Evidence-Grounded Forensic Reasoning for Detecting and Grounding Multi-Modal Media Manipulation

Yichun Yeh*
yeyijun2024@ia.ac.cn
MAIS, CASIA; SAI, UCAS
Beijing, China

Yiheng Li*
liyiheng2024@ia.ac.cn
SAI, UCAS; MAIS, CASIA
Beijing, China

Xiaobo Hu
huxiaobo2024@ia.ac.cn
MAIS, CASIA; SAI, UCAS
Beijing, China

Zhen Lei†
zhen.lei@ia.ac.cn
MAIS, CASIA; SAI, UCAS
Beijing, China

Yang Yang†
yang.yang@nlpr.ia.ac.cn
MAIS, CASIA; SAI, UCAS
Beijing, China

arXiv:2608.08009v1 [cs.CV] 8 Aug 2026

Abstract

Fake news increasingly relies on cross-modal image-text forgeries, making transparent and verifiable reasoning chains an urgent need for Detecting and Grounding Multi-Modal Media Manipulation (DGM4). Existing methods produce black-box detection results without any decision rationale, limiting their reliability in forensic practice. Multi-modal Large Language Models (MLLMs) offer a natural path toward explainability, but applying them to DGM4 raises two difficulties. First, models tend to generate explanations disconnected from predicted evidence locations, producing unverified attribution. Second, enforcing evidence-conclusion consistency requires active optimization, yet uniform training signals fail to distinguish localization tokens from classification tokens, making multi-head joint training unreliable. We propose a multi-modal manipulation detector based on an Evidence-Grounded Forensic Reasoning (EFR) framework. EFR introduces an Anchor-and-Verify reasoning chain that enforces modality-isolated perception before cross-modal comparison, with conclusion coordinates as explicit anchors to which downstream evidence must spatially correspond. A verifiable reward system then enforces evidence-conclusion consistency during training, while a Modality-Decoupled Advantage (MDA) routing mechanism mitigats credit misassignment across prediction tasks. Experiments show that EFR achieves state-of-the-art performance while producing structured forensic reasoning records that explicitly bind explanations to evidence.

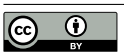
CCS Concepts

• **Computing methodologies** → **Biometrics**; **Natural language generation**.

Keywords

Media Manipulation Detection, Multi-Modal Large Language Model, Reasoning, Reinforcement Learning, Multi-Modal

*Both authors contributed equally to the paper.
†Corresponding authors.



This work is licensed under a Creative Commons Attribution 4.0 International License.
MM '26, Rio de Janeiro, Brazil
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2213-4/2026/11
https://doi.org/10.1145/3767308.3835015

ACM Reference Format:
Yichun Yeh, Yiheng Li, Xiaobo Hu, Zhen Lei, and Yang Yang. 2026. Evidence-Grounded Forensic Reasoning for Detecting and Grounding Multi-Modal Media Manipulation. In *Proceedings of the 34th ACM International Conference on Multimedia (MM '26), November 10–14, 2026, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3767308.3835015

1 Introduction

The rapid spread of deepfake technology and large language models [2, 32] has made multi-modal misinformation increasingly easy to produce and hard to detect [29]. Isolated face manipulation [52] and textual fabrication [27] have expanded into coordinated cross-modal forgeries where images and their associated texts are manipulated simultaneously, posing escalating threats to the credibility of public media. To address these threats, the DGM4 task [34] requires models to jointly perform authenticity classification, manipulation type identification, and fine-grained localization of manipulated faces and text tokens, a combination that far exceeds the complexity of single-modal detection. Existing DGM4 methods follow two main directions: cross-modal feature alignment and semantic interaction [18, 51], and fine-grained localization through richer supervision such as frequency-domain features and multi-scale objectives [23, 25, 47]. Despite steady progress, they treat detection as a closed prediction problem, outputting labels and bounding boxes without any decision rationale. Yet real-world forensic workflows such as editorial fact-checking and legal content authentication demand decisions that can be audited and challenged. These limitations call for a detection paradigm that produces structured, verifiable reasoning alongside its predictions.

Beyond these task-specific methods, MLLMs [21, 22, 42] have recently been applied to single-modality manipulation detection with localization and explanation [9, 14, 24, 44]. Leveraging cross-modal understanding and language generation, these approaches [11, 20, 45] offer a natural path toward explainable detection, with recent efforts adding pattern-aware reasoning, self-reflection, and two-stage training [5, 13, 38, 40]. Rather than pursuing generic reasoning, we identify a critical yet overlooked flaw shared by these methods: their explanations are never tied to the model's own predicted localization, so their interpretability remains superficial. This motivates a new direction, and applying MLLMs to DGM4 further exposes two problems amplified by its cross-modal complexity. First, *unverified attribution*: without structural constraints, MLLMs

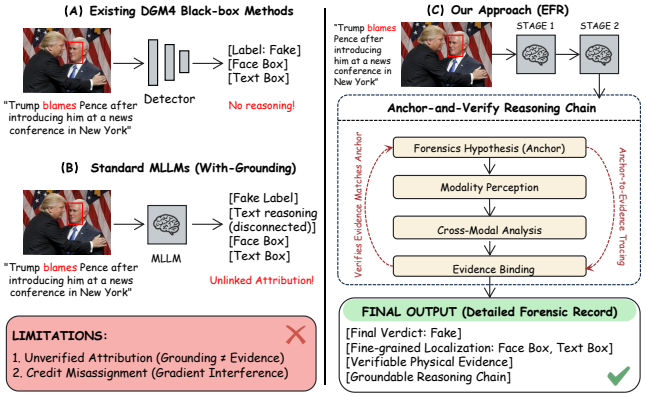


Figure 1: Comparison of existing approaches and our EFR framework. (A) Existing DGM4 methods produce fine-grained results but lack decision rationale. (B) MLLMs offer explanations, yet the reasoning is disconnected from predicted evidence locations, leaving attribution unverified. (C) Our EFR introduces an Anchor-and-Verify reasoning chain that places conclusion coordinates as explicit anchors and requires downstream evidence to spatially match them, enforced by a verifiable reward system during training.

produce coherent-looking reasoning yet cannot verify that cited evidence matches the predicted conclusion coordinates [55, 56]. Second, *credit misassignment*: DGM4 packs heterogeneous heads (classification, face localization, and text localization) into one output sequence, so a uniform signal over all tokens lets gradients from one head interfere with others, destabilizing joint training.

To address both problems, we propose the Evidence-Grounded Forensic Reasoning (EFR) framework. Since existing benchmarks provide no forensic rationale and contain labeling noise, we first construct a high-quality reasoning dataset as the training foundation. EFR then operates in two coordinated stages. The first stage introduces an Anchor-and-Verify reasoning chain that organizes outputs into modality-isolated perception, cross-modal conflict analysis, and evidence binding, placing conclusion coordinates as explicit anchors that downstream evidence must match, thereby closing the attribution gap. The second stage then optimizes the model with reinforcement learning: a five-component verifiable reward system reinforces this coordinate-level consistency, while a Modality-Decoupled Advantage (MDA) routing mechanism sends task-specific advantages to their corresponding tokens, eliminating the cross-head interference. Our main contributions are as follows:

- We propose EFR, an evidence-grounded forensic reasoning framework for DGM4 that, unlike black-box detectors and unconstrained MLLM explanations, grounds every forensic conclusion in a spatially verifiable location through anchor-first verifiable binding.
- We introduce an Anchor-and-Verify reasoning chain that constrains cited evidence to match the predicted anchors, together with a Modality-Decoupled Advantage routing scheme that delivers task-specific signals to their corresponding prediction tokens; both are trained through a two-stage pipeline over a curated 50K forensic reasoning dataset.
- Extensive experiments on DGM4 show that EFR attains state-of-the-art detection and competitive grounding, while producing structured reasoning records whose cited evidence is spatially consistent with the predicted anchors.

2 Related Works

Multi-Modal Manipulation Detection. Early efforts addressed coarse-grained binary classification for fake-news and out-of-context misinformation detection, treating an image-text pair as a single real-or-fake decision. HAMMER [34] reframes the problem as DGM4, which jointly detects manipulation, classifies its fine-grained type, and grounds the tampered image regions and text tokens. Subsequent work advances the task along two directions. One strengthens cross-modal alignment so that subtle semantic inconsistencies become separable: ASAP [51] adapts prompts to the input pair, while CSCL [18] enforces consistency-based objectives across modalities. The other sharpens localization through richer supervision, as in UFAFormer [23] with frequency-domain cues, IDseq [25] with decoupled grounding, and FMSN [47] with multi-scale signals. Beyond these, RamDG [37] targets coordinated semantic manipulations, and LADLE-MM [3] and CIEC [48] tackle limited-annotation settings, while DGM4+ [39] and MDSM [50] extend the benchmark with global scene inconsistencies and MLLM-crafted narratives. Despite steady progress, these methods remain prediction-oriented, producing bounding boxes and labels without any rationale behind their decisions.

MLLMs for Forgery Detection. The strong cross-modal reasoning of MLLMs has shifted forgery detection toward explainability, moving beyond a binary verdict to a human-readable justification. Early work [12] probes their zero-shot capability through prompt engineering, and later methods [4, 11, 30, 45] tie fine-grained tampered regions to textual explanations via multi-granularity prompts. To make the reasoning process explicit, AIGI-Holmes [54] and FakeReasoning [5] impose chain-of-thought formats aligned with forgery attributes, while VERITAS [40] and RAIDX [17] improve generalization through pattern-aware reasoning and retrieval-augmented evidence. More recently, DF-LLaVA [38] injects external knowledge with conflict-driven self-reflection for single-image detection, and TriDF [13] evaluates perception and hallucination to gauge interpretability. Yet all of these generate explanations without enforcing consistency between the cited evidence and the predicted conclusion coordinates, so a plausible-sounding rationale may still point elsewhere, leaving attribution unverified; applied to the multi-head DGM4 task, they further suffer credit misassignment under a uniform training signal. In contrast, EFR enforces anchor-first, coordinate-level verifiable binding and decouples per-task optimization, addressing both limitations.

Reinforcement Learning for LLMs. RL-based post-training has evolved from PPO-based RLHF toward scalable reasoning optimization [28, 31, 33]. GRPO [36] removes the explicit value model, estimating advantages from group-level reward comparisons. Later variants refine this scheme: DAPO [46] counters entropy collapse

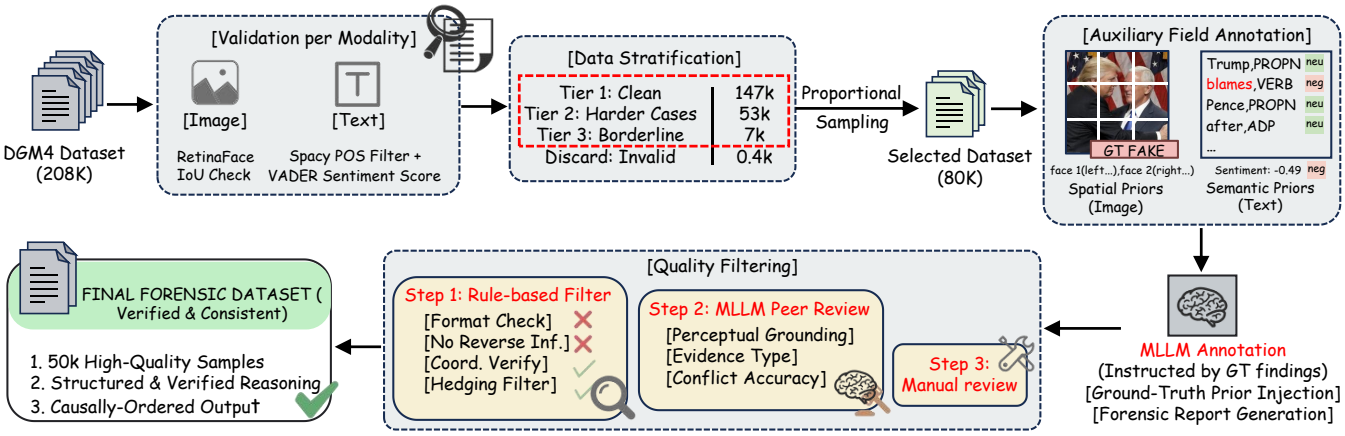


Figure 2: The forensic reasoning annotation data construction pipeline. Raw DGM4 samples (208K) are filtered and stratified by quality, with spatial and semantic priors extracted to guide MLLM annotation. A three-step quality filtering stage combining rule-based checks, MLLM peer review, and manual inspection yields 50K verified forensic reasoning samples.

and training instability, SRPO [49] scales via a two-stage curriculum, GSPO [53] stabilizes updates with sequence-level importance ratios, and GDPO [26] decouples reward normalization for multi-reward settings. All of them, however, apply the reward uniformly across output tokens, ignoring the semantic boundaries of structured multi-modal outputs. Our MDA routes task-specific advantages along these boundaries within the reasoning chain, mitigating credit misassignment in multi-head training.

3 Methodology

3.1 Overview

Directly applying MLLMs to DGM4 exposes two fundamental problems: *unverified attribution*, where explanations become disconnected from predicted manipulation locations, and *credit misassignment*, where joint optimization across heterogeneous prediction heads suffers under uniform advantage estimation. As illustrated in Figure 3, EFR addresses both through two coordinated stages. Stage 1 (Sec. 3.2) tackles unverified attribution with a structured reasoning chain that binds forensic conclusions to spatially grounded evidence, internalized via supervised fine-tuning. Stage 2 (Sec. 3.3) mitigates credit misassignment by replacing uniform advantage estimation with task-specific signals routed to their corresponding prediction tokens.

3.2 Structured Reasoning Cold-Start

To equip the model with structured forensic reasoning capability, we construct a high-quality forensic reasoning dataset from the DGM4 benchmark and use it to initialize the model via supervised fine-tuning to internalize the Anchor-and-Verify structure.

3.2.1 Anchor-and-Verify Reasoning Chain. Without structural constraints, MLLMs produce reasoning chains that appear coherent but lack a mechanism to verify that cited evidence corresponds to predicted conclusion coordinates. We address this by designing the Anchor-and-Verify reasoning chain, a structured chain-of-thought format that organizes model outputs into four ordered stages, each serving a distinct forensic function, as illustrated in Figure 1.

Forensic Hypothesis. All predicted outputs are placed at the beginning of the sequence as a Forensic Hypothesis Block. The predicted coordinates serve as explicit anchors that all downstream stages must spatially correspond to, establishing a verifiable reference before any explanation is generated.

Modality-Isolated Perception. Independent analysis is enforced through two strictly separated sub-modules. The visual module describes only physically observable image features, prohibited from referencing caption content or making authenticity judgments; the text module independently analyzes caption semantics and logical structure without referencing the image. This separation reduces the influence of language priors on visual analysis.

Cross-Modal Conflict Analysis. The two perceptual outputs are explicitly compared and their disagreement is graded into two levels: strong conflicts, where the image and text assert mutually exclusive facts, and weak mismatches, where the two modalities are thematically related but neither corroborates nor contradicts the other. This graded assessment produces fine-grained, evidence-grounded conflict judgments instead of a coarse binary decision.

Physical Evidence Binding. Conflict analysis is grounded in concrete physical locations. Let $\hat{b}$ and $\hat{s}$ denote the predicted face bounding box and text token span from the Forensic Hypothesis Block. Visual evidence must cite a bounding box $\hat{b}'$ with verifiable observations, and text evidence must cite a token span $\hat{s}'$ with an explanation of the semantic discontinuity. Both must satisfy $\hat{b}' \approx \hat{b}$ and $\hat{s}' \subseteq \hat{s}$, ensuring every forensic conclusion is traceable to a specific location. This constraint is enforced by the verifiable reward system during reinforcement learning (Sec. 3.3.1).

3.2.2 Forensic Reasoning Data Construction. With the Anchor-and-Verify chain established as the target structure, we construct the training data needed to instantiate it. The DGM4 dataset provides manipulation labels and localization annotations but no forensic rationale, and its 208K samples include substantial labeling noise.

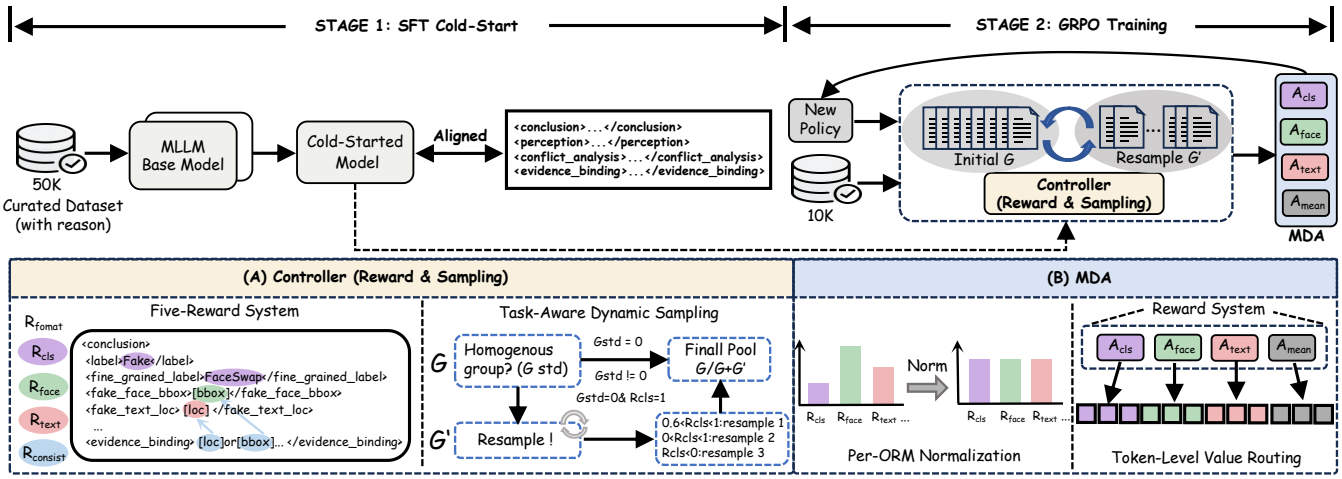


Figure 3: Overview of the two-stage EFR training framework. Stage 1 fine-tunes a base MLLM on 50K curated forensic reasoning samples to initialize structured output following the Anchor-and-Verify format. Stage 2 optimizes the cold-started model via GRPO. (A) A five-component verifiable reward system combined with task-aware dynamic resampling maintains training diversity. (B) MDA performs per-ORM normalization across modality-specific reward heads and routes task-specific advantage values to their corresponding prediction tokens, mitigating credit misassignment.

We design a systematic pipeline to clean, stratify, annotate, and filter the data, as shown in Figure 2.

Data cleaning and stratification. Samples are first validated per modality for face-localization and text-modification quality (details in the supplementary). Based on these quality dimensions and manual review of 6K critical samples, the dataset is stratified into three tiers: Tier 1 (147K clean), Tier 2 (53K harder multi-face or complex-token cases), and Tier 3 (7K borderline); invalid samples (0.4K) are discarded. Proportional sampling over Tier 1 and Tier 2 yields an 80K dataset for annotation.

Auxiliary field annotation. Before reasoning generation, each sample is enriched with structured spatial and semantic priors: per-face position and area-ranked indices with an explicit <FAKE> tag on the ground-truth fake face for images, and a token map with polarity scores for eligible candidate tokens for text. These priors give the generation model unambiguous references that mitigate positional confusion in multi-face scenes and ground text localization in precise token structure.

Reasoning generation and quality filtering. To generate reasoning chains, we condition a large-scale MLLM on the ground-truth findings, so that it reports physically consistent evidence for a known verdict rather than inferring the verdict itself. Each generated chain then passes three filters: rule-based checks on format and coordinate references, an independent MLLM review of perceptual grounding and conflict accuracy, and manual inspection of flagged cases. This yields the final 50K forensic reasoning samples.

3.2.3 Structured Reasoning Cold-Start. Direct RL training on the Anchor-and-Verify format is unstable, as the complex structured output requires the model to produce parseable multi-block responses before any reward signal can be meaningfully applied. We therefore first perform supervised fine-tuning (SFT) on our curated 50K forensic dataset $\mathcal{D}$, establishing a reliable reference policy π_{ref} that can consistently adhere to the formatting constraints. Following standard instruction-tuning paradigms [28], our optimization objective is to minimize the autoregressive cross-entropy loss:

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

where x represents the multi-modal prompt, $y = (y_1, \ldots, y_L)$ denotes the target sequence of length L comprising the complete Anchor-and-Verify reasoning chain, and π_θ is the policy parameterized by θ. Through this phase, the model internalizes the Anchor-and-Verify reasoning structure and learns to produce physically consistent observations and cross-modal evidence, providing a robust initialization for the subsequent RL stage.

3.3 Modality-Decoupled Policy Optimization

While cold-start SFT equips the model with format adherence, supervised imitation provides no explicit feedback on spatial consistency between conclusion anchors and cited evidence, nor does it resolve credit misassignment across multi-head outputs. We therefore apply reinforcement learning on top of the cold-start model π_{ref}, building on Group Relative Policy Optimization (GRPO) [36], which estimates policy gradients from group-level reward comparisons without a separate value network. For each prompt x, a group of G completions $\{y_i\}_{i=1}^{G}$ is sampled from the current policy π_θ, and the objective is:

$$\mathcal{L}_{\text{GRPO}} = -\mathbb{E}\Bigg[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{T_i}\sum_{t=1}^{T_i}\min\Big(\rho_{i,t}A_i, \text{clip}(\rho_{i,t}, 1-\varepsilon, 1+\varepsilon)A_i\Big) - \beta\mathbb{D}_{\text{KL}}\left[\pi_\theta \| \pi_{\text{ref}}\right]\Bigg], \tag{2}$$

where T_i is the token length of completion i; ε is the clipping threshold that limits the update of the policy per-step; and β is the KL penalty coefficient that regularizes π_θ with the reference policy for cold-start π_{ref}. The importance sampling ratio $\rho_{i,t}$ and the group-normalized advantage A_i are:

$$\rho_{i,t} = \frac{\pi_\theta(y_{i,t} \mid y_{i,<t}, x)}{\pi_{\text{ref}}(y_{i,t} \mid y_{i,<t}, x)}, \tag{3}$$

$$A_i = \frac{R_i - \mu_G}{\sigma_G + \epsilon}, \quad \mu_G = \frac{1}{G}\sum_{i=1}^{G} R_i, \quad \sigma_G = \sqrt{\frac{1}{G}\sum_{i=1}^{G}(R_i - \mu_G)^2}, \tag{4}$$

where R_i is the total scalar reward for completion i and ϵ is a small constant for numerical stability.

Applying GRPO directly to DGM4, however, is problematic: aggregating heterogeneous prediction targets operating on different reward scales into a single R_i produces noisy advantage estimates, and the uniform scalar advantage A_i applied identically to all tokens conflates gradients across classification, localization, and reasoning tokens. We address both with a fine-grained verifiable reward system and an MDA routing mechanism.

3.3.1 Verifiable Reward System. Designing effective rewards for DGM4 requires covering output format, classification accuracy, multi-modal localization, and reasoning consistency in a single framework. Inspired by verifiable reward design in mathematical reasoning [7, 36], we define five independent, deterministic rule-based Outcome Reward Models (ORMs), which assess our structured outputs precisely while keeping training signals stable and reproducible. To decouple signals across tasks, samples are partitioned by manipulation type (*face*, *text*, *compound*, *orig*); inapplicable ORMs are excluded from the per-sample sum, so gradients flow only through relevant heads.

Format Compliance (R_{format}) gates the reward: it enforces XML structural completeness and penalizes placeholder copying, ensuring outputs are parseable before other ORMs score them.

Hierarchical Classification (R_{cls}) jointly evaluates binary authenticity and fine-grained manipulation type with a cost-sensitive design, where missed detections are penalized more heavily than false positives. For valid predictions, fine-grained reward follows a hierarchical partial-order design:

$$R_{\text{cls}} = r_{\text{base}} + w_{\text{fam}} \cdot F1_{\text{fam}} + w_{\text{sub}} \cdot r_{\text{sub}} + w_{\text{exact}} \cdot \mathbb{1}[\hat{y} = y], \tag{5}$$

where r_{base} is a base score, $F1_{\text{fam}}$ measures family-level overlap (e.g., face vs. text family), r_{sub} captures within-family subtype accuracy, and $w_{\text{fam}}, w_{\text{sub}}, w_{\text{exact}}$ are fixed weights reflecting the taxonomy.

Localization (R_{face}, R_{text}) score spatial grounding for the relevant categories: R_{face} maps predicted-box IoU to a tiered reward, and R_{text} measures token-span overlap via Jaccard similarity with subtype-specific curves.

Consistency ($R_{\text{consistency}}$) verifies internal coherence of the reasoning chain, rewarding cases where the evidence cited in the `evidence_binding` block spatially and semantically matches the conclusion anchors and where evidence types match the predicted category, thereby instantiating the physical binding constraints of Sec. 3.2.1. The full decomposition is given in the supplementary.

The five components are integrated into a final scalar reward for each response:

$$R = \lambda_f R_{\text{format}} + \lambda_c R_{\text{cls}} + \lambda_v R_{\text{face}} + \lambda_t R_{\text{text}} + \lambda_r R_{\text{consistency}}, \tag{6}$$

where $\lambda_f, \lambda_c, \lambda_v, \lambda_t, \lambda_r$ balance the contribution of each component. While these ORMs provide task-specific signals, translating them into precise gradient updates requires routing each signal to its corresponding prediction tokens, addressed by MDA next.

3.3.2 Modality-Decoupled Advantage Routing. The verifiable reward system provides fine-grained signals across five task dimensions; MDA routes per-task advantages to their corresponding tokens to turn these into precise gradient updates (Figure 3).

Per-ORM Normalization. Since R_{format}–$R_{\text{consistency}}$ operate on different ranges and distributions, aggregating them into a single advantage distorts the relative contribution of each task. MDA normalizes each ORM independently within the group:

$$\hat{r}_k^{(i)} = \frac{r_k^{(i)} - \mu_k}{\sigma_k + \epsilon}, \tag{7}$$

where $r_k^{(i)}$ is the reward from ORM k for completion i, and μ_k, σ_k are the group mean and standard deviation of that ORM. Task-specific advantages are then derived: A_{cls} from $\hat{r}_{\text{cls}}$; A_{face} from $\hat{r}_{\text{face}}$ and the face-relevant portion of $\hat{r}_{\text{consistency}}$; and A_{text} from $\hat{r}_{\text{text}}$ and the text-relevant portion of $\hat{r}_{\text{consistency}}$.

Token-Level Routing. This routing is made possible by the structured XML boundaries established in Sec. 3.2.1: the Anchor-and-Verify format provides explicit block-level semantic boundaries that map each output token to a specific prediction task. MDA parses these boundaries at training time and assigns token-level advantages as:

$$A_t = \begin{cases} \sum_{k \in \mathcal{K}} A_k \cdot \mathbb{1}[t \in \mathcal{T}_k] & \text{if } t \in \bigcup_k \mathcal{T}_k \\ \bar{A} & \text{otherwise} \end{cases}, \mathcal{K} = \{\text{cls}, \text{face}, \text{text}\}, \tag{8}$$

where $\mathcal{T}_k$ is the set of token positions belonging to task k, and $\bar{A}$ is the mean advantage across all tasks, applied to shared reasoning-chain tokens that do not belong to any specific prediction block. Task-specific tokens receive the advantage of their corresponding head; reasoning-chain tokens receive $\bar{A}$, reflecting their shared role across all prediction tasks. As a global parseability gate, R_{format} is not routed to any single block but scales the whole-sequence reward.

Following DAPO [46], we exclude the KL penalty term from the objective (i.e., $\beta D_{\text{KL}}[\pi_\theta \| \pi_{\text{ref}}]$), as the model distribution is expected to diverge substantially from the cold-start reference during forensic reasoning, making this constraint unnecessary and potentially harmful to exploration. The final objective replaces the uniform scalar advantage in Eq. 2 with the token-level routed advantage:

$$\mathcal{L}_{\text{MDA}} = -\mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{T_i}\sum_{t=1}^{T_i}\min\Big(A_t^{(i)}\rho_{i,t}, A_t^{(i)}\,\text{clip}(\rho_{i,t}, 1-\varepsilon, 1+\varepsilon)\Big)\right], \tag{9}$$

Task-Aware Dynamic Resampling. Effective token-level routing requires stable and informative advantage estimates for each task

Table 1: Comparison of state-of-the-art methods on DGM4.↓ means lower is better. Bold/underline: best/second best in each group. † denotes zero-shot evaluation without fine-tuning. ‡For MLLM-based methods, mAP is reported as a/b: from token log probabilities and from the hard 0/1 prediction, respectively (Sec. 4.1).

Method	Binary Cls			Multi-label Cls			Image Grounding			Text Grounding		
	AUC	EER↓	ACC	mAP‡	CF1	OF1	IoU_m	IoU_{50}	IoU_{75}	PR.	RE.	F1
HAMMER [34] (CVPR'23)	93.19	14.10	86.39	86.22	79.37	80.37	76.45	83.75	76.06	75.01	68.02	71.35
HAMMER++ [35] (TPAMI'24)	93.33	14.06	86.66	86.41	79.73	80.71	76.46	83.77	76.03	73.05	72.14	72.59
ViKI [16] (IF'24)	93.51	13.87	86.67	86.58	81.07	80.10	76.51	83.95	75.77	77.79	66.06	73.44
UFAFormer [23] (IJCV'24)	93.81	13.60	86.80	87.85	80.31	81.48	78.33	85.39	79.20	73.35	70.73	72.02
MSF [41] (ICASSP'24)	95.11	11.36	88.75	**91.42**	83.60	84.38	80.83	88.35	80.39	76.51	70.61	73.44
IDseq [25] (AAAI'25)	94.55	11.40	88.94	90.01	83.00	84.90	83.33	89.39	**86.10**	75.96	71.23	73.52
ASAP [51] (CVPR'25)	94.38	12.73	87.71	88.53	81.72	82.89	77.35	84.75	76.54	**79.38**	73.86	**76.52**
Qwen3-VL-8B† [1]	64.42	40.06	54.71	30.34/26.48	0.02	0.02	32.65	32.54	32.54	21.66	1.59	2.97
InternVL3.5-8B† [43]	50.00	50.00	33.28	28.26/20.76	0.00	0.00	13.00	12.60	12.55	14.78	0.84	1.59
EFR (cold-start)	96.87	9.38	90.65	89.60/81.50	86.94	86.68	83.22	89.24	85.21	78.68	72.14	75.27
EFR (ours)	**96.97**	**9.18**	**90.82**	90.41/81.58	**87.19**	**87.05**	**83.49**	**89.41**	85.43	76.00	**75.61**	75.80

dimension. DGM4's heterogeneous task structure makes this non-trivial: original samples activate only a subset of ORMs and tend to produce low reward variance within a group, causing advantage estimates to degenerate. To stabilize training, MDA adopts a task-aware dynamic resampling strategy. Groups meeting task-specific quality thresholds across all active ORMs are exempted; for others, additional completions are sampled with intensity proportional to per-ORM error severity, ensuring every prediction head receives informative gradient signals.

By integrating per-ORM normalization, token-level routing, and task-aware resampling, MDA ensures that each prediction head receives optimization signals derived only from task-relevant reward dimensions. This mechanism is made possible by the explicit structural boundaries of the Anchor-and-Verify reasoning chain. Finally, the reasoning format and the optimization mechanism are tightly co-designed to be mutually reinforcing.

4 Experiments

4.1 Setup

Datasets. We evaluate EFR on the DGM4 dataset [34], a large-scale benchmark for detecting and grounding multi-modal media manipulation in the news domain. It comprises 230,000 image-text pairs (77,426 genuine and 152,574 manipulated) from four major outlets (The Guardian, BBC, USA TODAY, and The Washington Post), with manipulations spanning *face swap*, *face attribute*, *text swap*, and *text attribute*, covering both visual and textual forgery.

Evaluation Metrics. Following prior works [18, 25, 34, 35, 51], we adopt a comprehensive metric suite covering both coarse-grained detection and fine-grained grounding. Binary detection is assessed by AUC, EER (↓), and ACC; multi-label classification by mAP, CF1, and OF1. Image grounding is measured by IoU_m, IoU_{50}, and IoU_{75}; text grounding by token-level Precision (PR), Recall (RE), and F1.

As EFR generates predictions autoregressively, AUC is computed from the token-level log probabilities of the predicted authenticity label. Fine-grained labels are discrete and provide no continuous confidence score. For a fair comparison with prior work, we recover the per-class confidences from the token-level log probabilities at the fine-grained-label position and evaluate mAP under the standard protocol; we additionally report mAP on the hard 0/1 prediction, which is used in our ablations.

Implementation Details. We build EFR on Qwen3-VL-7B [1], and use Qwen3-VL-32B [1] as the annotation model for forensic reasoning data construction, generating evidence reports conditioned on ground-truth findings via confirmed-findings injection. Training proceeds in two stages [7, 36, 46], both using LoRA [10] (r=128, α=256). Stage 1 performs cold-start SFT on 50K forensic reasoning samples for 3 epochs (learning rate 5×10^{-5}, batch size 64) to initialize the Anchor-and-Verify structure. Stage 2 applies GRPO [36] on 10K samples for 1 epoch (learning rate 1×10^{-6}, batch size 16, group size G=8, temperature 1.0). Following DAPO [46], both KL coefficients β and β' are set to 0, as the policy is expected to diverge substantially from the cold-start reference during optimization.

4.2 Results

Table 1 compares EFR against dedicated DGM4 detectors and general-purpose MLLMs. EFR achieves state-of-the-art binary detection and multi-label F1, remains competitive on mAP and grounding.

Comparison with State-of-the-Art Methods. The advantage of EFR is concentrated on fine-grained attribution rather than detection. Binary detection is close to saturation for all recent methods, and EFR advances it by a modest 1.86 AUC over the strongest baseline. In multi-label classification the picture changes: CF1 and OF1 improve by 3.59 and 2.15 points, roughly twice the margin. Determining whether an item has been manipulated reduces to a low-level consistency check, whereas identifying which of the four manipulations occurred requires each hypothesis to be evaluated against its own evidence, the procedure the Anchor-and-Verify chain makes explicit. The grounding results are consistent with this account. EFR obtains the highest IoU_m and IoU_{50} despite emitting coordinates as text tokens, and falls behind IDseq only at the

Table 2: Ablation on MDA routing components. Each row adds one component to the previous. Bold/underline: best/second best. mAP is the hard 0/1 variant (Sec. 4.1).

Method	ACC	mAP	IoU$_m$	F1
SFT + GRPO	90.74	78.88	82.68	75.16
+ Task-Aware Dyn. Resampling	**90.86**	80.26	82.50	75.69
+ Per-ORM Normalization	90.41	77.37	82.78	74.17
+ Token-Level Routing (*Ours*)	90.82	**81.58**	**83.49**	**75.80**
Δ vs. SFT + GRPO	+0.08	+2.70	+0.81	+0.64

Table 3: Ablation on reward components in MDPO. ✓denotes an active reward. Bold/underline: best/second best. mAP is the hard 0/1 variant (Sec. 4.1).

R_{fmt}	R_{acc}	R_{bbox}	R_{text}	R_{consist}	ACC	mAP	IoU$_m$	F1
✓	✓				**90.84**	77.21	82.13	75.49
✓	✓	✓			90.78	81.35	83.26	75.80
✓	✓		✓		90.83	81.08	81.83	75.60
✓	✓	✓	✓		90.78	**81.59**	83.42	**75.89**
✓	✓	✓	✓	✓	90.82	81.58	**83.49**	75.80

strictest threshold, where a dedicated localization head retains sub-pixel precision beyond the resolution of a language decoder. mAP is the only metric where a discriminative head stays ahead, and the cause is protocol, not capability: prior methods read mAP off continuous multi-label scores, while ours comes from token log probabilities (Sec. 4.1), a coarser signal that hurts ranking but not the decisions, as CF1 and OF1 confirm.

Effect of Stage 2 Optimization. Stage 2 does not make the model uniformly better; it moves the model to a different operating point. Detection barely changes, as expected from a metric already above 96.8 after cold-start. What changes is text grounding, which after cold-start is strongly precision-skewed: the model reports only the spans it is certain about and misses the rest. Stage 2 trades 2.68 points of precision for 3.47 points of recall, turning an asymmetric operating point into a balanced one and improving F1. This is the intended effect of the reward design. Static annotations supervise the exact span and provide no signal that separates a partially recovered span from a missed one, so supervised training has no reason to extend a confident partial prediction; the verifiable rewards score coverage directly, and the model responds by widening its evidence.

Zero-shot MLLMs. General-purpose MLLMs can rank but cannot decide. Qwen3-VL-8B and InternVL3.5-8B obtain mAP of 30.34 and 28.26, close to what the class priors alone would yield, yet their CF1 and OF1 collapse to near zero: the models occasionally place manipulated items above authentic ones, but almost never emit the correct label. Without task-specific training, visual-linguistic competence does not translate into forensic attribution.

4.3 Ablation Studies

All ablations report mAP under the hard 0/1 protocol. EFR generates fine-grained labels autoregressively as text instead of scoring a fixed class set, so no continuous per-class confidence is available. Token log probabilities are only an indirect surrogate, since a single label spans several tokens whose distributions have to be recombined. We therefore use the log-probability variant only in Table 1, where comparability with prior work requires it.

Table 4: Reasoning quality of SFT vs. MDPO, on the <evidence_binding> and <perception> blocks of correctly predicted samples. ↑/↓: higher/lower better.

Module	Metric	Dir.	SFT	MDPO	Δ
Evidence	NLI Coherence [8]	↑	38.73	**41.07**	+2.34
	Redundancy [6]	↓	0.35	**0.22**	-0.13
	ROUGE-L [19]	↑	42.99	**43.56**	+0.57
Perception	Distinct-2 [15]	↑	8.99	**9.21**	+0.22
	Distinct-3 [15]	↑	18.99	**19.31**	+0.32

Ablation on MDA. Table 2 builds MDA over an SFT + GRPO baseline. The baseline shares a single scalar advantage across all reward heads, so gradients from easily satisfied heads dominate and mAP remains at 78.88. Task-Aware Dynamic Resampling draws rollouts in proportion to per-ORM error severity and raises mAP by 1.38 points. Uniformly correct groups carry no advantage signal, and replacing them restores usable gradients. Per-ORM Normalization then lowers mAP to 77.37, 1.51 points below the baseline it extends. Normalization rescales each reward into its own range, yet the resulting advantages are still broadcast over all tokens, so the rescaled signals compete on shared parameters and partially cancel. Token-Level Routing delivers each advantage only to the tokens that produced the corresponding output, recovering 4.21 points of mAP and reaching the best IoU$_m$ and F1 in the table. Normalization and routing therefore form one mechanism rather than two independent gains. Separating reward scales helps only when the separated signals stay separate in the backward pass.

Ablation on Reward Components. Table 3 shows that detection and attribution respond to different rewards. ACC varies by 0.06 points across all five configurations, so the binary decision is settled once R_{acc} is active and no grounding reward alters it. Attribution behaves in the opposite way. With R_{fmt} and R_{acc} alone the model attains its highest ACC together with the weakest mAP and IoU$_m$ in the table, having learned to name a manipulation without locating it. Supervising location repairs attribution even though it adds no label information, with R_{bbox} alone raising mAP by 4.14 points and R_{text} alone by 3.87. Being required to point at the evidence constrains which label the model can justify. The two grounding rewards are not interchangeable. Applied in isolation, R_{text} pushes IoU$_m$ below the configuration without any grounding reward, so optimizing text localization alone draws capacity away from face localization, and enabling both rewards removes the interference and yields the best mAP and F1. Adding R_{consist} finally gives the best IoU$_m$ while leaving mAP and F1 unchanged to within 0.1 point, tightening anchor-evidence agreement at no measurable cost, and this configuration is retained as the default.

4.4 Analysis of Generated Reasoning

Reasoning Quality Analysis. To assess the reasoning itself independently of label accuracy, we restrict the evaluation to correctly predicted manipulated samples and score the <evidence_binding>

Figure 4: Forensic reasoning output of EFR on a compound `face_attribute & text_attribute` sample. Zero-shot collapses to an authentic prediction; SFT recovers the face label but misses the text manipulation; EFR identifies both, localizes the altered face and token, and produces modality-specific evidence grounding for each.

sub-block of the ground-truth modality and the <perception> block with NLI entailment [8], repeated-4-gram redundancy [6], Distinct-2/3 [15], and ROUGE-L [19]. MDPO improves every metric in Table 4. NLI entailment gains the most (+2.34), indicating that the cited evidence follows more tightly from the anchor, and redundancy falls by over a third. Distinct-2/3 both rise, so the shorter chains do not collapse onto a template, a common outcome of reward optimization. ROUGE-L is the only reference-based metric and is biased toward SFT, whose outputs closely follow the surface form of the reference chains. MDPO improves it nonetheless.

Qualitative Analysis. Figure 4 shows a representative compound `face_attribute & text_attribute` sample under three conditions.Orange marks the labels, localization, and reasoning evidence that the weaker baselines miss or predict incorrectly. The zero-shot model collapses to an `authentic` verdict with empty localization and no evidence binding, failing to engage the forensic task at all. SFT cold-start recovers the `face_attribute` label and localizes the manipulated face, but overlooks the co-occurring text manipulation entirely, leaving the fine-grained label and text span incomplete, a typical failure mode on cross-modal compound forgeries. In contrast, EFR identifies both manipulation types, localizes the altered face region and the tampered token ("bad", token 9) simultaneously, and grounds each conclusion in modality-specific evidence: a visual-evidence block that ties the face anchor to a concrete attribute deviation, and a text-evidence block that ties the token anchor to the sentiment shift it induces. Crucially, every cited box and span matches the coordinates declared in the conclusion, illustrating the coordinate-level anchor-evidence consistency that EFR enforces.

5 Conclusion

In this work, we presented EFR, an evidence-grounded forensic reasoning framework for multi-modal manipulation detection. Existing DGM4 methods offer no rationale for their predictions, while MLLM-based approaches generate explanations that remain disconnected from the predicted locations and suffer credit misassignment when a uniform signal optimizes heterogeneous heads. EFR resolves both with two coordinated components. The Anchor-and-Verify reasoning chain treats conclusion coordinates as explicit spatial anchors that downstream evidence must match, enforced by a five-component verifiable reward system during training. The Modality-Decoupled Advantage routing mechanism then delivers only task-relevant signals to each prediction head. Trained on a curated 50K forensic reasoning dataset, EFR attains state-of-the-art detection and competitive grounding on DGM4, producing reasoning records whose cited evidence is spatially consistent with the predicted anchors. We hope this perspective can inspire future work on interpretable multi-modal forensic reasoning.

Acknowledgments

This work was supported in part by the New Generation Artificial Intelligence-National Science and Technology Major Project (No. 2025ZD0123501), Chinese National Natural Science Foundation Projects U23B2054, 62276254.

References

[1] Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, et al. 2025. Qwen3-vl technical report. *arXiv preprint arXiv:2511.21631* (2025).

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

[3] Daniele Cardullo, Simone Teglia, and Irene Amerini. 2025. LADLE-MM: Limited Annotation based Detector with Learned Ensembles for Multimodal Misinformation. *arXiv preprint arXiv:2512.20257* (2025).

[4] Tao Chen, Jingyi Zhang, Decheng Liu, and Chunlei Peng. 2025. Mgffd-vlm: Multi-granularity prompt learning for face forgery detection with vlm. *arXiv preprint arXiv:2507.12232* (2025).

[5] Yueying Gao, Dongliang Chang, Bingyao Yu, Haotian Qin, Lei Chen, Kongming Liang, and Zhanyu Ma. 2025. Fakereasoning: Towards generalizable forgery detection and reasoning. *arXiv e-prints* (2025), arXiv–2503.

[6] Olga Golovneva, Moya Peng Chen, Spencer Poff, Martin Corredor, Luke Zettlemoyer, Maryam Fazel-Zarandi, and Asli Celikyilmaz. 2023. ROSCOE: A Suite of Metrics for Scoring Step-by-Step Reasoning. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

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

[8] Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. DeBERTa: Decoding-Enhanced BERT with Disentangled Attention. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

[9] Xinan He, Yue Zhou, Bing Fan, Bin Li, Guopu Zhu, and Feng Ding. 2025. Vlforgery face triad: Detection, localization and attribution via multimodal large language models. *arXiv preprint arXiv:2503.06142* (2025).

[10] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Liang Wang, Weizhu Chen, et al. 2022. Lora: Low-rank adaptation of large language models. *Iclr* 1, 2 (2022), 3.

[11] Zhenglin Huang, Jinwei Hu, Xiangtai Li, Yiwei He, Xingyu Zhao, Bei Peng, Baoyuan Wu, Xiaowei Huang, and Guangliang Cheng. 2025. Sida: Social media image deepfake detection, localization and explanation with large multimodal model. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 28831–28841.

[12] Shan Jia, Reilin Lyu, Kangran Zhao, Yize Chen, Zhiyuan Yan, Yan Ju, Chuanbo Hu, Xin Li, Baoyuan Wu, and Siwei Lyu. 2024. Can chatgpt detect deepfakes? a study of using multimodal large language models for media forensics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4324–4333.

[13] Jian-Yu Jiang-Lin, Kang-Yang Huang, et al. 2026. TriDF: Evaluating Perception, Detection, and Hallucination for Interpretable DeepFake Detection. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2026).

[14] Rohit Kundu, Shan Jia, Vishal Mohanty, Athula Balachandran, and Amit K Roy-Chowdhury. 2025. TruthLens: Visual Grounding for Universal DeepFake Reasoning. *arXiv preprint arXiv:2503.15867* (2025).

[15] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016. A Diversity-Promoting Objective Function for Neural Conversation Models. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*.

[16] Qilei Li, Mingliang Gao, Guisheng Zhang, Wenzhe Zhai, Jinyong Chen, and Gwanggil Jeon. 2024. Towards multimodal disinformation detection by vision-language knowledge interaction. *Information Fusion* 102 (2024), 102037.

[17] Tianxiao Li, Zhenglin Huang, Haiquan Wen, Yiwei He, Shuchang Lyu, Baoyuan Wu, and Guangliang Cheng. 2025. RAIDX: A Retrieval-Augmented Generation and GRPO Reinforcement Learning Framework for Explainable Deepfake Detection. In *Proceedings of the 33rd ACM International Conference on Multimedia*. 11746–11755.

[18] Yiheng Li, Yang Yang, Zichang Tan, Huan Liu, Weihua Chen, Xu Zhou, and Zhen Lei. 2025. Unleashing the Potential of Consistency Learning for Detecting and Grounding Multi-Modal Media Manipulation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 9242–9252.

[19] Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out: Proceedings of the ACL Workshop*.

[20] Kaiqing Lin, Zhiyuan Yan, Ruoxin Chen, Junyan Ye, Ke-Yue Zhang, Yue Zhou, Peng Jin, Bin Li, Taiping Yao, and Shouhong Ding. 2025. Seeing before reasoning: A unified framework for generalizable and explainable fake image detection. *arXiv preprint arXiv:2509.25502* (2025).

[21] Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. 2024. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 26296–26306.

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

[23] Huan Liu, Zichang Tan, Qiang Chen, Yunchao Wei, Yao Zhao, and Jingdong Wang. 2025. Unified frequency-assisted transformer framework for detecting and grounding multi-modal manipulation. *International Journal of Computer Vision* 133, 3 (2025), 1392–1409.

[24] Jiawei Liu, Fanrui Zhang, Jiaying Zhu, Esther Sun, Qiang Zhang, and Zheng-Jun Zha. 2024. Forgerygpt: Multimodal large language model for explainable image forgery detection and localization. *arXiv preprint arXiv:2410.10238* (2024).

[25] Runxin Liu, Tian Xie, Jiaming Li, Lingyun Yu, and Hongtao Xie. 2025. IDseq: Decoupled and Sequentially Detecting and Grounding Multi-Modal Media Manipulation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 496–504.

[26] Shih-Yang Liu, Xin Dong, Ximing Lu, Shizhe Diao, Peter Belcak, Mingjie Liu, Min-Hung Chen, Hongxu Yin, Yu-Chiang Frank Wang, Kwang-Ting Cheng, et al. 2026. Gdpo: Group reward-decoupled normalization policy optimization for multi-reward rl optimization. *arXiv preprint arXiv:2601.05242* (2026).

[27] Qiong Nan, Juan Cao, Yongchun Zhu, Yanyan Wang, and Jintao Li. 2021. MDFEND: Multi-domain fake news detection. In *Proceedings of the 30th ACM international conference on information & knowledge management*. 3343–3347.

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

[29] Gan Pei, Jiangning Zhang, Menghan Hu, Zhenyu Zhang, Chengjie Wang, Yunsheng Wu, Guangtao Zhai, Jian Yang, and Dacheng Tao. 2024. Deepfake generation and detection: A benchmark and survey. *Comput. Surveys* (2024).

[30] Lixiong Qin, Yang Zhang, Mei Wang, Jiani Hu, Weihong Deng, and Weiran Xu. 2025. Fake-in-Facext: Towards Fine-Grained Explainable DeepFake Analysis. *arXiv preprint arXiv:2510.20531* (2025).

[31] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems* 36 (2023), 53728–53741.

[32] Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. 2018. Faceforensics: A large-scale video dataset for forgery detection in human faces. *arXiv preprint arXiv:1803.09179* (2018).

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

[34] Rui Shao, Tianxing Wu, and Ziwei Liu. 2023. Detecting and grounding multi-modal media manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6904–6913.

[35] Rui Shao, Tianxing Wu, Jianlong Wu, Liqiang Nie, and Ziwei Liu. 2024. Detecting and grounding multi-modal media manipulation and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 8 (2024), 5556–5574.

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

[37] Jinjie Shen, Yaxiong Wang, Lechao Cheng, Nan Pu, and Zhun Zhong. 2025. Beyond Artificial Misalignment: Detecting and Grounding Semantic-Coordinated Multimodal Manipulations. In *Proceedings of the 33rd ACM International Conference on Multimedia*. 11308–11317.

[38] Zhuokang Shen, Kaisen Zhang, Bohan Jia, Heming Jia, Yuan Fang, Zhou Yu, and Shaohui Lin. 2025. DF-LLaVA: Unlocking MLLMs for Synthetic Image Detection via Knowledge Injection and Conflict-Driven Self-Reflection. *arXiv preprint arXiv:2509.14957* (2025).

[39] Gagandeep Singh, Samudi Amarsinghe, Priyanka Singh, and Xue Li. 2025. DGM4+: Dataset Extension for Global Scene Inconsistency. *arXiv preprint arXiv:2509.26047* (2025).

[40] Hao Tan, Jun Lan, Zichang Tan, Ajian Liu, Chuanbiao Song, Senyuan Shi, Huijia Zhu, Weiqiang Wang, Jun Wan, and Zhen Lei. 2025. Veritas: Generalizable deepfake detection via pattern-aware reasoning. *arXiv preprint arXiv:2508.21048* (2025).

[41] Jiazhen Wang, Bin Liu, Changtao Miao, Zhiwei Zhao, Wanyi Zhuang, Qi Chu, and Nenghai Yu. 2024. Exploiting modality-specific features for multi-modal manipulation detection and grounding. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4935–4939.

[42] Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, et al. 2024. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *arXiv preprint*

arXiv:2409.12191 (2024).

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

[44] Wenbo Xu, Wei Lu, Xiangyang Luo, and Jiantao Zhou. 2026. MARE: Multimodal Alignment and Reinforcement for Explainable Deepfake Detection via Vision-Language Models. *arXiv preprint arXiv:2601.20433* (2026).

[45] Zhipei Xu, Xuanyu Zhang, Runyi Li, Zecheng Tang, Qing Huang, and Jian Zhang. 2024. Fakeshield: Explainable image forgery detection and localization via multi-modal large language models. *arXiv preprint arXiv:2410.02761* (2024).

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

[47] Xinquan Yu, Wei Lu, and Xiangyang Luo. 2025. Fine-grained Multiple Supervisory Network for Multi-modal Manipulation Detecting and Grounding. *arXiv preprint arXiv:2508.02479* (2025).

[48] Xinquan Yu, Wei Lu, Xiangyang Luo, and Rui Yang. 2026. CIEC: Coupling Implicit and Explicit Cues for Multimodal Weakly Supervised Manipulation Localization. *arXiv preprint arXiv:2602.02175* (2026).

[49] Xiaojiang Zhang, Jinghui Wang, Zifei Cheng, Wenhao Zhuang, Zheng Lin, Minglei Zhang, Shaojie Wang, Yinghan Cui, Chao Wang, Junyi Peng, et al. 2025. Srpo: A cross-domain implementation of large-scale reinforcement learning on llm. *arXiv preprint arXiv:2504.14286* (2025).

[50] Yuchen Zhang, Yaxiong Wang, Yujiao Wu, Lianwei Wu, Li Zhu, and Zhedong Zheng. 2025. The Coherence Trap: When MLLM-Crafted Narratives Exploit Manipulated Visual Contexts. *arXiv preprint arXiv:2505.17476* (2025).

[51] Zhenxing Zhang, Yaxiong Wang, Lechao Cheng, Zhun Zhong, Dan Guo, and Meng Wang. 2025. Asap: Advancing semantic alignment promotes multi-modal manipulation detecting and grounding. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 4005–4014.

[52] Hanqing Zhao, Wenbo Zhou, Dongdong Chen, Tianyi Wei, Weiming Zhang, and Nenghai Yu. 2021. Multi-attentional deepfake detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2185–2194.

[53] Chujie Zheng, Shixuan Liu, Mingze Li, Xiong-Hui Chen, Bowen Yu, Chang Gao, Kai Dang, Yuqiong Liu, Rui Men, An Yang, et al. 2025. Group sequence policy optimization. *arXiv preprint arXiv:2507.18071* (2025).

[54] Ziyin Zhou, Yunpeng Luo, Yuanchen Wu, Ke Sun, Jiayi Ji, Ke Yan, Shouhong Ding, Xiaoshuai Sun, Yunsheng Wu, and Rongrong Ji. 2025. Aigi-holmes: Towards explainable and generalizable ai-generated image detection via multimodal large language models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 18746–18758.

[55] Tinghui Zhu, Qin Liu, Fei Wang, Zhengzhong Tu, and Muhao Chen. 2024. Unraveling cross-modality knowledge conflicts in large vision-language models. *arXiv preprint arXiv:2410.03659* (2024).

[56] Zhihao Zhu, Jiafeng Liang, Shixin Jiang, Jinlan Fu, Ming Liu, Guanglu Sun, See-Kiong Ng, and Bing Qin. 2026. Analyzing Reasoning Consistency in Large Multimodal Models under Cross-Modal Conflicts. *arXiv preprint arXiv:2601.04073* (2026).