LEARNING TO VERIFY: A COUNTERFACTUAL-GUIDED DUAL-LOOP AGENT FRAMEWORK FOR MULTIMODAL HALLUCINATION MITIGATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Mitigating multimodal hallucination is essential for the reliable utilization of multimodal large language models in image-grounded tasks. Existing approaches perform mitigation by adjusting decoding, strengthening visual information during generation, or using verification to guide response correction. However, verification itself may accept unsupported statements or challenge correct ones, allowing repeated revision to retain hallucinations and remove valid information. In this paper, we propose Learning to Verify (L2V), a dual-loop agent framework for multimodal hallucination mitigation that learns a verifier's criteria through counterfactual analysis of its judgments. The inner loop performs inference by checking a main agent's response against the image and guiding revision until the response is accepted or an uncertainty response is returned. The outer loop operates only during training, using reference responses and counterfactual interventions to identify verification errors whose correction improves the final response. The identified verification errors guide targeted revisions to the verifier's criteria, which are evaluated through held-out replay on examples excluded from criteria refinement and adopted only when they improve response quality over the current criteria. Experiment results and analysis on multimodal hallucination benchmarks show that the proposed L2V outperforms strong baselines and existing approaches, demonstrating its effectiveness in mitigating multimodal hallucination.[1]

1 INTRODUCTION

Multimodal hallucination refers to cases where generated content is not grounded in the information provided by the input modalities (Li et al., 2023). As multimodal large language models (MLLMs) advance in capability and scope of application, mitigating hallucination becomes increasingly important for their reliable use (Sahoo et al., 2024). Compositional evaluations also expose weaknesses in recognizing attributes, relations, and word order in vision–language models (Yuksekgonul et al., 2022). For image-grounded tasks, the challenge lies in reducing unsupported content while preserving visually supported information and expressing uncertainty when visual evidence is insufficient.

Existing approaches mitigate MLLM hallucination by adjusting decoding (Leng et al., 2024; Huang et al., 2024b), strengthening visual information during generation (Zou et al., 2024), or correcting completed responses (Zhou et al., 2024; Yin et al., 2024). Agentic workflows provide a complementary perspective by organizing generation, feedback, and revision around a large language model (LLM) (Madaan et al., 2023; Shinn et al., 2023). The instructions and control mechanisms governing these operations form an agent harness, within which verification provides feedback for response revision. For multimodal hallucination mitigation, verification assesses whether generated statements are supported by the image and identifies content requiring correction (Yin et al., 2024; Wu et al., 2026). However, introducing verification does not ensure reliable correction, because the verifier itself may accept unsupported statements or challenge correct ones. Repeating this process without improving its verification criteria risks retaining hallucinations and removing valid information. Studies of multimodal dialogue show that earlier hallucinations can mislead later responses (Zhong et al., 2024), while analyses of language-model reasoning find that self-correction without external

[1] Code is available at `https://anonymous.4open.science/r/L2V-4244`.

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

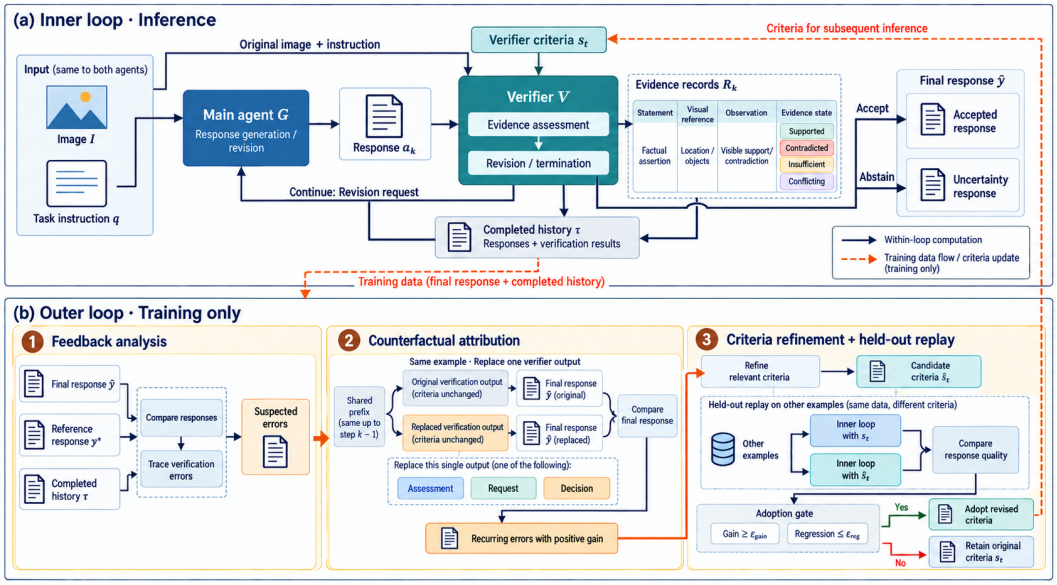

Figure 1: The overall architecture of L2V, which consists of two loops. (a) shows the inner loop that verifies and revises responses using image evidence. (b) shows the outer loop that identifies verification errors, tests alternative verifier outputs through counterfactual interventions, and evaluates revised criteria through held-out replay. Counterfactual attribution uses the same example with unchanged criteria, whereas replay compares original and candidate criteria on other examples.

feedback can fail to improve accuracy (Huang et al., 2024a). This limitation motivates learning the verifier's criteria from the consequences of its judgments, identifying which corrections improve final responses and whether the resulting criteria benefit other examples.

In this paper, we propose Learning to Verify (L2V), a dual-loop agent framework that uses counterfactual attribution to guide the refinement of visual verification criteria. An inner loop performs inference by coupling a main agent's response generation with a verifier's evidence assessment, revision requests, and termination decisions. An outer loop operates only during training and compares completed responses with reference responses to locate potential verification errors. To refine the verifier's criteria, we propose counterfactual interventions to identify verification errors whose correction improves the final response. These findings guide targeted criteria revisions, which undergo held-out replay to assess their effectiveness on other examples. The inner loop then uses the improved criteria to guide response verification and revision on new inputs, aiming to reduce unsupported content while preserving visually supported information. We evaluate L2V on English multimodal hallucination benchmarks where L2V achieves better performance than strong baselines and existing studies, demonstrating its effectiveness for multimodal hallucination mitigation.

2 The Approach

L2V is a dual-loop agent framework for multimodal hallucination mitigation, with an inner loop for inference and an outer loop for optimization used only during training. L2V comprises two multimodal agents, namely, a main agent G for response generation and a verifier V for response verification. The overall architecture of our approach is presented in Figure 1. Given an image I and a task instruction q, the inner loop uses V to assess the visual support for the response generated by G and guide response revision, producing the final response $\widehat{y}$. The outer loop uses reference responses and counterfactual interventions to identify verification errors that warrant criteria refinement, then evaluates the revised criteria through held-out replay for use in subsequent inner-loop inference. The following texts describe the inner and outer loops in detail.

2.1 Inner Loop for Verifier-Guided Inference

The inner loop mitigates multimodal hallucination through response generation, evidence assessment, and response revision. Given the image I and the task instruction q, the main agent G gener-

ates an initial response. The verifier V assesses whether the image supports the factual statements in the response and identifies statements with insufficient visual support. Based on this assessment, V challenges unsupported statements and guides G to revise the response, repeating verification and revision until V accepts the response or determines that further revision does not resolve the insufficient visual support. The inner loop returns the accepted response or an uncertainty response as the final response $\widehat{y}$.

Response Generation. At inner-loop step $k \geq 1$, let h_{k-1} denote the history of responses generated by G and verification results produced by V in the preceding steps. The history is initially empty, with $h_0 = \varnothing$. The main agent G generates the current response a_k from the image I, the task instruction q, and the history h_{k-1}:

$$a_k = G(I, q, h_{k-1}) \tag{1}$$

For $k > 1$, G uses the verification results in h_{k-1} to reassess the visual support for challenged statements, revise unsupported statements, and retain supported statements. The main agent returns a complete revised response so that V checks all factual statements, including those introduced during revision.

Evidence Assessment. The verifier V decomposes the current response a_k into factual statements $\mathcal{C}_k = \{c_{k,j}\}_{j=1}^{J_k}$, where J_k denotes the number of statements and j indexes them. Each statement expresses an independently assessable assertion about an object, attribute, relation, or quantity in the image. For each statement $c_{k,j}$, V produces a visual reference $\ell_{k,j}$ identifying the relevant location or objects in I, an observation $o_{k,j}$ describing the visible information that supports or contradicts the statement, and an evidence state $r_{k,j}$ indicating whether that information supports the statement. The evidence assessment operation V^{assess} computes these quantities from the image, task instruction, statement, and preceding history:

$$(\ell_{k,j}, o_{k,j}, r_{k,j}) = V^{\text{assess}}(I, q, c_{k,j}, h_{k-1}) \tag{2}$$

The verifier applies its evidence assessment criteria to assign one of four evidence states: Sup for sufficient support, Con for contradiction, Ins for insufficient evidence, and Mix for unresolved conflicting observations. When relevant visual information is unavailable or ambiguous, $o_{k,j}$ describes what information remains missing or unclear. The evidence assessment criteria specify the visual information required to support a statement at its stated level of detail. For example, the presence of a person and a bicycle supports statements about their existence, while a statement that the person is riding the bicycle requires visual support for the riding relation. We collect the statement, visual reference, observation, and evidence state into an evidence record $(c_{k,j}, \ell_{k,j}, o_{k,j}, r_{k,j})$ and denote the set of evidence records at step k by $\mathcal{R}_k$. The verifier uses $\mathcal{R}_k$ to identify statements requiring revision and determine whether the response has sufficient visual support. Across steps, V uses the history h_{k-1} to track repeated descriptions of the same image content as a single source of visual support.

Response Revision and Termination. The verifier V uses the evidence records $\mathcal{R}_k$ to determine whether to accept the current response, request further revision, or terminate the inner loop with an uncertainty response. Let $K_{\max}$ denote the maximum number of responses generated by G for the current input, and let $b_k = K_{\max} - k$ denote the number of remaining revision steps. The revision and termination operation V^{control} produces a decision d_k and a revision request ι_{k+1} specifying what G should reconsider in the next step:

$$(d_k, \iota_{k+1}) = V^{\text{control}}(I, q, \mathcal{R}_k, h_{k-1}, b_k) \tag{3}$$

The decision d_k takes one of three values, namely, Accept, Continue, or Abstain. The verifier selects Accept when the image sufficiently supports the factual statements in the response and the response satisfies the task instruction q. The verifier selects Continue when $b_k > 0$ and further revision is needed to address insufficient visual support. For Continue, the revision request ι_{k+1} identifies the challenged statements, the observations motivating the challenge, and the image content that G should reexamine. The verifier selects Abstain when the response still lacks sufficient visual support and either $b_k = 0$ or V determines that further revision does not resolve the insufficient support. For Accept and Abstain, the revision request is empty.

After step k, L2V appends $(a_k, \mathcal{R}_k, d_k, \iota_{k+1})$ to h_{k-1} to obtain h_k. For Continue, G uses the updated history h_k to generate the next response according to Equation 1. Let $K \leq K_{\max}$ denote the final inner-loop step and $\tau = h_K$ denote the completed history, which records the responses and verification results for subsequent outer-loop analysis. If $d_K = \mathsf{Accept}$, the inner loop returns the accepted response a_K as the final response $\widehat{y}$. If $d_K = \mathsf{Abstain}$, the inner loop returns an uncertainty response as $\widehat{y}$, indicating what the image does not sufficiently support. The final evidence records $\mathcal{R}_K$ accompany $\widehat{y}$ and state the visual basis for acceptance or the insufficient support that leads to abstention.

2.2 Outer Loop for Verifier Optimization

The outer loop optimizes the verifier's criteria during training through feedback analysis, counterfactual failure attribution, and criteria refinement with held-out replay. L2V first compares the final response $\widehat{y}$ with a reference response $y^\star$ to obtain feedback and examines the recorded responses and verification results to locate potential verification errors. Counterfactual failure attribution assesses whether correcting these errors improves the final response, identifying targets for criteria refinement. L2V then revises the relevant criteria and uses held-out replay to compare the original and revised criteria on other examples. L2V adopts only revisions that improve response quality and applies the revised criteria in subsequent inner-loop inference to mitigate hallucination.

Feedback Analysis. Let s_t denote the verifier's criteria at outer-loop optimization round t. We construct a feedback batch $\mathcal{B}_t$ containing records $(I, q, \widehat{y}, y^\star, \tau)$, each comprising an image, a task instruction, a final response, a reference response, and the completed inner-loop history. L2V compares $\widehat{y}$ with $y^\star$ to obtain feedback and traces the discrepant statements through the responses, evidence records, revision requests, and decisions recorded in τ. The verifier V analyzes this feedback to identify potential errors, including acceptance of unsupported statements, unnecessary challenges to supported statements, ineffective revision requests, premature termination, and unnecessary continuation. For each potential error, V produces a discrepancy record identifying the affected statement, the supporting information in I or $y^\star$, the relevant inner-loop step, and the suspected verification error. The feedback analysis operation V^{analyze} collects these discrepancy records into $\mathcal{D}_t$:

$$\mathcal{D}_t = \bigcup_{(I,q,\widehat{y},y^\star,\tau)\in\mathcal{B}_t} V^{\text{analyze}}(I, q, \widehat{y}, y^\star, \tau; s_t) \tag{4}$$

The verifier uses both the image and the reference response to assess each discrepancy, since omission from the reference response alone does not establish that a statement is false. The verifier retains discrepancy records with supporting evidence in the image or reference response for subsequent counterfactual failure attribution.

Counterfactual Failure Attribution. Counterfactual failure attribution tests whether correcting a suspected verification error improves the final response. For each discrepancy record in $\mathcal{D}_t$, L2V selects a verifier output at the identified step k, namely, an evidence state $r_{k,j}$, a revision request ι_{k+1}, or a decision d_k. Let m denote the type of output selected for replacement, with

$$\mathcal{M} = \{\mathsf{Assess}, \mathsf{Revise}, \mathsf{Stop}\} \tag{5}$$

Here, Assess, Revise, and Stop refer to the evidence state, revision request, and decision, respectively, and $m \in \mathcal{M}$. The verifier V proposes an alternative output using I, q, the preceding history h_{k-1}, the current response a_k, and any verification results preceding the selected output at step k. L2V replaces the selected output, recomputes the verification results that depend on it, and continues the inner loop under the original criteria s_t. The counterfactual intervention preserves a_k and the preceding history and retains the original limit $K_{\max}$ on the number of responses. L2V reserves the reference response $y^\star$ for feedback analysis and outcome evaluation, excluding it from the inputs used to generate the alternative output and subsequent responses. The counterfactual continuation produces a final response $\widehat{y}^{(k,m)}$ and a completed history $\tau^{(k,m)}$.

For the same image I, task instruction q, and reference response $y^\star$, we evaluate the original and counterfactual outcomes using an evaluation function $U(\widehat{y}, \tau)$. The evaluation function accounts for task correctness, unsupported statements, retention of supported statements, and the number of inner-loop steps. The counterfactual gain $\Delta_{k,m}$ measures the change in evaluation score caused by

replacing the selected verifier output:

$$\Delta_{k,m} = U\left(\widehat{y}^{(k,m)}, \tau^{(k,m)}\right) - U(\widehat{y}, \tau) \tag{6}$$

A positive counterfactual gain indicates that the replacement improves the evaluated outcome for the same input. Within the feedback batch $\mathcal{B}_t$, L2V groups the original and counterfactual outcomes by the verifier criterion associated with the suspected error. Each group represents a potential recurring verification error. L2V retains a group for criteria refinement when its mean counterfactual gain exceeds a positive threshold ϵ_{attr} across multiple independent inputs. Each retained group links the relevant criterion in s_t to the original verifier outputs, their replacements, and the resulting changes in final responses.

Criteria Refinement with Held-out Replay. Using the retained groups of verification errors, V revises the relevant criteria in s_t to produce candidate criteria $\widetilde{s}_t$. Each revision specifies how V should assess visual support, challenge statements, or terminate the inner loop in cases associated with the recurring error, while preserving unrelated criteria. L2V associates the candidate criteria with the original and counterfactual outcomes supporting the revision.

Let $\mathcal{H}_t$ denote a replay set of examples $(I, q, y^\star)$ whose images are excluded from the feedback batch $\mathcal{B}_t$ used to construct the candidate criteria. For each example, L2V runs the inner loop under s_t and $\widetilde{s}_t$, using the same initial response a_1, response limit $K_{\max}$, decoding settings, and random seeds. For any verifier criteria s, let $\widehat{y}(s)$ and $\tau(s)$ denote the final response and completed history produced for the current image I and task instruction q. L2V uses the corresponding reference response $y^\star$ exclusively for outcome evaluation through the evaluation function U. The replay objective $J(s; \mathcal{H}_t)$ averages the evaluation scores across the replay set:

$$J(s; \mathcal{H}_t) = \frac{1}{|\mathcal{H}_t|} \sum_{(I,q,y^\star)\in\mathcal{H}_t} U(\widehat{y}(s), \tau(s)) \tag{7}$$

Each summand evaluates the final response and completed history for the corresponding image, task instruction, and reference response. Held-out replay assesses whether the candidate criteria improve response quality on examples beyond those used for criteria refinement.

To identify decreases in performance that the average score might conceal, L2V partitions $\mathcal{H}_t$ into predefined task groups $\mathcal{H}_{t,g}$, where g indexes the groups. Using $[z]_+ = \max(z, 0)$, we define the maximum replay regression R_t as the largest decrease in the replay objective across task groups:

$$R_t = \max_g \left[J(s_t; \mathcal{H}_{t,g}) - J(\widetilde{s}_t; \mathcal{H}_{t,g})\right]_+ \tag{8}$$

The replay gain $\Delta J_t = J(\widetilde{s}_t; \mathcal{H}_t) - J(s_t; \mathcal{H}_t)$ measures the improvement in the average evaluation score. Let $\epsilon_{\text{gain}} > 0$ denote the minimum required replay gain and $\epsilon_{\text{reg}} \geq 0$ denote the maximum permitted replay regression. L2V adopts the candidate criteria only when both requirements hold:

$$s_{t+1} = \begin{cases} \widetilde{s}_t & \text{if } \Delta J_t \geq \epsilon_{\text{gain}} \ \wedge \ R_t \leq \epsilon_{\text{reg}} \\ s_t & \text{otherwise} \end{cases} \tag{9}$$

After T outer-loop optimization rounds, L2V uses the final criteria s_T in subsequent inner-loop inference. Final test examples remain disjoint from all feedback batches and replay sets.

3 Experiment Setting

3.1 Datasets

We evaluate hallucination mitigation on two English benchmarks, POPE (Li et al., 2023) and CHAIR (Rohrbach et al., 2018), for object-existence questions and image captions, respectively. POPE tests object existence with random, popular, and adversarial negative-object sampling on COCO (Lin et al., 2014) and GQA (Hudson & Manning, 2019) images. We use 3,024 questions per sampling protocol for POPE-COCO and 7,675 for POPE-GQA, each covering 500 images. For CHAIR, we evaluate object hallucination in one generated caption for each of 1,000 COCO images.

For outer-loop training, we use image–instruction pairs and reference responses from LLaVA-Instruct-150K (Liu et al., 2023). We select 1,000 images disjoint from the evaluation sets and divide them into 800 feedback images for criteria refinement and 200 held-out replay images for evaluating candidate criteria.

3.2 Baselines and Comparing Models

We compare our approach with the following baselines to examine the contributions of response revision, verification, and criteria refinement. "**Base**" generates a single response directly from the image and task instruction. "**+ Self-Refine**" (Madaan et al., 2023) uses the main agent to critique its initial response and revise it once. "**+ Single-pass Verifier**" uses a separate verifier to assess the initial response and request at most one revision. "**+ Inner Loop**" runs the full inner loop with the initial verifier criteria s_0. "**+ Inner + Outer (L2V)**" is our full model uses the outer loop to learn verifier criteria s_T and applies them in inner-loop inference. All baselines and our approach use the same settings (e.g., task-specific input prompts).

We also compare L2V with existing models to multimodal hallucination mitigation. **VCD** (Leng et al., 2024) contrasts predictions from original and distorted images to reduce reliance on language priors. **HALC** (Chen et al., 2024) combines adaptive selection of visual regions with contrastive decoding and beam search to correct hallucinated tokens. **LURE** (Zhou et al., 2024) uses a learned revisor to correct object hallucinations in generated descriptions. **DeCo** (Wang et al., 2025a) incorporates predictions from selected preceding layers into the final output logits to recover visual information during decoding. **AdaVIB** (Bai et al., 2025) applies an adaptive variational information bottleneck to suppress irrelevant visual information. **CausalMM** (Zhou et al., 2025) uses counterfactual reasoning over visual and language attention to reduce hallucinations from modality priors.

3.3 Implementation Details

We evaluate L2V with four MLLM backbones: Qwen3.5-2B, Qwen3.5-4B, Qwen3.5-35B-A3B (Team, 2026), and Ministral-3-3B (Liu et al., 2026a). Qwen3.5-2B and Qwen3.5-4B use dense LLM backbones with 24 and 32 layers and hidden dimensions of 2,048 and 2,560, respectively. Qwen3.5-35B-A3B has 40 Transformer layers with a hidden dimension of 2,048 and a mixture-of-experts (MoE) architecture comprising 256 routed experts per layer, of which 8 are selected for each token, alongside one shared expert; each expert has an intermediate dimension of 512. Ministral-3-3B uses a 26-layer language backbone with a hidden dimension of 3,072 and a Pixtral vision encoder. For each backbone, the main agent G and verifier V use the same checkpoint with separate contexts and role-specific instructions, and both receive the original image. We use Qwen3.5-2B as the primary configuration for comparisons with existing approaches and ablation studies. The Qwen3.5 models are used with thinking disabled.

All experiments are conducted on NVIDIA A40 GPUs. During outer-loop training, we update only the verifier's criteria, keeping the parameters of both agents and the main-agent instructions fixed. Counterfactual comparisons and held-out replay use matched decoding settings, random seeds, and response limits within each pair of runs. The trajectory utility scoring rules and criteria acceptance thresholds remain fixed throughout optimization. During inference, the inner loop allows at most $K_{\max} = 5$ main-agent responses per input. Response generation uses beam search with a beam size of 5 and maximum output lengths of 4 new tokens for POPE and 64 for CHAIR.

For evaluation, we use different metraics for vairous datasets. For POPE, we report accuracy (Acc) and positive-class F1 for the random, popular, and adversarial sampling settings of POPE-COCO and POPE-GQA, with higher values indicating better performance. For CHAIR dataset, we use CHAIR_S and CHAIR_I to measure the proportions of captions and object mentions containing hallucinated objects, respectively, with lower values indicating fewer hallucinations.

4 Results and Analysis

4.1 Overall Results

Table 1 reports the results of comparing baselines and our approach across four backbones on POPE and CHAIR. There are several observations. First, single-pass verification consistently outperforms Self-Refine on both object-existence prediction and caption hallucination. This comparison supports assigning visual evidence assessment to a separate verifier rather than relying solely on the main agent to critique its own response. Second, the inner loop improves all reported metrics over single-pass verification. Reassessing revised responses provides an opportunity to address unsupported content that remains after the first correction. Third, full L2V improves on the inner loop with initial

Table 1: Comparison of our approach (i.e., L2V) and baselines across different MLLM backbones on POPE-COCO, POPE-GQA, and CHAIR.

Approach	POPE-COCO						POPE-GQA						CHAIR	
	Random		Popular		Adversarial		Random		Popular		Adversarial			
	Acc↑	F1↑	Acc↑	F1↑	Acc↑	F1↑	Acc↑	F1↑	Acc↑	F1↑	Acc↑	F1↑	CH_S ↓	CH_I ↓
Qwen3.5-2B	82.2	84.1	81.5	83.4	75.7	80.1	86.2	89.6	82.1	84.7	78.4	85.5	13.8	11.2
+ Self-Refine	85.2	85.7	85.7	85.0	80.6	82.7	87.8	90.6	84.0	86.6	80.0	86.5	12.0	10.9
+ Single-pass Verifier	88.5	88.6	86.8	86.9	82.6	84.2	89.2	92.5	84.9	89.5	80.9	87.3	10.8	10.1
+ Inner Loop	91.7	91.8	88.6	88.6	86.7	86.5	91.0	93.0	86.7	90.6	83.6	88.0	9.6	9.6
+ Inner + Outer (L2V)	**93.9**	**94.8**	**90.7**	**90.1**	**88.1**	**87.7**	**92.2**	**94.8**	**88.7**	**92.9**	**84.7**	**89.2**	**8.2**	**9.0**
Qwen3.5-4B	84.8	86.3	83.7	84.8	80.8	82.5	88.3	91.5	80.4	86.9	81.5	87.8	13.1	11.0
+ Self-Refine	87.0	87.8	85.9	86.6	84.2	83.8	89.4	92.2	83.4	88.7	82.7	88.2	10.6	10.3
+ Single-pass Verifier	90.5	90.9	88.3	88.5	86.6	85.8	91.1	93.6	85.6	89.7	83.7	89.0	8.6	9.1
+ Inner Loop	91.3	92.3	91.5	91.7	87.3	87.9	92.2	94.3	87.2	91.2	84.1	89.3	7.1	7.6
+ Inner + Outer (L2V)	**94.6**	**95.3**	**94.5**	**93.7**	**90.8**	**90.3**	**93.1**	**95.5**	**89.6**	**92.7**	**85.3**	**89.8**	**5.1**	**6.6**
Qwen3.5-35B-A3B	88.3	87.9	86.2	86.0	84.3	84.7	88.8	92.0	81.9	87.8	79.3	86.7	12.7	10.9
+ Self-Refine	90.7	89.4	88.1	88.7	85.6	86.3	90.2	92.9	83.6	89.6	80.3	88.0	10.6	9.3
+ Single-pass Verifier	91.6	92.2	90.8	90.4	88.5	87.9	90.7	94.4	86.8	90.6	82.7	88.6	8.4	8.4
+ Inner Loop	92.9	94.0	93.0	91.9	89.5	89.6	92.0	95.2	88.9	91.9	84.9	90.4	6.0	7.0
+ Inner + Outer (L2V)	**95.3**	**95.8**	**94.9**	**93.8**	**91.4**	**90.9**	**93.4**	**96.3**	**90.1**	**93.1**	**85.9**	**91.1**	**4.6**	**6.2**
Ministral-3-3B	82.5	84.4	81.6	83.9	75.3	79.8	81.5	84.7	74.9	81.3	68.8	79.2	17.6	16.1
+ Self-Refine	84.6	85.3	83.1	85.0	76.7	81.2	82.7	86.2	76.0	83.1	72.0	80.6	15.2	14.8
+ Single-pass Verifier	87.0	86.9	85.4	86.8	79.6	82.3	83.8	87.0	77.2	83.9	73.9	82.9	12.9	12.5
+ Inner Loop	87.7	87.4	86.8	88.0	82.8	82.9	84.7	88.9	77.8	86.2	76.5	83.9	11.8	11.2
+ Inner + Outer (L2V)	**90.1**	**88.9**	**89.2**	**89.5**	**84.4**	**84.3**	**85.9**	**90.2**	**78.9**	**87.7**	**78.6**	**85.9**	**9.6**	**9.8**

criteria while retaining the same inference procedure, supporting the benefit of learning verification criteria beyond repeated verification. Fourth, L2V consistently improves all reported metrics over the corresponding base models across the evaluated Qwen3.5 and Ministral backbones. These gains span different model sizes and both dense and MoE architectures, indicating that the approach is effective across the evaluated backbone configurations.

Table 2 compares our approach with existing approaches under Qwen3.5-2B. L2V achieves the highest accuracy and F1 across all six POPE settings and the lowest CHAIR hallucination rates. The POPE gains hold across random, popular, and adversarial sampling on both COCO and GQA images. On CHAIR, the reductions at both the caption and object-mention levels show that the benefit also extends to responses containing multiple object descriptions. These results support the effectiveness of L2V across object-existence questions and caption generation.

4.2 Ablation Study

We examine how individual components contribute to L2V under Qwen3.5-2B. For the inner loop, we replace statement-level assessment with a response-level judgment, targeted revision requests with generic rechecking, or adaptive termination with revision to the response limit, keeping the learned criteria fixed. For the outer loop, we remove counterfactual attribution or held-out replay while retaining the initial criteria and feedback data. Table 3 reports POPE F1 and CHAIR hallucination rates for these variants.

Removing statement-level assessment causes the largest F1 decrease in most POPE settings and the largest increase in caption-level hallucination, supporting the value of locating unsupported content within a response. Removing targeted revision requests causes the largest increase in object-mention-level hallucination, highlighting the importance of directing correction toward specific statements. Replacing adaptive termination with revision to the response limit also degrades all metrics, indicating that additional revision alone does not ensure better responses.

Removing either counterfactual attribution or held-out replay worsens all reported metrics. Counterfactual attribution connects a suspected verification error to its effect on the final response, helping target criteria revisions at errors whose correction is beneficial. Held-out replay evaluates whether those revisions also help on examples excluded from their construction. The losses from removing either component support using both outcome-based attribution and evaluation on separate examples to refine the verifier.

Table 2: Comparison of L2V with existing studies using Qwen3.5-2B backbone on POPE-COCO, POPE-GQA, and CHAIR. Best results are shown in bold, and second-best results are underlined.

Approach	POPE-COCO						POPE-GQA						CHAIR	
	Random		Popular		Adversarial		Random		Popular		Adversarial			
	Acc↑	F1↑	Acc↑	F1↑	Acc↑	F1↑	Acc↑	F1↑	Acc↑	F1↑	Acc↑	F1↑	$CH_S \downarrow$	$CH_I \downarrow$
Qwen3.5-2B	82.2	84.1	81.5	83.4	75.7	80.1	86.2	89.6	82.1	84.7	78.4	85.5	13.8	11.2
VCD (Leng et al., 2024)	92.9	92.6	89.2	89.5	85.0	87.4	90.4	90.7	84.5	86.5	83.4	87.3	9.3	10.8
HALC (Chen et al., 2024)	92.9	92.6	89.1	88.9	86.5	86.4	89.4	90.9	83.4	88.3	82.3	86.4	13.2	9.8
LURE (Zhou et al., 2024)	92.7	92.2	90.1	88.8	86.6	84.6	89.9	91.9	83.7	88.8	82.7	87.3	13.0	10.1
DeCo (Wang et al., 2025a)	92.4	91.7	89.8	88.8	87.8	86.7	89.1	90.3	83.3	87.2	81.9	87.0	16.7	10.6
AdaVIB (Bai et al., 2025)	92.5	92.7	78.1	81.7	72.7	78.3	88.4	91.6	83.1	85.9	80.0	85.7	12.2	9.9
CausalMM (Zhou et al., 2025)	93.3	93.7	89.8	89.4	87.3	87.0	91.1	93.8	87.4	91.8	83.8	89.0	12.4	9.8
L2V (Ours)	**93.9**	**94.8**	**90.7**	**90.1**	**88.1**	**87.7**	**92.2**	**94.8**	**88.7**	**92.9**	**84.7**	**89.2**	**8.2**	**9.0**

Table 3: Ablation study of our approach under Qwen3.5-2B on POPE-COCO, POPE-GQA, and CHAIR. We only report the F1 scores of different settings on POPE-COCO and POPE-GQA. The performance of Qwen3.5-2B is also reported for reference.

Approach	POPE-COCO (F1↑)			POPE-GQA (F1↑)			CHAIR	
	Random	Popular	Adversarial	Random	Popular	Adversarial	$CH_S \downarrow$	$CH_I \downarrow$
L2V	**94.8**	**90.1**	**87.7**	**94.8**	**92.9**	**89.2**	**8.2**	**9.0**
w/o statement-level assessment	91.3	87.6	84.2	93.1	89.5	87.4	10.7	9.8
w/o targeted revision requests	93.2	88.7	85.1	92.6	91.8	88.3	9.6	10.1
w/o adaptive termination	94.1	89.5	86.8	94.3	91.4	88.7	9.3	9.4
w/o counterfactual attribution	92.4	88.1	85.8	93.5	89.9	87.8	10.2	9.7
w/o held-out replay	93.7	88.9	86.1	93.8	90.6	88.0	9.1	9.9
Qwen3.5-2B	84.1	83.4	80.1	89.6	84.7	85.5	13.8	11.2

4.3 Case Study

We examine how verification corrects object-category errors using two POPE examples in Figure 2. Each panel shows the input and reference answer, followed by the initial response, verifier feedback, and revised response. In (a), the main agent recognizes a truck but incorrectly answers "Yes" to a question about a car. The verifier distinguishes the two categories, leading to a negative answer that retains the truck description. In (b), the verifier uses the bag's visible straps and design to challenge the backpack label. The revised response rejects that label while retaining the bag's color and pattern. Both answers match the references after one revision. These cases illustrate how feedback directed at a specific unsupported statement corrects the answer.

5 Related Work

Multimodal hallucination research examines both visual grounding failures and interventions that improve factual responses (Sahoo et al., 2024). Evaluation spans general perception and cognition in MME (Fu et al., 2025), language hallucination and visual illusion in HallusionBench (Guan et al., 2024), recognition of absent objects in NOPE (Lovenia et al., 2024), and sensitivity to object removal (He et al., 2025); HalLoc provides token-level annotations for localizing hallucinations (Park et al., 2025). Generation-time approaches adjust token selection through contrastive predictions (Leng et al., 2024; Wang et al., 2024) or reinforce visual grounding through attention and representation interventions (Huang et al., 2024b; Zou et al., 2024; Jiang et al., 2025). Related interventions smooth abrupt changes between layer representations (Wang et al., 2025b), steer the visual key–value cache during prefill (Zhang et al., 2026), or inject context embeddings during decoding (Fazli et al., 2026). Training-based approaches augment reward models with factual information (Sun et al., 2024) or learn from human and model-generated preferences (Yu et al., 2024; 2025; Peng et al., 2025). Preference objectives further address noisy supervision (Zhang et al., 2025), rank images by object visibility (Zadeh et al., 2025), or regularize self-generated response triplets (Guo et al., 2026). Response-level approaches assess generated statements against visual evidence (Jing et al., 2024; Sahu et al., 2024) and use verification feedback to revise or resample

(a) Category correction



Task instruction
Is there a car in the image?
Reference answer *No*

Round 1 / Single Response
Main agent output G
*"**Yes**, there is a vintage truck in the image."*
Verifier output V *Continue*
"The vehicle is a vintage truck, not a car, but it is a motor vehicle."

Round 2 / Final response
Main agent output G
*"**No, there is no car** in the image. The image contains a vintage truck."*
Verifier output V *Accept*
"The vehicle is a truck, not a car."

(b) Bag category correction



Task instruction
Is there a backpack in the image?
Reference answer
No

Round 1 / Single Response
Main agent output G
*"**Yes, there is a backpack** in the image. It is orange with a zebra-print pattern and is being carried by the woman."*
Verifier output V (revision request) *Continue*
"Specify that the bag is a tote or duffel bag, not a backpack, based on its visible straps and design."

Round 2 / Final response
Main agent output G
*"**No, there is not a backpack** in the image. The bag is an orange tote or duffel bag with a zebra-print pattern, carried by the woman."*
Verifier output V (observation) *Accept*
"The bag has visible straps and a design consistent with a tote or duffel bag, not a backpack."

Figure 2: Case study of our approach with two examples. (a) The verifier corrects a car-existence judgment by distinguishing a truck from a car. (b) The verifier corrects a backpack-existence judgment using the bag's visible straps and design. Each example presents the initial response, verifier feedback, and final response. Red text highlights incorrect statements, while green text highlights corrections and verifier feedback. Continue indicates a request for further revision, and Accept indicates acceptance of the response.

unsupported content (Yin et al., 2024; Wu et al., 2026; Liu et al., 2026b). Instruction optimization uses errors, textual feedback, and execution trajectories to propose and evaluate prompt revisions across examples (Pryzant et al., 2023; Yuksekgonul et al., 2025; Agrawal et al., 2026). Complementary approaches generate candidate instructions from scored optimization histories (Yang et al., 2024), evolve task prompts and mutation prompts (Fernando et al., 2023), optimize instructions and demonstrations for multi-stage programs (Opsahl-Ong et al., 2024), or refine evaluator prompts from annotated examples (Zhen et al., 2025). Compared with these studies, L2V focuses on learning visual verification criteria from the downstream consequences of verifier judgments. Counterfactual interventions test whether correcting a specific judgment improves the final response, providing evidence for targeted criteria refinement; held-out replay then assesses whether the revised criteria benefit other examples.

6 Conclusion

In this paper, we propose Learning to Verify (L2V), a dual-loop agent framework that learns visual verification criteria for multimodal hallucination mitigation. The inner loop uses these judgments to guide response revision, while the outer loop learns from completed responses to improve the criteria during training. Counterfactual interventions identify verification errors whose correction improves the final response, and held-out replay selects criteria revisions that benefit other examples. Experiments on POPE and CHAIR show that L2V improves object-existence predictions and reduces hallucinated object mentions across the evaluated backbones, outperforming strong baselines and existing studies. Ablation results support the contributions of counterfactual attribution and held-out replay, indicating that learning verification criteria provides benefits beyond repeated verification with fixed criteria.

AI USE STATEMENT

We used generative AI tools to polish the manuscript and assist with collecting and organizing research information. For manuscript polishing, these tools helped refine sentence structure, improve clarity, and maintain consistent terminology. For research information, they assisted with identifying relevant literature and organizing information by topic to support comparison and synthesis.

REPRODUCIBILITY STATEMENT

The configuration and experimental execution settings are described in Section 3.3 (Implementation Details). We release the code at `https://anonymous.4open.science/r/L2V-4244` to facilitate reproducibility.

REFERENCES

Lakshya A. Agrawal, Shangyin Tan, Dilara Soylu, Noah Ziems, Rishi Khare, Krista Opsahl-Ong, Arnav Singhvi, Herumb Shandilya, Michael J. Ryan, Meng Jiang, Christopher Potts, Koushik Sen, Alexandros G. Dimakis, Ion Stoica, Dan Klein, Matei Zaharia, and Omar Khattab. GEPA: Reflective prompt evolution can outperform reinforcement learning. In *International Conference on Learning Representations*, pp. 8479–8565, 2026.

Jiaqi Bai, Hongcheng Guo, Zhongyuan Peng, Jian Yang, Zhoujun Li, Mohan Li, and Zhihong Tian. Mitigating hallucinations in large vision-language models by adaptively constraining information flow. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 23442–23450, 2025.

Zhaorun Chen, Zhuokai Zhao, Hongyin Luo, Huaxiu Yao, Bo Li, and Jiawei Zhou. Halc: Object hallucination reduction via adaptive focal-contrast decoding. *arXiv preprint arXiv:2403.00425*, 2024.

Mehrdad Fazli, Bowen Wei, and Ziwei Zhu. Inject to heal: Alleviating hallucination in lvlms via context embedding injection. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 41177–41193, 2026.

Chrisantha Fernando, Dylan Banarse, Henryk Michalewski, Simon Osindero, and Tim Rocktäschel. Promptbreeder: Self-referential self-improvement via prompt evolution. *arXiv preprint arXiv:2309.16797*, 2023.

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

Tianrui Guan, Fuxiao Liu, Xiyang Wu, Ruiqi Xian, Zongxia Li, Xiaoyu Liu, Xijun Wang, Lichang Chen, Furong Huang, Yaser Yacoob, et al. Hallusionbench: an advanced diagnostic suite for entangled language hallucination and visual illusion in large vision-language models. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14375–14385, 2024.

Yingnan Guo, Kejia Chen, Xiaofeng Zhang, Zifei Wu, and Yu Zhang. Trident: Self-supervised preference alignment via triplet regularization. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 31668–31683, 2026.

Yixiao He, Haifeng Sun, Pengfei Ren, Jingyu Wang, Huazheng Wang, Qi Qi, Zirui Zhuang, and Jing Wang. Evaluating and mitigating object hallucination in large vision-language models: Can they still see removed objects? In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 6841–6858, 2025.

Jie Huang, Xinyun Chen, Swaroop Mishra, Huaixiu Steven Zheng, Adams Yu, Xinying Song, and Denny Zhou. Large language models cannot self-correct reasoning yet. In *International conference on learning representations*, volume 2024, pp. 32808–32824, 2024a.

Qidong Huang, Xiaoyi Dong, Pan Zhang, Bin Wang, Conghui He, Jiaqi Wang, Dahua Lin, Weiming Zhang, and Nenghai Yu. Opera: Alleviating hallucination in multi-modal large language models via over-trust penalty and retrospection-allocation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13418–13427, 2024b.

Drew A Hudson and Christopher D Manning. GQA: A new dataset for real-world visual reasoning and compositional question answering. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6693–6702, 2019.

Zhangqi Jiang, Junkai Chen, Beier Zhu, Tingjin Luo, Yankun Shen, and Xu Yang. Devils in middle layers of large vision-language models: Interpreting, detecting and mitigating object hallucinations via attention lens. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25004–25014, 2025.

Liqiang Jing, Ruosen Li, Yunmo Chen, and Xinya Du. Faithscore: Fine-grained evaluations of hallucinations in large vision-language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 5042–5063, 2024.

Sicong Leng, Hang Zhang, Guanzheng Chen, Xin Li, Shijian Lu, Chunyan Miao, and Lidong Bing. Mitigating object hallucinations in large vision-language models through visual contrastive decoding. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13872–13882, 2024.

Yifan Li, Yifan Du, Kun Zhou, Jinpeng Wang, Xin Zhao, and Ji-Rong Wen. Evaluating object hallucination in large vision-language models. In *Proceedings of the 2023 conference on empirical methods in natural language processing*, pp. 292–305, 2023.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision*, pp. 740–755, 2014.

Alexander H. Liu, Kartik Khandelwal, Sandeep Subramanian, Victor Jouault, Abhinav Rastogi, Adrien Sadé, Alan Jeffares, Albert Jiang, Alexandre Cahill, Alexandre Gavaudan, et al. Ministral 3. *arXiv preprint arXiv:2601.08584*, 2026a.

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

Jiarui Liu, Weihao Xuan, Zhijing Jin, and Mona Diab. Taming object hallucinations with verified atomic confidence estimation. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 5430–5444, 2026b.

Holy Lovenia, Wenliang Dai, Samuel Cahyawijaya, Ziwei Ji, and Pascale Fung. Negative object presence evaluation (nope) to measure object hallucination in vision-language models. In *Proceedings of the 3rd Workshop on Advances in Language and Vision Research (ALVR)*, pp. 37–58, 2024.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, et al. Self-refine: Iterative refinement with self-feedback. *Advances in neural information processing systems*, 36:46534–46594, 2023.

Krista Opsahl-Ong, Michael J Ryan, Josh Purtell, David Broman, Christopher Potts, Matei Zaharia, and Omar Khattab. Optimizing instructions and demonstrations for multi-stage language model programs. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 9340–9366, 2024.

Eunkyu Park, Minyeong Kim, and Gunhee Kim. Halloc: Token-level localization of hallucinations for vision language models. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 29893–29903, 2025.

Shangpin Peng, Senqiao Yang, Li Jiang, and Zhuotao Tian. Mitigating object hallucinations via sentence-level early intervention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 635–646, 2025.

Reid Pryzant, Dan Iter, Jerry Li, Yin Lee, Chenguang Zhu, and Michael Zeng. Automatic prompt optimization with "gradient descent" and beam search. In *Proceedings of the 2023 conference on empirical methods in natural language processing*, pp. 7957–7968, 2023.

Anna Rohrbach, Lisa Anne Hendricks, Kaylee Burns, Trevor Darrell, and Kate Saenko. Object hallucination in image captioning. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pp. 4035–4045, 2018.

Pranab Sahoo, Prabhash Meharia, Akash Ghosh, Sriparna Saha, Vinija Jain, and Aman Chadha. A comprehensive survey of hallucination in large language, image, video and audio foundation models. In *Findings of the association for computational linguistics: EMNLP 2024*, pp. 11709–11724, 2024.

Pritish Sahu, Karan Sikka, and Ajay Divakaran. Pelican: Correcting hallucination in vision-LLMs via claim decomposition and program of thought verification. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 8228–8248, 2024.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. In *Advances in neural information processing systems*, volume 36, pp. 8634–8652, 2023.

Zhiqing Sun, Sheng Shen, Shengcao Cao, Haotian Liu, Chunyuan Li, Yikang Shen, Chuang Gan, Liangyan Gui, Yu-Xiong Wang, Yiming Yang, Kurt Keutzer, and Trevor Darrell. Aligning large multimodal models with factually augmented RLHF. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 13088–13110, 2024.

Qwen Team. Qwen3.5-omni technical report. *arXiv preprint arXiv:2604.15804*, 2026.

Chenxi Wang, Xiang Chen, Ningyu Zhang, Bozhong Tian, Haoming Xu, Shumin Deng, and Huajun Chen. Mllm can see? dynamic correction decoding for hallucination mitigation. In *International Conference on Learning Representations*, volume 2025, pp. 13712–13736, 2025a.

Sudong Wang, Yunjian Zhang, Yao Zhu, Enci Liu, Jianing Li, Yanwei Liu, and Xiangyang Ji. Shift: Smoothing hallucinations by information flow tuning for multimodal large language models. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3639–3649, 2025b.

Xintong Wang, Jingheng Pan, Liang Ding, and Chris Biemann. Mitigating hallucinations in large vision-language models with instruction contrastive decoding. In *Findings of the association for computational linguistics: ACL 2024*, pp. 15840–15853, 2024.

Tsung-Han Patrick Wu, Heekyung Lee, Jiaxin Ge, Joseph Gonzalez, Trevor Darrell, and David Chan. Generate, but verify: Reducing hallucination in vision-language models with retrospective resampling. *Advances in Neural Information Processing Systems*, 38:65749–65777, 2026.

Chengrun Yang, Xuezhi Wang, Yifeng Lu, Hanxiao Liu, Quoc V Le, Denny Zhou, and Xinyun Chen. Large language models as optimizers. In *International Conference on Learning Representations*, volume 2024, pp. 12028–12068, 2024.

Shukang Yin, Chaoyou Fu, Sirui Zhao, Tong Xu, Hao Wang, Dianbo Sui, Yunhang Shen, Ke Li, Xing Sun, and Enhong Chen. Woodpecker: Hallucination correction for multimodal large language models. *Science China Information Sciences*, 67(12):220105, 2024.

Tianyu Yu, Yuan Yao, Haoye Zhang, Taiwen He, Yifeng Han, Ganqu Cui, Jinyi Hu, Zhiyuan Liu, Hai-Tao Zheng, Maosong Sun, et al. Rlhf-v: Towards trustworthy mllms via behavior alignment from fine-grained correctional human feedback. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13807–13816, 2024.

Tianyu Yu, Haoye Zhang, Qiming Li, Qixin Xu, Yuan Yao, Da Chen, Xiaoman Lu, Ganqu Cui, Yunkai Dang, Taiwen He, et al. Rlaif-v: Open-source ai feedback leads to super gpt-4v trustworthiness. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19985–19995, 2025.

Mert Yuksekgonul, Federico Bianchi, Pratyusha Kalluri, Dan Jurafsky, and James Zou. When and why vision-language models behave like bags-of-words, and what to do about it? *arXiv preprint arXiv:2210.01936*, 2022.

Mert Yuksekgonul, Federico Bianchi, Joseph Boen, Sheng Liu, Pan Lu, Zhi Huang, Carlos Guestrin, and James Zou. Optimizing generative ai by backpropagating language model feedback. *Nature*, 639(8055):609–616, 2025.

Fatemeh Pesaran Zadeh, Yoojin Oh, and Gunhee Kim. Lpoi: Listwise preference optimization for vision language models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 26830–26844, 2025.

Chengsheng Zhang, Chenghao Sun, Xinyan Jiang, Wei Li, and Xinmei Tian. Prefill-time intervention for mitigating hallucination in large vision-language models. *arXiv preprint arXiv:2604.25642*, 2026.

Zefeng Zhang, Hengzhu Tang, Jiawei Sheng, Zhenyu Zhang, Yiming Ren, Zhenyang Li, Dawei Yin, Duohe Ma, and Tingwen Liu. Debiasing multimodal large language models via noise-aware preference optimization. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9423–9433, 2025.

Cheng Zhen, Ervine Zheng, Jilong Kuang, and Geoffrey Jay Tso. Enhancing llm-as-a-judge through active-sampling-based prompt optimization. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)*, pp. 960–970, 2025.

Weihong Zhong, Xiaocheng Feng, Liang Zhao, Qiming Li, Lei Huang, Yuxuan Gu, Weitao Ma, Yuan Xu, and Bing Qin. Investigating and mitigating the multimodal hallucination snowballing in large vision-language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 11991–12011, 2024.

Guanyu Zhou, Yibo Yan, Xin Zou, Kun Wang, Aiwei Liu, and Xuming Hu. Mitigating modality prior-induced hallucinations in multimodal large language models via deciphering attention causality. In *International Conference on Learning Representations*, volume 2025, pp. 54415–54439, 2025.

Yiyang Zhou, Chenhang Cui, Jaehong Yoon, Linjun Zhang, Zhun Deng, Chelsea Finn, Mohit Bansal, and Huaxiu Yao. Analyzing and mitigating object hallucination in large vision-language models. In *International Conference on Learning Representations*, volume 2024, pp. 56969–56998, 2024.

Xin Zou, Yizhou Wang, Yibo Yan, Yuanhuiyi Lyu, Kening Zheng, Sirui Huang, Junkai Chen, Peijie Jiang, Jia Liu, Chang Tang, et al. Look twice before you answer: Memory-space visual retracing for hallucination mitigation in multimodal large language models. *arXiv preprint arXiv:2410.03577*, 2024.