification labels (Huang et al., 2025; Shalabi et al., 2024). Although these models bring significant improvements in detection accuracy, they struggle to effectively provide misinformation analysis for judgment. In contrast, our proposed MMD-Thinker traces forgery trajectories of misinformation, and thereby enhancing detection capabilities to AI-generated multimodal misinformation.

2.2 Reasoning in MLLMs

Remedying the absence of reasoning processes in classification tasks, some approaches are divided into two fronts. Early work adopts standard prompting with the uniform reasoning template to achieve both explanation and judgment (Xu et al., 2025b; Wen et al., 2025; Xu et al., 2025a). However, they often lack task-specific knowledge, resulting in inaccurate reasoning. Some recent studies extend general-purpose MLLMs to visual and multimodal tasks by constructing instruction-following datasets with both explanations and judgments (Qi et al., 2024; Guo et al., 2025a; Tong et al., 2025; Wu et al., 2025). However, they follow a single thinking mode with a predefined reasoning style for judgment, resulting in inflexible thinking across scenarios. In this work, we focus on adaptive thinking with multi-dimensional reasoning during detection.

3 Methodology

In this section, we introduce the MMD-Thinker framework, as illustrated in Figure 2. First, we provide the specifics of *Multi-Dimensional Thinking Mode Design*, which aligns with human cognitive behavior in multimodal misinformation detection. Second, we elaborate on *Thinking Mode Learning* and *Adaptive Thinking Mode Policy Optimization* for establishing foundational reasoning capabilities and thinking mode switching in misinformation analysis. Finally, we provide a data synthesis pipeline to support training the models with appropriate reasoning processes in the task of MMD.

3.1 Multi-Dimensional Thinking Mode Design

Aligning with human cognitive behavior in multimodal misinformation detection, we follow Cognitive Control Theory (Hoshi, 2006; Koechlin and Summerfield, 2007; Badre, 2008; Wang et al., 2025b; Zhu and Li, 2025) to design multi-dimensional thinking modes along both depth and breadth, as illustrated in Figure 2(a).

In depth, we observe that cognitive control operates a posterior-to-anterior hierarchy within the prefrontal cortex (PFC), spanning the posterior, mid-lateral, and anterior regions. Each region manages its respective goals and actions. The posterior region is associated with sensory control for rapid stimulus-response implementation. The mid-lateral region supports contextual control for multi-step coordination. The anterior region manages abstract goals and actions beyond both posterior and mid-lateral regions. Therefore, we devise three different thinking modes $\mathbf{M} = \{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$ corresponding to different regions of PFC. The $\mathbf{m}_1$ represents intuitive response for the posterior region. The $\mathbf{m}_2$ represents shallow deliberation for the mid-lateral region. The $\mathbf{m}_3$ represents deep contemplation for the anterior region.

In breadth, we investigate the characteristics of current MLLMs in debunking forgery trajectories of misinformation. A set of dimensions (i.e., image, text, cross-modal, and attribution) are crucial for effective detection. The image understanding refers to key visual entities and events. The text analysis also refers to key textual entities and events. The cross-modal alignment indicates the consistency between image and text at both entity-level and event-level. The attribution means the generation category of multimodal content, which includes AI-generated and Human-authored.

Specifically, the designed thinking modes with their reasoning processes are presented as follows:

- Intuitive Response $\mathbf{m}_1$: A basic thinking mode only contains the predicted classification label without any reasoning processes in a small token budget.
- Shallow Deliberation $\mathbf{m}_2$: A shallow thinking mode emphasizes the semantic understanding in pairs of image-text. The mode consists of a reasoning process and its judgment, where the reasoning process covers the perspectives of image, text and cross-modal in sequence.
- Deep Contemplation $\mathbf{m}_3$: A deep thinking mode based on shallow deliberation further introduces attribution into the reasoning process for identifying the veracity of image-text pairs. The mode effectively promotes deeper thinking to cope with multimodal misinformation stemming from AIGC.

Once multi-dimensional thinking modes for MMD have been designed, we leverage the large-scale MLLM Qwen2.5-VL-72B (Bai et al., 2025) to perform a data synthesis pipeline for supervised data. The details for constructing the dataset $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2\}$ can be found in Sec. 3.4. In $\mathcal{D}$, the dataset $\mathcal{D}_1$ is utilized for task-specific instruction fine-tuning, and the dataset $\mathcal{D}_2$ is leveraged for reinforcement learning.

3.2 Thinking Mode Learning

To learn all the thinking modes for establishing foundational reasoning capabilities in detection models, we employ instruction fine-tuning to incorporate task-specific knowledge into general-purpose MLLMs for end-to-end training, as illustrated in Figure 2(b).

Specifically, we extend the small-scale MLLM Qwen2.5-VL (Bai et al., 2025) to the task of multimodal misinformation detection. Based on the constructed instruction-following dataset $\mathcal{D}_1$, the loss function of this stage can be formulated as follows:

$$\mathcal{L}_1 = -\mathbb{E}_{(x,y)\sim\mathcal{D}_1}\left[\sum_{t=1}^{|y|} \log \pi_\theta(y_t \mid x, y_{<t})\right], \quad (1)$$

where y_t represents the t-th token for the given input x and the previous tokens $y_{<t}$. The symbol θ represents the updated parameters of MLLMs during training. After instruction fine-tuning, the model acquires both task-specific knowledge and diverse thinking modes, enabling it to establish the initial relationship between reasoning and classification in detection.

3.3 Adaptive Thinking Mode Policy Optimization

Although the model trained in the prior stage has basic reasoning capabilities, it still faces the dilemma related to the predefined thinking mode for each sample. This may result in a suboptimal path for the final judgment. To address this challenge, we introduce Reinforcement Learning (RL) to enable the model to dynamically adjust its thinking mode for different samples, as illustrated in Figure 2(c).

Specifically, we employ the Group-Relative Policy Optimization algorithm (GRPO) (Shao et al., 2024) to refine the model for potential self-reflection. Given the constructed dataset $\mathcal{D}_2$, the

objective of GRPO can be defined as follows:

$$\mathcal{L}_2 = -\mathbb{E}_{(x,y)\sim\mathcal{D}_2}\Bigg[\frac{1}{G}\sum_{i=1}^{G}\Big(\min\Big(\frac{\pi_\theta(y_i)}{\pi_{\text{old}}(y_i)}A_i, \text{clip}\Big(\frac{\pi_\theta(y_i)}{\pi_{\text{old}}(y_i)}, 1-\epsilon,\, 1+\epsilon\Big)A_i\Big) -\beta\,\mathcal{D}_{\text{KL}}(\pi_\theta \,\|\, \pi_{\text{ref}})\Big)\Bigg], \tag{2}$$

where π_θ, π_{old}, and π_{ref} denote the current policy model, the sampling policy model, and the fixed reference model, respectively. G denotes the total number of responses. ϵ denotes the hyperparameter that controls the clipping range. β denotes the hyperparameter for controlling the KL penalty strength. $\mathcal{D}_{\text{KL}}(\pi_\theta \,\|\, \pi_{\text{ref}})$ denotes the KL divergence constraint. Here, A represents the mixed advantage for each response y in both sample and mode:

$$A_i = A_i^{\mathbf{S}} + A_i^{\mathbf{M}}, \tag{3}$$

where $A^{\mathbf{S}}$ and $A^{\mathbf{M}}$ represent the normalized advantages in sample and mode, respectively. The $A^{\mathbf{S}}$ is calculated as:

$$A_i^{\mathbf{S}} = \frac{R_i^{\mathbf{s}} - \text{mean}\left(\{R_1^{\mathbf{s}}, \ldots, R_G^{\mathbf{s}}\}\right)}{\text{std}\left(\{R_1^{\mathbf{s}}, \ldots, R_G^{\mathbf{s}}\}\right)}, \tag{4}$$

where mean and std denote the arithmetic mean and the standard deviation, respectively. $R_i^{\mathbf{s}}$ represents the reward value of the i-th response from the policy model. It integrates detection accuracy reward R_{acc} with reasoning format reward R_{format}. The former ensures the consistency between the ground truth and the predicted classification label. The latter ensures the model to follow the standard structure of the reasoning process. The reasoning process is enclosed within the "<think> </think>" tokens, and the classification label is contained within "<answer> </answer>" tokens. The $A^{\mathbf{M}}$ is computed as:

$$A_i^{\mathbf{M}} = \frac{R_i^{\mathbf{m}} - \text{mean}\left(\{R^{\mathbf{m}_1}, \ldots, R^{\mathbf{m}_n}\}\right)}{\text{std}\left(\{R^{\mathbf{m}_1}, \ldots, R^{\mathbf{m}_n}\}\right)}, \tag{5}$$

where $n \in \{1, 2, 3\}$ denotes the number of thinking mode. $R^{\mathbf{m}}$ denotes the average reward within a certain thinking mode. The mixed advantage function A guides the model to adaptively select a thinking mode with its reasoning process for judgment.

In summary, compared to the instruction fine-tuning stage which strictly regulates the output of the model to establish foundational reasoning capabilities, the GRPO stage emphasizes adaptive thinking capabilities across various scenarios.

3.4 Data Synthesis

Considering the scarcity of reasoning processes in multimodal misinformation detection datasets created by AIGC, we develop a data synthesis pipeline to construct a MMD dataset, namely MMD-Instruct, which provides supervision for both reasoning and classification in the task of MMD, as shown in Figure 2(d). The data synthesis pipeline consists of the following two parts: Data Generation and Data Quality Control.

Data Generation First, we follow three designed multi-dimensional thinking modes with their respective reasoning styles to construct an instruction $\mathbf{q}$ for multimodal misinformation detection with the designed instruction template:

$$\mathbf{q} = \text{template}(\mathcal{X}, gt), \tag{6}$$

where gt denotes the ground truth corresponding to the veracity of an image-text pair, enabling MLLMs reasoning correctly. $\mathcal{X}$ denotes the basic concept of multiple dimensions that guide MLLMs to perform goals and actions in a certain thinking mode. Second, we utilize the large-scale MLLM Qwen2.5-VL-72B (Bai et al., 2025) to generate high-quality multimodal misinformation data with both reasoning and classification. The process can be formulated as follows:

$$y = \text{MLLM}(v, t, \mathbf{q}), \tag{7}$$

where (v, t) denotes the image-text pair. Here, the image-text pairs are sourced from the MM-FakeBench (Liu et al., 2025b) dataset, which contains a large volume of texts and images generated by AIGC. This is beneficial for current detection models to advance the realm of MMD in the era of AIGC. Furthermore, we introduce an operation of data quality control into the data synthesis pipeline for high-quality samples.

Data Quality Control To eliminate the noise (e.g., erroneous reasoning logic) related to MLLMs outputs in Data Generation, we introduce a rigorous and robust two-step filtering protocol, i.e., automatic filtering followed by manual verification. **(1) Automatic Filtering.** The generated samples are automatically validated against four key criteria: classification labels, thinking modes, multiple

dimensions, and reasoning format:

$$\hat{y} = \mathbb{I}\left[M_{cls}(y) \wedge M_{mode}(y) \wedge M_{dimension}(y) \wedge M_{format}(y)\right], \quad (8)$$

where $M_{cls}(y)$ represents the predicted binary classification such as "True" and "Fake", verifying whether it aligns with the ground truth. $M_{mode}(y)$ and $M_{dimension}(y)$ represent a certain thinking mode and its multi-dimensional reasoning process. $M_{format}(y)$ assesses whether the output follows the required structure. $\mathbb{I}[\cdot]$ denotes the indicator function. If all the conditions are satisfied, it returns 1. Otherwise, it returns 0. For a given input consisting of an image-text pair (v, t) and instruction $\mathbf{q}$, we use a mode-controlled strategy to generate a valid response y through thinking modes $\mathbf{m}_1$, $\mathbf{m}_2$, and $\mathbf{m}_3$, until the maximum retry limit is reached. **(2) Manual Verification.** After automatic filtering, we achieve the MMD-Instruct dataset $\mathcal{D}$ that is divided into a training set ($\sim$7K samples) and a test set ($\sim$1K samples). For the training set, we further split $\mathcal{D}$ into two parts: $\mathcal{D}_1$ and $\mathcal{D}_2$. The former contains $\sim$6K samples, and the latter includes $\sim$1K samples. Additionally, for the test set, we employ expert annotators to conduct a thorough manual evaluation, and thereby ensuring the logical coherence and consistency of reasoning processes.

4 Experiment

In this section, we perform experiments on different benchmark datasets to illustrate the effectiveness and generalization of MMD-Thinker, and further provide ablation study for evaluating the contribution of each core component in MMD-Thinker.

4.1 Experimental Setups

Datasets. To demonstrate the efficiency of MMD-Thinker, we evaluate it on the constructed MMD-Instruct, Twitter (Boididou et al., 2015), and PHEME (Zubiaga et al., 2017), as presented in Table 1. The MMD-Instruct features diverse thinking modes with reasoning processes, which makes it a more rigorous benchmark for measuring the detection capabilities of MLLMs. Twitter is a large-scale dataset with 17K unique tweets spanning two classification labels. PHEME encompasses 2K tweets, primarily includes five significant breaking news events.

Evaluation Metrics. To evaluate detection performance, we follow existing evaluation protocols (Qian et al., 2021; Chen et al., 2023b; Liu et al., 2025b) to adopt the four primary metrics: Accuracy (Acc.), Precision (Pre.), Recall (Rec.), and macro-F1 Score. In these evaluation metrics, the F1 score balances the precision and recall through the harmonic mean, where higher values represent better performance. Additionally, we report average tokens per response (Tokens) as an auxiliary evaluation metric for measuring the model's flexibility.

Dataset	Training	Test	Total
Twitter	15,000	2,000	17,000
PHEME	1,414	608	2,022
MMD-Instruct (Ours)	7,174	1,000	8,174

Table 1: Data distribution of Twitter, PHEME and MMD-Instruct.

Baselines. We compare MMD-Thinker against both general-purpose MLLMs and task-specific detection models. General-purpose MLLMs include GPT-5-Mini (OpenAI, 2025), Gemma 3-12B (Team et al., 2025), Llama-3.2-11B-Vision (Grattafiori et al., 2024), Qwen2-VL-7B (Wang et al., 2024a), and Qwen2.5-VL with two parameters: 3B and 7B (Bai et al., 2025). Task-specific detection models include SNIFFER (Qi et al., 2024) and TRUST-VL (Yan et al., 2025). To make a fair comparison, all methods consider image-text pairs as input for detecting multimodal misinformation. The specifics of baselines are provided in the appendix.

Implementation Details. All the experiments are performed using PyTorch (Paszke et al., 2019) on 2×NVIDIA A800 GPUs (80G), with Qwen2.5-VL-3B and Qwen2.5-VL-7B as baseline models (Bai et al., 2025). In Thinking Mode Learning, we employ full-parameter fine-tuning to train for 3 epochs on 6K image-text pairs with a batch size of 8 and a learning rate of 2×10^{-5}. In Adaptive Thinking Mode Policy Optimization, we train for 8 epochs on 1K image-text pairs with 12 candidate responses per group.

4.2 Overall Evaluation

Table 2 provides overall evaluation across different benchmark datasets. From the results, we can achieve the following findings:

- Our proposed MMD-Thinker with Qwen2.5-VL-7B significantly outperforms other com-

Method	In-Domain (MMD-Instruct)				Out-of-Domain				Tokens	Avg. F1 ↑
					PHEME		Twitter			
	Acc. ↑	F1 ↑	Pre. ↑	Rec. ↑	Acc. ↑	F1 ↑	Acc. ↑	F1 ↑		
GPT-5-Mini	80.48	73.87	75.32	72.84	44.67	44.07	57.05	54.48	99	57.47
Gemma 3-12B	68.34	67.54	72.23	77.63	34.88	33.30	61.93	61.35	116	54.06
Llama-3.2-11B-Vision	63.95	63.39	66.27	68.49	42.31	41.76	57.72	57.65	166	54.27
Qwen2-VL-7B	73.80	71.18	70.97	76.20	41.34	40.57	61.41	61.41	87	57.72
Qwen2.5-VL-7B	82.10	78.46	77.34	80.20	43.76	39.74	56.06	55.06	124	57.75
Qwen2.5-VL-3B	75.30	69.47	68.98	70.12	51.49	44.27	57.67	56.34	86	56.69
SNIFFER	55.50	55.14	63.21	65.46	34.81	33.69	50.22	47.23	72	45.35
TRUST-VL	83.00	80.72	79.27	85.43	33.33	33.15	64.18	62.07	141	58.64
MMD-Thinker										
Qwen2.5-VL-3B	90.70	88.40	87.65	89.27	52.31	46.80	62.84	59.12	192	64.77
Qwen2.5-VL-7B	**92.90**	**90.74**	**91.78**	**89.83**	**58.25**	**50.86**	**68.60**	**62.53**	**175**	**68.04**

Table 2: Performance (%) comparison between MMD-Thinke and other MLLMs-assisted detection methods across various benchmark datasets. Best results are in bold.

pared methods on three datasets across all evaluation metrics. In the term of token usage, while our method shows a slight increase in inference, it strengthens multimodal analysis for advanced detection capabilities. This is because deep contemplation guides the model to trace forgery trajectories of multimodal misinformation from multiple dimensions. This strong performance highlights the efficacy of MMD-Thinker and its potential to advance reasoning capabilities in detection.

- General-purpose MLLMs, even the close-source GPT-5-Mini (OpenAI, 2025), exhibit unsatisfactory performance. Specifically, the GPT-5-Mini achieves an average F1 metric of 57.47%, which performs worse than MMD-Thinker. This is attributed to the lack of task-specific knowledge. This suggests that the task-specific instruction fine-tuning can benefit the models for misinformation detection.

- Task-specific detection models, such as SNIFFER and TRUST-VL, show the opposite results. SNIFFER exhibits performance drops on these benchmark datasets, indicating poor generalization. In the meantime, SNIFFER employs a simple thinking mode with entity analysis that struggles to cope with complex scenarios, especially for AI-generated multimodal misinformation. TRUST-VL provides performance improvements compared to general-purpose MLLMs, but its performance seriously relies on web-retrieved external knowledge. This significantly increases the burden of model deployment and dataset storage, especially in low-resource conditions.

Method	MMD-Instruct	PHEME	Twitter	Tokens
Qwen2.5-VL-3B	69.47	44.27	56.34	86
w/ TML (6K)	85.15	44.63	54.54	220
w/ TML (7K)	86.07	44.08	56.63	229
w/ TML+GRPO (6K+1K)	87.11	45.57	57.83	212
w/ TML+ATMPO (6K+1K)	88.40	46.80	59.12	192
Qwen2.5-VL-7B	78.46	39.74	55.06	124
w/ TML (6K)	87.28	45.89	61.70	223
w/ TML (7K)	87.90	44.01	59.15	225
w/ TML+GRPO (6K+1K)	89.89	47.88	59.45	210
w/ TML+ATMPO (6K+1K)	90.74	50.86	62.53	175

Table 3: Ablation study of TML and ATMPO in MMD-Thinker across various benchmark datasets, evaluated using the F1 metric. Data sizes used for TML and ATMPO are annotated in blue and red, respectively. TML denotes Thinking Mode Leaning. ATMPO denotes Adaptive Thinking Mode Policy Optimization.

4.3 Ablation Study

To evaluate the contribution of each component in the proposed MMD-Thinker, we provide ablation experiments in Table 3 and Table 4.

Effect of TML. First, we conduct the variant of MMD-Thinker, where the model is trained only by Thinking Mode Learning (TML), denoted by *w/* TML. As shown in 1-th row and 2-th row, 6-th row and 7-th row of Table 3, the results exhibit that the variant achieves substantial improvements while increasing the cost of token usage in inference. The results demonstrate that TML injects task-specific knowledge into the model to perform appropriate reasoning processes for more accurate

Method	MMD-Instruct	PHEME	Twitter	Tokens
MMD-Thinker	88.40	46.80	59.12	192
Effect of Each Thinking Mode				
Intuitive Response $\mathbf{m}_1$	81.67	45.47	42.47	22
Shallow Deliberation $\mathbf{m}_2$	83.95	44.75	55.97	215
Deep Contemplation $\mathbf{m}_3$	85.63	45.70	54.02	233

Table 4: Ablation study of MDTM in MMD-Thinker across various benchmark datasets, evaluated using the F1 metric. MDTM denotes Multi-Dimensional Thinking Mode.

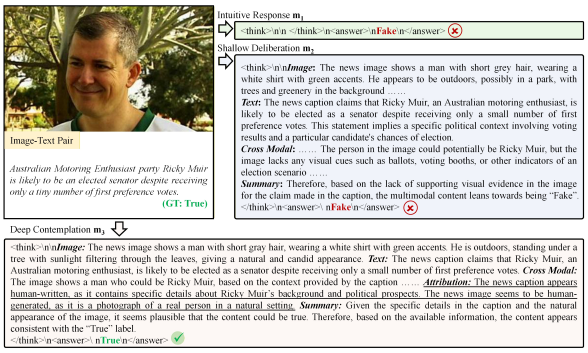

Figure 3: Example multimodal misinformation detection spanning Intuitive Response $\mathbf{m}_1$, Shallow Deliberation $\mathbf{m}_2$, and Deep Contemplation $\mathbf{m}_3$ in Adaptive Thinking Mode Policy Optimization of MMD-Thinker.

detection. On the other hand, we also use samples of different size to perform TML for comparison (see 2-th row and 3-th row, 7-th row and 8-th row in Table 3). Increasing the number of samples used during TML yields marginal performance gains while bringing additional computational resource usage. The results highlight that TML, together with instruction-following supervision, ensures the model to establish basic reasoning capabilities.

Effect of ATMPO. Second, the comparison between Vanilla GRPO and *w/o* GRPO (see 2-th row and 4-th row, 7-th row and 9-th row in Table 3) can observe that the former performs better than the latter. It surpasses the variant (*w/o* GRPO) by approximate 2% of average F1 score across three datasets. This verifies that reinforcement learning is conducive to refine reasoning processes for judgment. Furthermore, we provide the comparison between Adaptive Thinking Mode Policy Optimization (ATMPO) and Vanilla GRPO (see 4-th row and 5-th row, 9-th row and 10-th row in Table 3). The performance of the former outperforms that of the latter by a large margin, while reducing the cost of token usage in inference. This underscores the significance of ATMPO in MMD-Thinker. More discussions in terms of dynamic thinking mode are provided in Appendices A.3.

Effect of MDTM. Finally, we provide the performance of MMD-Thinker with each thinking mode, as shown in Table 4. In Intuitive Response $\mathbf{m}_1$, this mode uses the least tokens and achieves the lowest average F1 score. This quick decision without explicit reasoning limits the model in its capability to distinguish between true and fake content. In both Shallow Deliberation $\mathbf{m}_2$ and Deep Contemplation $\mathbf{m}_3$, the models achieve significant performance improvements compared to Intuitive Response $\mathbf{m}_1$. This verifies the effectiveness of different thinking modes in detection. Together with different thinking modes that contain multi-dimensional reasoning (MDTM), MMD-Thinker achieves the highest results on various benchmark datasets.

4.4 Case Study

Figure 3 depicts the case that different thinking modes with their respective reasoning processes in Adaptive Thinking Mode Policy Optimization. In the process of training, MMD-Thinker could conduct different reasoning chains for reaching possible classification labels. For example, in Intuitive Response $\mathbf{m}_1$, the model attempts to provide an inaccurate classification label without any reasoning processes. In Shallow Deliberation $\mathbf{m}_2$, the model emphasizes the semantic analysis between image and text, which struggles to cope with the complex scenario. In Deep Contemplation $\mathbf{m}_3$, the model follows a deep reasoning process for the accurate predicted label. The cases demonstrate the necessity for thinking mode switching of MMD-Thinker across different scenarios.

5 Conclusion

In this paper, we introduce MMD-Thinker, the first adaptive multi-dimensional thinking framework for multimodal misinformation detection. While MLLMs-assisted detection methods face inflexible thinking in different scenarios, our method effectively integrates different thinking modes with multi-dimensional reasoning processes into general-purpose MLLMs via task-specific instruction fine-tuning and reinforcement learning. Experiments on different benchmark datasets demonstrate that MMD-Thinker outperforms state-of-the-art methods with significant improvements, highlighting the effectiveness of MMD-Thinker in multimodal misinformation detection.

Limitations

Despite the effectiveness of the proposed MMD-Thinker, these are some limitations. First, the thinking mode with its reasoning process is directly generated by the designed task queries, rather than iteratively being learned by the model. Leveraging the trained model expands the training dataset to enable scalable self-improvement of the model. Second, while our designed three thinking modes with their respective reasoning processes are effective, more thinking modes could be considered for more complex scenarios. Finally, the model's practical applicability in real-world deployment could be for future exploration.

References

2025. Large visual-language models are also good classifiers: A study of in-context multimodal fake news detection. *arXiv preprint arXiv:2407.12879*.

David Badre. 2008. Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, 12(5):193–200.

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, and 8 others. 2025. Qwen2.5-vl technical report. *arXiv preprint arXiv:2502.13923*.

Christina Boididou, Katerina Andreadou, Symeon Papadopoulos, Duc-Tien Dang-Nguyen, Giulia Boato, Michael Riegler, , and Yiannis Kompatsiaris. 2015. Verifying multimedia use at mediaeval 2015.

Jun Chen, Deyao Zhu, Xiaoqian Shen, Xiang Li, Zechu Liu, Pengchuan Zhang, Raghuraman Krishnamoorthi, Vikas Chandra, Yunyang Xiong, and Mohamed Elhoseiny. 2023a. Minigpt-v2: large language model as a unified interface for vision-language multi-task learning. *arXiv preprint arXiv:2310.09478*.

Ziwei Chen, Linmei Hu, Weixin Li, Yingxia Shao, and Liqiang Nie. 2023b. Causal intervention and counterfactual reasoning for multi-modal fake news detection. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 627–638, Toronto, Canada. Association for Computational Linguistics.

Zebang Cheng, Zhi-Qi Cheng, Jun-Yan He, Jingdong Sun, Kai Wang, Yuxiang Lin, Zheng Lian, Xiaojiang Peng, and Alexander G. Hauptmann. 2024. Emotion-llama: Multimodal emotion recognition and reasoning with instruction tuning. In *Advances in Neural Information Processing Systems*, volume 37, pages 110805–110853. Curran Associates, Inc.

Wenlong Fang, Qiaofeng Wu, Jing Chen, and Yun Xue. 2025. Notes-guided mllm reasoning: Enhancing mllm with knowledge and visual notes for visual question answering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19597–19607.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, and 542 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Jiawei Guo, Tianyu Zheng, Yizhi Li, Yuelin Bai, Bo Li, Yubo Wang, King Zhu, Graham Neubig, Wenhu Chen, and Xiang Yue. 2025a. MAmmoTH-VL: Eliciting multimodal reasoning with instruction tuning at scale. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13869–13920, Vienna, Austria. Association for Computational Linguistics.

Yaoqi Guo, Zhenpeng Chen, Jie M. Zhang, Yang Liu, and Yun Ma. 2025b. Personality-guided code generation using large language models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1068–1080, Vienna, Austria. Association for Computational Linguistics.

Eiji Hoshi. 2006. Functional specialization within the dorsolateral prefrontal cortex: A review of anatomical and physiological studies of non-human primates. *Neuroscience Research*, 54(2):73–84.

Dawei Huang, Qing Li, Chuan Yan, Zebang Cheng, Zihao Han, Yurong Huang, Xiang Li, Bin Li, Xiaohui Wang, Zheng Lian, Zhi-Qi Cheng, and Xiaojiang Peng. 2025. Emotion-qwen: A unified framework for emotion and vision understanding. *arXiv preprint arXiv:2505.06685*.

Etienne Koechlin and Christopher Summerfield. 2007. An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, 11(6):229–235.

Shanglin Lei, Guanting Dong, Xiaoping Wang, Keheng Wang, Runqi Qiao, and Sirui Wang. 2024. Instructerc: Reforming emotion recognition in conversation with multi-task retrieval-augmented large language models. *arXiv preprint arXiv:2309.11911*.

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

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual instruction tuning. In *Advances in*

Neural Information Processing Systems, volume 36, pages 34892–34916. Curran Associates, Inc.

Xiangyan Liu, Jinjie Ni, Zijian Wu, Chao Du, Longxu Dou, Haonan Wang, Tianyu Pang, and Michael Qizhe Shieh. 2025a. Noisyrollout: Reinforcing visual reasoning with data augmentation. *arXiv preprint arXiv:2504.13055*.

Xuannan Liu, Zekun Li, Peipei Li, Huaibo Huang, Shuhan Xia, Xing Cui, Linzhi Huang, Weihong Deng, and Zhaofeng He. 2025b. Mmfakebench: A mixed-source multimodal misinformation detection benchmark for lvlms. In *ICLR*.

Zhiwei Liu, Xin Zhang, Kailai Yang, Qianqian Xie, Jimin Huang, and Sophia Ananiadou. 2025c. Fmdllama: Financial misinformation detection based on large language models. In *Companion Proceedings of the ACM on Web Conference 2025*, WWW '25, page 1153–1157, New York, NY, USA. Association for Computing Machinery.

Ziyang Luo, Kaixin Li, Hongzhan Lin, Yuchen Tian, Mohan Kankanhalli, and Jing Ma. 2025. Tree-of-evolution: Tree-structured instruction evolution for code generation in large language models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 297–316, Vienna, Austria. Association for Computational Linguistics.

Zihan Ma, Minnan Luo, Hao Guo, Zhi Zeng, Yiran Hao, and Xiang Zhao. 2024. Event-radar: Event-driven multi-view learning for multimodal fake news detection. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5809–5821, Bangkok, Thailand. Association for Computational Linguistics.

OpenAI. 2025. Gpt-5 system card.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, and 2 others. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Peng Qi, Zehong Yan, Wynne Hsu, and Mong Li Lee. 2024. Sniffer: Multimodal large language model for explainable out-of-context misinformation detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13052–13062.

Chen Qian, Fuli Feng, Lijie Wen, Chunping Ma, and Pengjun Xie. 2021. Counterfactual inference for text classification debiasing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5434–5445, Online. Association for Computational Linguistics.

Monika Shah, Sudarshan Balaji, Somdeb Sarkhel, Sanorita Dey, and Deepak Venugopal. 2025. Analyzing the sensitivity of vision language models in visual question answering. In *Proceedings of the Fourth Workshop on Generation, Evaluation and Metrics (GEM²)*, pages 431–438, Vienna, Austria and virtual meeting. Association for Computational Linguistics.

Fatma Shalabi, Hichem Felouat, Huy H. Nguyen, and Isao Echizen. 2024. Leveraging chat-based large vision language models for multimodal out-of-context detection. In *Advanced Information Networking and Applications*, pages 86–98, Cham. Springer Nature Switzerland.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, Louis Rouillard, Thomas Mesnard, Geoffrey Cideron, Jean bastien Grill, Sabela Ramos, Edouard Yvinec, Michelle Casbon, Etienne Pot, Ivo Penchev, and 197 others. 2025. Gemma 3 technical report. *arXiv preprint arXiv:2503.19786*.

Shengbang Tong, David Fan, Jiachen Li, Yunyang Xiong, Xinlei Chen, Koustuv Sinha, Michael Rabbat, Yann LeCun, Saining Xie, and Zhuang Liu. 2025. Metamorph: Multimodal understanding and generation via instruction tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 17001–17012.

Jin Wang, Chenghui Lv, Xian Li, Shichao Dong, Huadong Li, Kelu Yao, Chao Li, Wenqi Shao, and Ping Luo. 2025a. Forensics-bench: A comprehensive forgery detection benchmark suite for large vision language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4233–4245.

Minzheng Wang, Yongbin Li, Haobo Wang, Xinghua Zhang, Nan Xu, Bingli Wu, Fei Huang, Haiyang Yu, and Wenji Mao. 2025b. Adaptive thinking via mode policy optimization for social language agents. *arXiv preprint arXiv:2505.02156*.

Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Yang Fan, Kai Dang, Mengfei Du, Xuancheng Ren, Rui Men, Dayiheng Liu, Chang Zhou, Jingren Zhou, and Junyang Lin. 2024a. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *arXiv preprint arXiv:2409.12191*.

Shengkang Wang, Hongzhan Lin, Ziyang Luo, Zhen Ye, Guang Chen, and Jing Ma. 2024b. Mfc-bench: Benchmarking multimodal fact-checking with large vision-language models. *arXiv preprint arXiv:2406.11288*.

Congcong Wen, Yisiyuan Huang, Hao Huang, Yanjia Huang, Shuaihang Yuan, Yu Hao, Hui Lin, Yu-Shen Liu, and Yi Fang. 2025. Zero-shot object navigation with vision-language models reasoning. In *Pattern Recognition*, pages 389–404, Cham. Springer Nature Switzerland.

Jiulong Wu, Yucheng Shen, Lingyong Yan, Haixin Sun, Deguo Xia, Jizhou Huang, and Min Cao. 2025. Facial-r1: Aligning reasoning and recognition for facial emotion analysis. *arXiv preprint arXiv:2511.10254*.

Jiacong Xu, Shao-Yuan Lo, Bardia Safaei, Vishal M. Patel, and Isht Dwivedi. 2025a. Towards zero-shot anomaly detection and reasoning with multimodal large language models. pages 20370–20382.

Yifang Xu, Yunzhuo Sun, Benxiang Zhai, Ming Li, Wenxin Liang, Yang Li, and Sidan Du. 2025b. Zero-shot video moment retrieval via off-the-shelf multimodal large language models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(9):8978–8986.

Zehong Yan, Peng Qi, Wynne Hsu, and Mong-Li Lee. 2025. TRUST-VL: An explainable news assistant for general multimodal misinformation detection. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 5588–5604, Suzhou, China. Association for Computational Linguistics.

Yuan Yao, Tianyu Yu, Ao Zhang, Chongyi Wang, Junbo Cui, Hongji Zhu, Tianchi Cai, Haoyu Li, Weilin Zhao, Zhihui He, Qianyu Chen, Huarong Zhou, Zhensheng Zou, Haoye Zhang, Shengding Hu, Zhi Zheng, Jie Zhou, Jie Cai, Xu Han, and 4 others. 2024. Minicpm-v: A gpt-4v level mllm on your phone. *arXiv preprint arXiv:2408.01800*.

Erxin Yu, Jing Li, Ming Liao, Qi Zhu, Boyang Xue, Minghui Xu, Baojun Wang, Lanqing Hong, Fei Mi, and Lifeng Shang. 2025. Self-error-instruct: Generalizing from errors for llms mathematical reasoning. *arXiv preprint arXiv:2505.22591*.

Hang Zhang, Xin Li, and Lidong Bing. 2023. Video-LLaMA: An instruction-tuned audio-visual language model for video understanding. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 543–553, Singapore. Association for Computational Linguistics.

Litian Zhang, Xiaoming Zhang, Ziyi Zhou, Feiran Huang, and Chaozhuo Li. 2024. Reinforced adaptive knowledge learning for multimodal fake news detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(15):16777–16785.

Shengjia Zhang, Junjie Wu, Jiawei Chen, Changwang Zhang, Xingyu Lou, Wangchunshu Zhou, Sheng Zhou, Can Wang, and Jun Wang. 2025. Othink-r1: Intrinsic fast/slow thinking mode switching for over-reasoning mitigation. *arXiv preprint arXiv:2506.02397*.

Jason Zhu and Hongyu Li. 2025. Towards concise and adaptive thinking in large reasoning models: A survey. *arXiv preprint arXiv:2507.09662*.

Arkaitz Zubiaga, Maria Liakata, and Rob Procter. 2017. Exploiting context for rumour detection in social media. In *Social Informatics*, pages 109–123, Cham. Springer International Publishing.

A Appendix

A.1 Baseline Details

- **GPT-5-Mini** (OpenAI, 2025) is one of the most powerful multimodal large language models. We leverage GPT-5-Mini with reasoning templates in a zero-shot manner for multimodal misinformation detection.

- **Gemma 3-12B** (Team et al., 2025) is an advanced multimodal large language model while introducing key advances including post-training alignment for instruction following and safety, and architecture refinement for powerful performance.

- **Llama-3.2-11B-Vision** (Grattafiori et al., 2024) is a cutting-edge multimodal language model that extends the capabilities of Llama. It introduces two key innovations: Modular Vision for efficient multimodal reasoning, and Preference Alignment for robustness and faithfulness.

- **Qwen2-VL-7B** (Wang et al., 2024a) is a multimodal large language model that has visual and language understanding capabilities. It consists of two key innovations: Naive Dynamic Resolution, dynamically processing images of varying resolutions, and Multimodal Rotary Position Embedding, which fuses information across modals.

- **Qwen2.5-VL-3B** and **Qwen2.5-VL-7B** (Bai et al., 2025) are advanced multimodal large language models that build upon the foundation of Qwen2-VL. They further introduce two key innovations: Window Attention, optimizing inference efficiency, and Dynamic Frames Per Second (FPS) Sampling, which extends dynamic resolution.

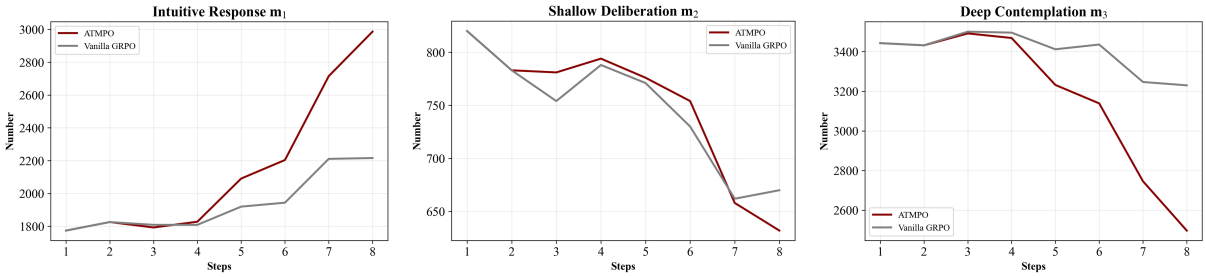

Figure 4: Comparison of vanilla GRPO and ATMPO on Qwen2.5-VL-3B (Bai et al., 2025) in terms of dynamic thinking mode. ATMPO denotes Adaptive Thinking Mode Policy Optimization.

Thinking Mode	Total	Training	Test
Intuitive Response $\mathbf{m}_1$	1349	1199	150
Shallow Deliberation $\mathbf{m}_2$	1267	1167	100
Deep Contemplation $\mathbf{m}_3$	5558	4808	750

Table 5: Statistics of MMD-Instruct regarding thinking mode.

- **SNIFFER** (Qi et al., 2024) extends task-specific instruction fine-tuning to the general-purpose multimodal large language model for end-to-end training. It utilizes internal and external checking to perform the consistency between images and texts.
- **TRUST-VL** (Yan et al., 2025) is a unified and explainable multimodal large language model, which integrates Question-Aware Visual Amplifier to extract features and support training across multiple modalities.

A.2 MMD-Instruct Details

As illustrated in Table 5, we provide the distribution of different thinking modes in our constructed MMD-Instruct dataset. Specifically, we maintain an equal number of samples for Intuitive Response $\mathbf{m}_1$ and Shallow Deliberation $\mathbf{m}_2$. This helps mitigate the mode-induced bias caused by both thinking modes. In contrast, we allocate more samples to Deep Contemplation $\mathbf{m}_3$ that guides the mode to conduct deep thinking for complex scenarios. Additionally, we maintain the ratio of true samples to false samples at 3:7. This is consistent with MMFakeBench (Liu et al., 2025b) in the era of AIGC.

A.3 Visualization of Thinking Mode

As shown in Figure 4, we further visualize dynamic thinking mode between vanilla GRPO and Adaptive Thinking Mode Policy Optimization (ATMPO) at the stage of Reinforcement Learning (RL). From the comparison, we achieve the following findings: (1) The variant trained with vanilla GRPO is sensitive to the data allocation across thinking modes. It exhibits a mode-induced bias toward Deep Contemplation $\mathbf{m}_3$, thereby leading to perform overthinking in most of scenarios. This prevents adaptive thinking capabilities of the model and increases additional computational resource usage in inference. However, (2) Our proposed ATMPO focuses on both mode and sample, which guides the model for thinking mode switching across different scenarios. For most of scenarios, MMD-Thinker adopts Intuitive Response $\mathbf{m}_1$ or Shallow Deliberation $\mathbf{m}_2$ for judgments. For complex scenarios, MMD-Thinker employs Deep Contemplation $\mathbf{m}_3$ to address them. Therefore, the results highlight the effectiveness of ATMPO of MMD-Thinker in the task of multimodal misinformation detection.