UNIFIED MULTIMODAL MODELS DO NOT LEARN IN UNISON: ACTIVE PROBING OF CROSS-CAPABILITY TRANSFER

Anonymous authors
Paper under double-blind review

ABSTRACT

Shared parameters in unified multimodal models do not imply shared learning. In our controlled setting, optimizing one capability can reduce the other by up to 3.70 percentage points on average, and similar passive gradient statistics can precede opposite receiver outcomes. We introduce Active Mechanism-Separating Probe (AMSP). AMSP first selects a bounded native low-rank probe that maximizes robust separation among action-distinct finite-response hypotheses. It measures source and receiver responses on a disposable branch, then commits a source-feasible intervention from the untouched optimizer state; insufficient evidence leads to abstention. On Show-o2-7B, AMSP reaches a joint understanding–generation harmonic score of 82.36%, exceeding Pareto LoRA by 1.68 points and equal-budget lookahead by 1.85 points; both margins remain positive under paired seed-and-combination resampling. On InternVL-U-4B, it again achieves the highest mean score at 84.42%. Lookahead repairs more destructive events, but AMSP preserves 98.15% of constructive updates versus 79.63% for lookahead.

1 INTRODUCTION

Unified multimodal models couple understanding and generation through shared computation (Diao et al., 2026; Liu et al., 2025; Peng et al., 2026). Show-o2 combines autoregressive language modeling with flow-based visual generation; InternVL-U places understanding and a specialized generation module in a unified contextual framework (Xie et al., 2025; Tian et al., 2026). Yet sharing parameters does not guarantee beneficial transfer. Related work finds architecture- and task-dependent learning synergy (Kang et al., 2026; Wu et al., 2026), while GapEval documents understanding–generation inconsistencies (Wang et al., 2026a). In our controlled experiments, understanding-only training lowers generation by 3.70 percentage points on average.

Update-level remedies project conflicting gradients, balance tasks, or optimize competing objectives (Yu et al., 2020; Liu et al., 2021; Chen et al., 2018; Wei et al., 2026; Navon et al., 2022; Liu et al., 2023). However, similar passive statistics can precede opposite receiver outcomes (Figure 1), and suppressing an apparently conflicting update can erase useful transfer. The missing quantity is the *value of intervention*: is changing the ordinary update better than leaving it alone?

AMSP uses a reversible low-rank probe to distinguish hypotheses that recommend different actions. Source and receiver responses are measured on a disposable optimizer branch; AMSP then commits a source-feasible action from the untouched state, or abstains when evidence is weak. Here, "mechanism" means an action-distinct finite-response hypothesis, not a unique neural circuit.

Our contributions are:

- We formulate interference as finite action identification, separating transfer sign from intervention value, and introduce AMSP to resolve action-relevant ambiguity.
- Across Show-o2-7B and InternVL-U-4B, AMSP achieves the highest mean joint understanding–generation score among the compared methods.
- Event-level analysis reveals a repair–retention trade-off: repairing more destructive updates can sacrifice more constructive transfer.

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.

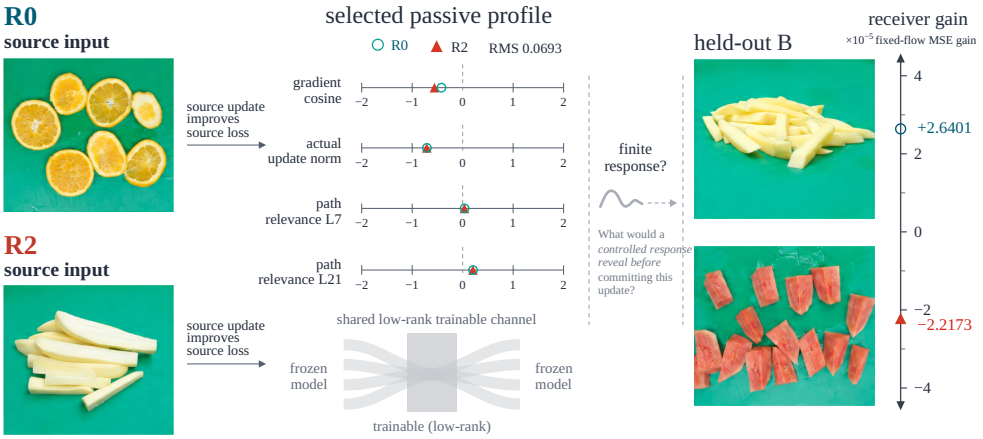


Figure 1: A measured observability gap. R0 and R2 have closely matched selected passive profiles ($d_{\mathrm{RMS}} = 0.0693$), yet their held-out receiver gains have opposite signs: $+2.6401$ and -2.2173 ($\times 10^{-5}$ fixed-flow MSE gain). The open finite-response aperture marks the additional measurement considered by AMSP; it is not a measured response for this pair.

2 Actionable Cross-Capability Interference

2.1 Cross-capability learning as a state transition

For the capability set $\mathcal{C}$, let $s, r \in \mathcal{C}$, $s \neq r$, denote source and receiver capabilities; s supplies the current learning signal. Our experiments use understanding U and generation G. The complete state is $x = (\phi, \omega)$, with shared trainable parameters ϕ and optimizer, stochastic, and other stateful quantities ω. Action $a \in \mathcal{A}(x)$ induces $x' = T_a(x)$; a_0 denotes the ordinary source update. Accounting for the optimizer matters because Adam can distort multi-objective update geometry (Liu et al., 2026a).

Here, ϕ comprises shared rank-4 LoRA factors on Transformer q, k, v, o projections (Hu et al., 2021); the base model and task-specific heads remain frozen.

2.2 Transfer sign is not intervention value

For any capability $c \in \mathcal{C}$, define the finite consequence of action a from the same pre-update state x as

$$\Delta_c(a \mid x) = L_c(T_a(x)) - L_c(x), \qquad g_c(a \mid x) = -\Delta_c(a \mid x). \tag{1}$$

Positive g_c denotes improvement. Intervention value compares a with the ordinary update:

$$V_r(a; a_0 \mid x) = g_r(a \mid x) - g_r(a_0 \mid x). \tag{2}$$

A destructive ordinary update can have $g_r(a_0) < 0$, while an alternative has $V_r(a; a_0) > 0$ but still leaves $g_r(a) < 0$; the intervention reduces damage without repairing it. Conversely, when $g_r(a_0) > 0$, an aggressive correction can have negative intervention value by erasing useful transfer. Detecting the sign of $g_r(a_0)$ therefore does not rank the alternatives.

When the ordinary source step makes progress, source-feasible actions satisfy

$$\mathcal{A}_s(x) = \{a \in \mathcal{A}(x) : g_s(a \mid x) \geq \eta\, g_s(a_0 \mid x)\}, \tag{3}$$

with $\eta \in [0, 1]$ fixed before evaluation. Receiver events are called constructive when $g_r(a_0 \mid x) > \epsilon$, destructive when $g_r(a_0 \mid x) < -\epsilon$, and uncertain inside the neutral band. For evaluation, a constructive event is retained when the selected action preserves at least a fraction κ of its ordinary receiver gain.

We fix $\eta = 0.6$, $\kappa = 0.8$, and $\epsilon = 0.003$ before held-out B evaluation. These thresholds govern source feasibility and event accounting in normalized-loss units, not task accuracy.

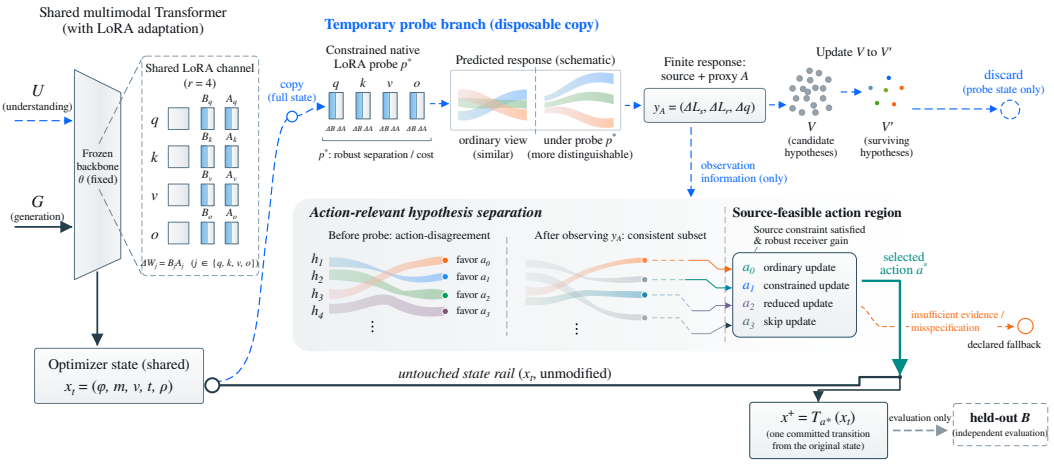


Figure 2: AMSP probes a disposable copy of x_t, uses source and proxy-A responses to contract $\mathcal{V}$ to $\mathcal{V}'$, and commits one source-feasible action from the untouched state. Held-out B is evaluated only after commitment.

2.3 The Observability Gap

For parameter displacement $\delta\phi_a$, the receiver response is

$$\Delta_r(a \mid x) = \nabla_\phi L_r(x)^\top \delta\phi_a + R_r(a, x), \tag{4}$$

where $R_r(a, x)$ collects finite-step curvature and state-dependent effects. Gradient cosine, inner products, update norms, and path summaries characterize parts of the first term, but do not directly observe the action-specific remainder or unexecuted alternatives. An actionable diagnostic must inform their relative value before commitment.

3 Active Mechanism-Separating Probe

AMSP probes a temporary copy of the source-learning state x_t, records finite source and receiver proxy-A responses, and discards the copy (Figure 2).

3.1 Probe Only Where the Decision Is Ambiguous

For each adapted projection $j \in \{q, k, v, o\}$, AMSP probes the native LoRA channel:

$$W_j' = W_j + B_j A_j, \qquad A_j \in \mathbb{R}^{r \times d_{\text{in}}}, \quad B_j \in \mathbb{R}^{d_{\text{out}} \times r}, \tag{5}$$

with $r = 4$. Low-rank optimization has parameterization-dependent geometry (Bogachev et al., 2025), and interactions between factor updates can reintroduce interference even under factor-level constraints (Chen et al., 2026).

Predeclared probes transform the proposed factor-space update, for example by bounded scaling, sign changes of selected atoms, or omission of a projection group. Enforcing relative-update, source-cost, and compute budgets gives $\mathcal{P}_{\text{safe}}(x_t)$.

The separate final action space is $\mathcal{A}(x_t) = \{a_0, a_1, a_2, a_3\}$: ordinary, constrained, reduced, and skip transitions. Informative probes need not resemble committed actions, as in query selection for downstream computational outputs (Neiswanger et al., 2021).

Hypotheses $\mathcal{V}(x_t) = \{h_1, \ldots, h_K\}$ predict legal-probe responses

$$P_h(Y_A \mid x_t, p), \qquad Y_A = (\Delta L_s, \Delta L_r, \Delta q), \tag{6}$$

where ΔL_r is measured on receiver proxy-A and Δq denotes optional predeclared internal summaries. Held-out B never appears in Y_A.

For hypothesis h, let

$$\pi_h(x_t) \in \arg\max_{a \in \mathcal{A}_{s,h}(x_t)} \widehat{V}_{r,h}(a; a_0 \mid x_t) \tag{7}$$

denote its preferred source-feasible action. AMSP separates only action-distinct pairs:

$$\mathcal{D}(\mathcal{V}) = \{(h, h') : \pi_h(x_t) \neq \pi_{h'}(x_t)\}. \tag{8}$$

This avoids separating decision-equivalent hypotheses, connecting to equivalence-class determination (Golovin et al., 2010) and task-driven experimental design (Rossa et al., 2026).

For a candidate p, let D_L denote a lower confidence bound on the separation of two predicted finite-response distributions, and let δ_p account for response-model error. We define

$$S(p; \mathcal{V}) = \min_{(h,h') \in \mathcal{D}(\mathcal{V})} \left[D_L(P_h(Y_A \mid x_t, p), P_{h'}(Y_A \mid x_t, p)) - 2\delta_p \right], \tag{9}$$

and choose

$$p^* = \arg\max_{p \in \mathcal{P}_{\text{safe}}(x_t)} \frac{S(p; \mathcal{V})}{C(p)}, \tag{10}$$

where $C(p)$ is the probe cost.

The max–min objective protects the weakest action-relevant distinction under model uncertainty (Harikumar et al., 2026). Large response magnitude alone is not enough: if every surviving hypothesis predicts the same change, it does not distinguish actions. If $\mathcal{D}(\mathcal{V})$ is empty, no active measurement is needed.

The selected probe is executed once on the disposable branch. After observing y_A, AMSP keeps the consistent hypotheses:

$$\mathcal{V}' = \{h \in \mathcal{V} : -\log P_h(y_A \mid x_t, p^*) \leq \tau_h\}, \tag{11}$$

with thresholds calibrated before held-out evaluation.

Failed robust separation, an out-of-envelope response, or an unreliable survivor set invokes the declared fallback.

3.2 From surviving hypotheses to one action

The remaining hypotheses first restrict the action family to transitions that robustly retain source progress:

$$\mathcal{A}_s(\mathcal{V}') = \left\{ a \in \mathcal{A}(x_t) : \min_{h \in \mathcal{V}'} [\widehat{g}_{s,h}(a \mid x_t) - \eta \widehat{g}_{s,h}(a_0 \mid x_t)] \geq 0 \right\}. \tag{12}$$

For every feasible action, AMSP evaluates the intervention value defined in Eq. (2) under the least favorable surviving hypothesis:

$$R(a \mid \mathcal{V}') = \min_{h \in \mathcal{V}'} \widehat{V}_{r,h}(a; a_0 \mid x_t), \tag{13}$$

and selects

$$a^* \in \arg\max_{a \in \mathcal{A}_s(\mathcal{V}')} R(a \mid \mathcal{V}'). \tag{14}$$

An alternative replaces a_0 only when its robust advantage clears the predeclared margin. If surviving hypotheses support constructive ordinary transfer and no alternative reliably improves it, AMSP retains a_0.

Unlike the candidate skip action a_3, which may be selected from inferred responses, fallback applies when inference cannot support a reliable decision. Its transition semantics are fixed by the implementation.

AMSP commits exactly once from the untouched state:

$$x^+ = T_{a^*}(x_t). \tag{15}$$

Algorithm 1 AMSP at one source-learning state

Require: untouched state x_t, proxy-A, hypotheses $\mathcal{V}$, safe probes $\mathcal{P}_{\text{safe}}(x_t)$, actions $\mathcal{A}(x_t)$, declared fallback

Compute $\pi_h(x_t)$ and merge decision-equivalent hypotheses
Form $\mathcal{D}(\mathcal{V})$
if $\mathcal{D}(\mathcal{V}) \neq \varnothing$ **then**
    Score safe probes by Eq. (9) and select p^* by Eq. (10)
    if $S(p^*; \mathcal{V}) \leq 0$ **then**
        return declared fallback from untouched x_t
    end if
    Clone the optimizer state and execute p^* on source/proxy-A
    Record y_A, discard the temporary state, and form $\mathcal{V}'$ by Eq. (11)
    if $\mathcal{V}'$ fails the prespecified consistency checks **then**
        return declared fallback from untouched x_t
    end if
else
    $\mathcal{V}' \leftarrow \mathcal{V}$
end if
Construct $\mathcal{A}_s(\mathcal{V}')$ and select a^* by Eq. (14)
if the declared decision criterion is not satisfied **then**
    return declared fallback from untouched x_t
end if
Recompute and commit $x^+ = T_{a^*}(x_t)$ exactly once
Evaluate held-out B after commitment only

3.3 State isolation and cost

For AdamW (Loshchilov & Hutter, 2019), isolation requires copying LoRA factors, moments (m, v), step counter, random state ρ, relevant buffers, and data position. The selected action is recomputed from the original x_t after discarding the copy.

Direct pairwise candidate scoring requires $O(|\mathcal{P}||\mathcal{D}(\mathcal{V})|d_Y)$ response-model operations, with worst-case complexity $O(|\mathcal{P}||\mathcal{V}|^2 d_Y)$. Only p^* is executed through the neural model, so candidate scoring does not require a resident ensemble of full model copies.

4 Experimental Setup

4.1 Models and task split

We evaluate Show-o2-7B and replicate on InternVL-U-4B (Xie et al., 2025; Tian et al., 2026), using Chop & Learn (Saini et al., 2023). Its 112 compositions are split into 70 training-side, 17 calibration, and 25 final-test compositions (838/204/296 images). The training-side split contains 54 source-learning and 16 receiver proxy-A compositions. Proxy-A is excluded from ordinary source gradients and queried for finite responses. The final 25 compositions form held-out B, inaccessible during probe and action selection.

Understanding predicts object and state in text form. Generation is scored with a frozen object/state classifier using 16 noise-conditioned samples for each held-out composition and seed. Like object-focused compositional evaluation (Ghosh et al., 2023), this metric targets semantic correctness; it does not measure general image realism or diversity.

4.2 Training and controlled baselines

All methods use the same shared rank-4 LoRA channel on the Transformer q, k, v, o projections. Unless otherwise stated, optimization uses AdamW for 600 source-learning steps with learning rate 1.5×10^{-5}. Experiments are run on two NVIDIA RTX 4090 GPUs.

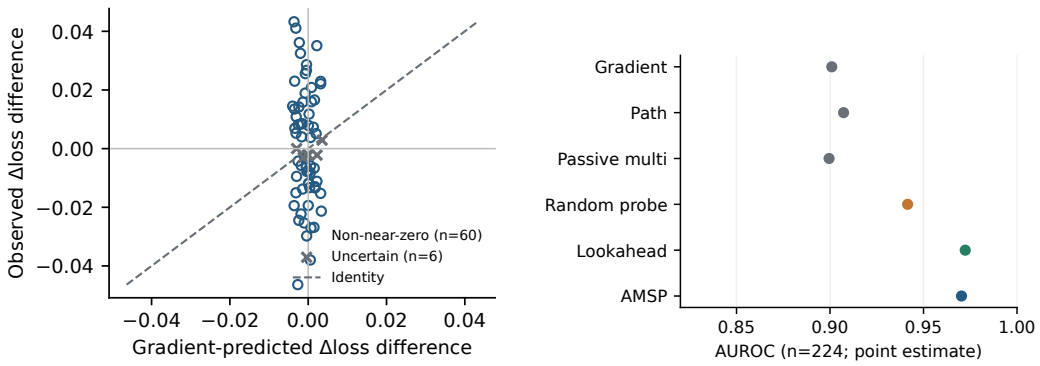


Figure 3: Passive observability and active response information on Show-o2. (a) Gradient-predicted versus observed receiver-loss differences for 66 prespecified matched pairs; six near-zero cases remain marked as uncertain. (b) Receiver-sign AUROC over 224 eligible events.

Show-o2 evaluates all baselines in Table 1; InternVL-U repeats the central subset in Table 3. The `-Adapt` suffix denotes implementation in the common rank-4 parameterization.

The active controls receive the same two-probe budget. Random probing removes AMSP's probe-selection rule while retaining the downstream controller; equal-budget lookahead receives the same number of additional response evaluations.

4.3 Metrics and statistical protocol

For each seed, understanding and generation scores are macro-averaged over the 25 held-out compositions. If U_s and G_s are the joint accuracies for seed s, we compute

$$H_s = \frac{2U_sG_s}{U_s + G_s} \tag{16}$$

before averaging H_s across seeds.

Show-o2 uses five seeds per main configuration; InternVL-U uses three. We report sample SD and match seed indices across methods. Principal Show-o2 contrasts use 1,000 paired seed-by-composition resamples, sharing sampled indices between methods.

The separate event study contains 400 source-learning events, 2,400 probe evaluations, and 2,000 candidate-action outcomes, not additional training runs. B-side potential outcomes serve only retrospective scoring. Repair requires a source-feasible action to improve a destructive receiver outcome.

5 Does Active Identification Improve Shared Multimodal Learning?

5.1 Passive evidence is informative, but not action-complete

We form 66 non-overlapping Show-o2 event pairs whose passive signatures are matched before held-out B is examined. Sixty pairs have a non-negligible receiver difference. A gradient inner-product predictor orders only 24/60 of these pairs correctly, compared with 45/60 for AMSP response information and 51/60 for equal-budget lookahead. On the full set of 224 signed Show-o2 events, gradient and multi-statistic passive predictors already reach AUROC 0.901 and 0.900.

As shown in Figure 3b, AMSP raises receiver-sign accuracy from 84.38% for the passive multi-statistic predictor to 91.07%, with AUROC increasing from 0.8995 to 0.9703. Equal-budget lookahead reaches AUROC 0.9723 but slightly lower sign accuracy at 89.73%. Their normalized-loss MAEs are almost identical, 0.00751 for AMSP and 0.00755 for lookahead.

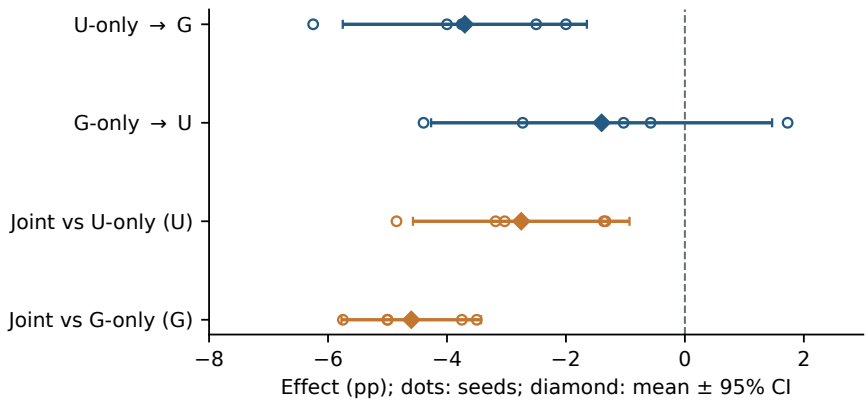


Figure 4: Cross-capability effects over five matched Show-o2 seeds. Open circles are individual seeds; diamonds denote means and horizontal bars show 95% paired-t intervals. Negative values indicate degradation of the reported capability.

Table 1: Show-o2-7B results over five seeds. U/G are joint understanding and generation accuracies (%); H is computed within each seed and then averaged. Bold denotes the highest trained-method mean.

Method	U-joint	G-joint	H
Frozen reference	78.48	56.75	65.87
U-only	88.91 ± 1.73	53.05 ± 1.65	66.44 ± 1.63
G-only	77.08 ± 2.31	77.95 ± 1.66	77.51 ± 1.81
Joint LoRA	84.76 ± 2.25	69.65 ± 2.40	76.45 ± 1.97
PCGrad-Adapt	84.62 ± 1.98	71.35 ± 2.32	77.41 ± 1.92
CAGrad-Adapt	85.04 ± 2.36	72.45 ± 2.02	78.22 ± 1.70
GradNorm-Adapt	84.62 ± 1.68	72.90 ± 2.01	78.31 ± 1.65
Pareto LoRA-Adapt	86.67 ± 2.26	75.50 ± 1.99	80.68 ± 1.54
Equal-budget lookahead	86.59 ± 2.10	75.25 ± 1.66	80.51 ± 1.46
Random probe + controller	84.95 ± 1.30	72.30 ± 2.15	78.10 ± 1.45
AMSP	**87.62 ± 2.50**	**77.70 ± 1.01**	**82.36 ± 1.56**

5.2 Active identification improves the joint learning trade-off

Figure 4 reports paired effects over the five Show-o2 seeds. Understanding-only training reduces generation by 3.70 percentage points on average, with a 95% paired interval of $[-5.75, -1.65]$. Generation-only training has a smaller mean effect on understanding, -1.40 points, with a wider interval that crosses zero.

Relative to Pareto LoRA-Adapt, AMSP gains 2.20 points on generation and 1.68 on H; paired seed-by-composition resampling gives 95% intervals of $[0.70, 3.70]$ and $[0.71, 2.64]$, respectively. Against equal-budget lookahead, the corresponding gains are 2.45 and 1.85 points, with intervals $[1.15, 3.90]$ and $[1.06, 2.78]$. The understanding differences are smaller and their intervals cross zero.

5.3 Better prediction does not guarantee a better intervention

Among 250 Show-o2 events (Figure 5), 116 ordinary updates are destructive, 108 are constructive, and 26 fall inside the neutral band.

Equal-budget lookahead is the most aggressive repair policy: it repairs 101/116 destructive events (87.07%). AMSP repairs 69/116 (59.48%). The ordering reverses on constructive transfer. Looka-

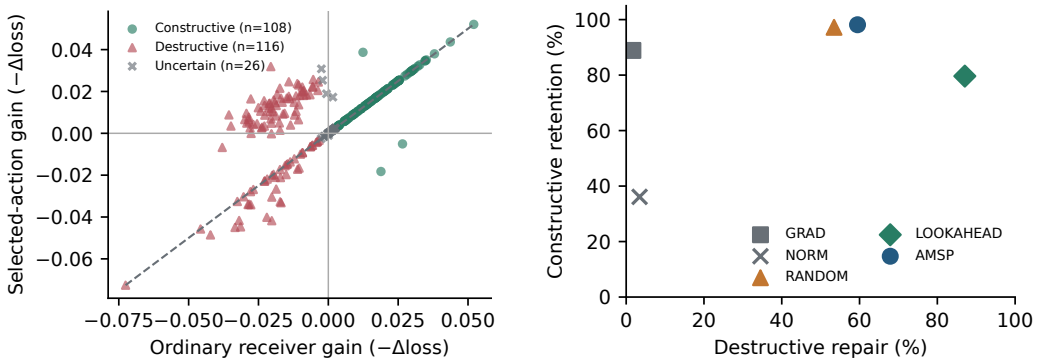


Figure 5: Event-level selective control on Show-o2. (a) Ordinary versus AMSP-selected receiver gain for all 250 events; the dashed line denotes no change. (b) Destructive repair versus constructive retention for five policies.

Table 2: Key Show-o2-7B ablations over five seeds. ΔH is relative to full AMSP; additional ablations are reported in the appendix.

Variant	U-joint	G-joint	H	ΔH
AMSP	87.62 ± 2.50	77.70 ± 1.01	82.36 ± 1.56	+0.00
Passive replacement	86.55 ± 2.42	73.15 ± 1.23	79.28 ± 1.50	-3.08
Random probe	84.95 ± 1.30	72.30 ± 2.15	78.10 ± 1.45	-4.26
Same action for all hypotheses	86.06 ± 2.49	74.30 ± 2.30	79.73 ± 1.85	-2.63
Linear-response replacement	86.34 ± 2.42	75.80 ± 0.91	80.72 ± 1.45	-1.64
No abstention	87.36 ± 2.25	76.15 ± 0.84	81.36 ± 1.13	-1.00
No constructive protection	85.33 ± 2.56	77.85 ± 1.24	81.40 ± 1.58	-0.95

head retains 86/108 constructive events (79.63%), whereas AMSP retains 106/108 (98.15%). Random probing is similarly conservative, retaining 97.22% while repairing 53.45%.

Predictive and control performance also diverge on InternVL-U: random probing obtains higher event-level AUROC than AMSP (0.9838 versus 0.9788), but its final joint score is lower (82.06 versus 84.42).

5.4 What parts of AMSP matter?

Table 2 isolates the active response, probe selection, action disagreement, and finite-response modeling. Forcing all hypotheses to recommend the same action directly tests Eq. (8).

Removing constructive protection raises generation accuracy slightly, from 77.70 to 77.85%, but understanding drops from 87.62 to 85.33%, so the joint score falls.

In the three-seed sensitivity study, moving from two to four probes changes H only from 82.34 to 82.38, and $K = 4$ versus $K = 8$ gives 82.34 versus 82.11. Doubling the update learning rate gives 82.09. Rank 8 reaches 82.93, while rank 2 and the single-probe setting show the clearest deterioration. We keep the rank-4, two-probe configuration for the controlled comparisons rather than selecting the best sensitivity point after evaluation.

5.5 The pattern replicates across architectures

InternVL-U provides an independent architectural check (Table 3): AMSP again obtains the highest mean joint score, while Pareto LoRA-Adapt is marginally higher on understanding.

With only three seeds, the H differences against Pareto LoRA-Adapt and lookahead remain uncertain under paired seed-by-composition resampling: $+0.83$ pp with interval $[-0.39, 2.34]$ and $+0.69$ pp with interval $[-0.58, 2.18]$. The generation gains are more stable, remaining positive against

Table 3: InternVL-U-4B validation over three seeds. Bold denotes the highest trained-method mean in each column.

Method	U-joint	G-joint	H
Frozen reference	80.91	69.00	74.48
Joint LoRA	85.97 ± 1.58	75.08 ± 2.24	80.14 ± 1.34
PCGrad-Adapt	86.86 ± 1.03	76.83 ± 0.95	81.53 ± 0.69
Pareto LoRA-Adapt	$\mathbf{88.56 \pm 1.79}$	79.17 ± 2.16	83.59 ± 1.69
Equal-budget lookahead	88.43 ± 2.28	79.50 ± 1.09	83.73 ± 1.60
Random probe + controller	87.56 ± 1.17	77.25 ± 2.38	82.06 ± 1.25
AMSP	88.44 ± 0.61	$\mathbf{80.75 \pm 1.39}$	$\mathbf{84.42 \pm 0.95}$

both Pareto ($+1.58$, $[0.25, 3.00]$) and lookahead ($+1.25$, $[0.33, 2.59]$). The result should therefore be read as a cross-architecture replication of the ordering, rather than a second statistically resolved benchmark.

On InternVL-U, lookahead repairs 88.75% of destructive events but retains 80.00% of constructive ones; AMSP repairs 52.50% and retains 98.33%. Across the available external capability checks on both backbones, AMSP stays within 0.43 percentage points of the corresponding frozen checkpoint, suggesting that the selective update policy does not obtain its task-level gain by broadly erasing the pretrained capabilities measured by those evaluations.

6 Limitations

Our operational hypotheses do not identify neural circuits, and misspecification checks detect only 16/32 injected Show-o2 cases and 5/20 InternVL-U cases. Distribution-free risk control requires separate valid calibration and assumptions (Angelopoulos et al., 2024). Evaluation covers two backbones, one benchmark, and five/three seeds; InternVL-U joint-score margins remain uncertain. The frozen semantic classifier does not establish generation realism or diversity. Broader joint-capability, multi-turn, and generation-assisted understanding evaluations remain outside this protocol (Liu et al., 2026b; Meng et al., 2026; Tong et al., 2026; Zhu et al., 2026). We claim no wall-clock or memory advantage without measured efficiency comparisons, including deployment-oriented settings (Shaker et al., 2026).

7 Conclusion

AMSP treats cross-capability interference as action identification: a reversible probe resolves action-relevant uncertainty before one source-feasible update is committed. Across two unified multimodal backbones, it achieves the highest mean joint score among compared methods while retaining more constructive transfer than aggressive lookahead. Effective control requires identifying when intervention is worthwhile, beyond predicting ordinary transfer.

REFERENCES

Anastasios N. Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal Risk Control. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2208.02814.

Vladimir Bogachev, Vladimir Aletov, Alexander Molozhavenko, Denis Bobkov, Vera Soboleva, Aibek Alanov, and Maxim Rakhuba. LoRA meets Riemannion: Muon Optimizer for Parametrization-independent Low-Rank Adapters, 2025. URL https://arxiv.org/abs/2507.12142.

Cheng Chen, Pengpeng Zeng, Yuyu Guo, Lianli Gao, Hengtao Shen, and Jingkuan Song. Janus-LoRA: A Balanced Low-Rank Adaptation for Continual Learning. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2605.28495.

Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 793–802, 2018. URL https://arxiv.org/abs/1711.02257.

Haiwen Diao, Penghao Wu, Hanming Deng, Jiahao Wang, Shihao Bai, Silei Wu, Weichen Fan, Wenjie Ye, Wenwen Tong, Xiangyu Fan, Yan Li, Yubo Wang, Zhijie Cao, Zhiqian Lin, Zhitao Yang, Zhongang Cai, Yuwei Niu, Yue Zhu, Bo Liu, Chengguang Lv, Haojia Yu, Haozhe Xie, Hongli Wang, Jianan Fan, Jiaqi Li, Jiefan Lu, Jingcheng Ni, Junxiang Xu, Kaihuan Liang, Lianqiang Shi, Linjun Dai, Linyan Wang, Oscar Qian, Peng Gao, Pengfei Liu, Qingping Sun, Rui Shen, Ruisi Wang, Shengnan Ma, Shuang Yang, Siyi Xie, Siying Li, Tianbo Zhong, Xiangli Kong, Xuanke Shi, Yang Gao, Yongqiang Yao, Yves Wang, Zhengqi Bai, Zhengyu Lin, Zixin Yin, Wenxiu Sun, Ruihao Gong, Quan Wang, Lewei Lu, Lei Yang, Ziwei Liu, and Dahua Lin. SenseNova-U1: Unifying Multimodal Understanding and Generation with NEO-unify Architecture, 2026. URL https://arxiv.org/abs/2605.12500.

Fengyi Fu, Mengqi Huang, Shaojin Wu, Yunsheng Jiang, Yufei Huo, Hao Li, Yinghang Song, Fei Ding, Jianzhu Guo, Qian He, Zheren Fu, Zhendong Mao, and Yongdong Zhang. Lance: Unified Multimodal Modeling by Multi-Task Synergy, 2026. URL https://arxiv.org/abs/2605.18678.

Dhruba Ghosh, Hanna Hajishirzi, and Ludwig Schmidt. GenEval: An Object-Focused Framework for Evaluating Text-to-Image Alignment. In *Advances in Neural Information Processing Systems*, 2023. URL https://arxiv.org/abs/2310.11513.

Daniel Golovin, Andreas Krause, and Debajyoti Ray. Near-Optimal Bayesian Active Learning with Noisy Observations. In *Advances in Neural Information Processing Systems*, 2010. URL https://arxiv.org/abs/1010.3091.

Jitai Hao, Hao Liu, Xinyan Xiao, Qiang Huang, and Jun Yu. Uni-X: Mitigating Modality Conflict with a Two-End-Separated Architecture for Unified Multimodal Models. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2509.24365.

Haripriya Harikumar, Sammie Katt, Yasir Zubayr Barlas, and Samuel Kaski. Robust Bayesian Decision Making under Adversarial Uncertainty, 2026. URL https://arxiv.org/abs/2607.08590.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, 2021. URL https://arxiv.org/abs/2106.09685.

Jiwon Kang, Heeji Yoon, Jaewoo Jung, Jaewon Min, Minkyeong Jeon, Biyeon Hwang, Sangwon Jung, and Seungryong Kim. Transferability between understanding and generation in unified multimodal models, 2026. URL https://arxiv.org/abs/2607.04423. Accepted at ECCV 2026.

Lijiang Li, Zuwei Long, Yunhang Shen, Heting Gao, Haoyu Cao, Xing Sun, Caifeng Shan, Ran He, and Chaoyou Fu. Omni-Diffusion: Unified Multimodal Understanding and Generation with Masked Discrete Diffusion. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2603.06577.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. In *Advances in Neural Information Processing Systems*, 2021. URL https://arxiv.org/abs/2110.14048.

Bo Liu, Yihao Feng, Peter Stone, and Qiang Liu. FAMO: Fast Adaptive Multitask Optimization. In *Advances in Neural Information Processing Systems*, 2023. URL https://arxiv.org/abs/2306.03792.

Fengbei Liu, Rachit Saluja, Sunwoo Kwak, Ruibo Wang, Ruining Deng, Heejong Kim, Johannes C. Paetzold, and Mert R. Sabuncu. MAdam: Metric-Aware Multi-Objective Adam, 2026a. URL https://arxiv.org/abs/2606.03904.

Jinyu Liu, Xincheng Shuai, Henghui Ding, and Yu-Gang Jiang. Unison: Benchmarking Unified Multimodal Models via Synergistic Understanding and Generation. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026b. URL https://arxiv.org/abs/2606.26984.

Zeyu Liu, Zanlin Ni, Yang Yue, Cheng Da, Huan Yang, Di Zhang, Kun Gai, and Gao Huang. Steering Visual Generation in Unified Multimodal Models with Understanding Supervision, 2026c. URL https://arxiv.org/abs/2605.05781.

Zhiheng Liu, Weiming Ren, Haozhe Liu, Zijian Zhou, Shoufa Chen, Haonan Qiu, Xiaoke Huang, Zhaochong An, Fanny Yang, Aditya Patel, Viktar Atliha, Tony Ng, Xiao Han, Chuyan Zhu, Chenyang Zhang, Ding Liu, Juan-Manuel Perez-Rua, Sen He, Jürgen Schmidhuber, Wenhu Chen, Ping Luo, Wei Liu, Tao Xiang, Jonas Schult, and Yuren Cong. TUNA: Taming Unified Visual Representations for Native Unified Multimodal Models, 2025. URL https://arxiv.org/abs/2512.02014.

Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1711.05101.

Renjie Lu, Xulong Zhang, Xiaoyang Qu, Shangfei Wang, and Jianzong Wang. DIVA: Harnessing the Representation Divergence in Unified Multimodal Models for Mutual Reinforcement. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2605.25328.

Yinyi Luo, Wenwen Wang, Hayes Bai, Marios Savvides, and Jindong Wang. LatentUMM: Dual Latent Alignment for Unified Multimodal Models, 2026. URL https://arxiv.org/abs/2605.17766.

Lingyi Meng, Zecong Tang, Haoran Li, Tengju Ru, Zhejun Cui, Weitong Lian, Qi Kang, Hangshuo Cao, Yichen Zhu, Yechi Liu, Kaixuan Wang, Yu-Jie Yuan, Chunwei Wang, Yu Zhang, and Bo Dai. IMUG-Bench: Benchmarking Unified Multimodal Models on Interleaved Understanding and Generation, 2026. URL https://arxiv.org/abs/2606.09169.

Aviv Navon, Aviv Shamsian, Idan Achituve, Haggai Maron, Kenji Kawaguchi, Gal Chechik, and Ethan Fetaya. Multi-Task Learning as a Bargaining Game. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 16428–16446, 2022. URL https://arxiv.org/abs/2202.01017.

Willie Neiswanger, Ke Alexander Wang, and Stefano Ermon. Bayesian Algorithm Execution: Estimating Computable Properties of Black-box Functions Using Mutual Information. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 8005–8015, 2021. URL https://arxiv.org/abs/2104.09460.

Wujian Peng, Lingchen Meng, Yuxuan Cai, Xianwei Zhuang, Yuhuan Yang, Rongyao Fang, Chenfei Wu, Junyang Lin, Zuxuan Wu, and Shuai Bai. Unified Multimodal Autoregressive Modeling with Shared Context-Visual Tokenizer is Key to Unification. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2606.18249.

Hector G. Rodriguez and Marcus Rohrbach. SIEVES: Selective Prediction Generalizes through Visual Evidence Scoring, 2026. URL https://arxiv.org/abs/2604.25855.

Tom Rossa, Angus Phillips, and Tom Rainforth. Action-BED: Task-Driven Bayesian Experimental Design with Singly Intractable Objectives, 2026. URL https://arxiv.org/abs/2606.23662.

Nirat Saini, Hanyu Wang, Archana Swaminathan, Vinoj Jayasundara, Bo He, Kamal Gupta, and Abhinav Shrivastava. Chop & learn: Recognizing and generating object-state compositions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 20247–20258, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Saini_Chop__Learn_Recognizing_and_Generating_Object-State_Compositions_ICCV_2023_paper.html.

Abdelrahman Shaker, Ahmed Heakl, Jaseel Muhammad, Ritesh Thawkar, Omkar Thawakar, Senmao Li, Hisham Cholakkal, Ian Reid, Eric P. Xing, Salman Khan, and Fahad Shahbaz Khan. Mobile-O: Unified Multimodal Understanding and Generation on Mobile Device, 2026. URL https://arxiv.org/abs/2602.20161.

Zihan Su, Hongyang Wei, Kangrui Cen, Yong Wang, Guanhua Chen, Chun Yuan, and Xiangxiang Chu. Generation Enhances Understanding in Unified Multimodal Models via Multi-Representation Generation, 2026. URL https://arxiv.org/abs/2601.21406.

Changyao Tian, Danni Yang, Guanzhou Chen, et al. Internvl-u: Democratizing unified multimodal models for understanding, reasoning, generation and editing, 2026. URL https://arxiv.org/abs/2603.09877.

Yujun Tong, Dongliang Chang, Zijin Yin, Xintong Liu, Yuanchen Fang, and Zhanyu Ma. Reversing the Flow: Generation-to-Understanding Synergy in Large Multimodal Models, 2026. URL https://arxiv.org/abs/2605.15792.

Chenlong Wang, Yuhang Chen, Zhihan Hu, Dongping Chen, Wenhu Chen, Sarah Wiegreffe, and Tianyi Zhou. Quantifying the Gap between Understanding and Generation within Unified Multimodal Models, 2026a. URL https://arxiv.org/abs/2602.02140.

Yuqing Wang, Zhijie Lin, Ceyuan Yang, Yang Zhao, Fei Xiao, Hao He, Qi Zhao, Zihan Ding, Fuyun Wang, Shuai Wang, Youliang Zhang, Haoqi Fan, and Xihui Liu. Representation Forcing for Bottleneck-Free Unified Multimodal Models, 2026b. URL https://arxiv.org/abs/2605.31604.

Xiwen Wei, Mark Nutter, Madhusudhanan Srinivasan, and Radu Marculescu. Pareto lora: Mitigating modality imbalance in unified multimodal models via pareto-optimal gradient integration, 2026. URL https://arxiv.org/abs/2606.17296.

Penghao Wu, Haiwen Diao, Weichen Fan, Lewei Lu, Dahua Lin, and Ziwei Liu. Uncovering Understanding-Generation Synergy in Native Unified Multimodal Models: From Representation, Task to System, 2026. URL https://arxiv.org/abs/2609.01607.

Jinheng Xie, Zhenheng Yang, and Mike Zheng Shou. Show-o2: Improved native unified multimodal models. In *Advances in Neural Information Processing Systems*, 2025. URL https://arxiv.org/abs/2506.15564.

Songsong Yu, Yuxin Chen, Ying Shan, and Yanwei Li. Semantic Generative Tuning for Unified Multimodal Models, 2026. URL https://arxiv.org/abs/2605.18714.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. In *Advances in Neural Information Processing Systems*, 2020. URL https://arxiv.org/abs/2001.06782.

Yubo Zhu, Zehan Kan, Jingyi Yang, Miaolin Chen, Jinbo Xing, Kai Zhu, Zijian Wang, Sheng Zhong, and Wei Tong. When Does Visual Generation Help Visual Understanding in Unified Multimodal Models?, 2026. URL https://arxiv.org/abs/2608.22174.

AI USE STATEMENT

Generative AI tools were used to assist with literature search and organization, drafting experimental scripts, and improving the clarity and readability of the manuscript. All AI-assisted outputs incorporated into the final submission were reviewed and verified by the authors. The authors take full responsibility for the content of this work.

ETHICS STATEMENT

This work uses an existing object–state dataset to study model-update decisions. It introduces no new human-subject study or collection of private personal data.

REPRODUCIBILITY STATEMENT

Algorithm 1 and Section 3 specify the probe–inference–commit procedure. Appendices B–D detail the finite-response interpretation, optimizer-state isolation, data partition, evaluation, and training settings. Appendices E–F report supplementary ablations and response diagnostics, and Appendix G lists the per-seed values underlying the aggregate results. Section 4 specifies macro-averaging, seed-level variability, and paired resampling.

A EXTENDED RELATED WORK

Unified computation and visual representations. Show-o2 and InternVL-U instantiate the shared understanding–generation setting evaluated here (Xie et al., 2025; Tian et al., 2026). Recent alternatives expose several different forms of unification. SenseNova-U1 builds on the native NEO-unify architecture (Diao et al., 2026); Lance combines shared context with separate capability pathways (Fu et al., 2026); and Omni-Diffusion models discrete tokens across modalities with masked diffusion (Li et al., 2026). TUNA uses a common continuous visual representation, UniAR shares a discrete visual tokenizer, and Representation Forcing predicts intermediate representations before pixel synthesis (Liu et al., 2025; Peng et al., 2026; Wang et al., 2026b). Mobile-O instead emphasizes compact vision-language–diffusion conditioning for deployment (Shaker et al., 2026). These architecture choices determine where cross-capability interaction can occur; our controlled experiments hold the backbone and adaptation channel fixed.

Synergy, interference, and alignment. Studies of transfer and native-model synergy show that shared computation can support either cooperation or competition, depending on task and architecture (Kang et al., 2026; Wu et al., 2026). Uni-X separates early and late modality-specific computation to address gradient conflict (Hao et al., 2026). DIVA separates shared and unique visual components, while LatentUMM aligns generation and re-encoding transformations (Lu et al., 2026; Luo et al., 2026). UNO supplies understanding supervision to generative representations; UniMRG and Semantic Generative Tuning use auxiliary visual generation to improve perception and generation (Liu et al., 2026c; Su et al., 2026; Yu et al., 2026). These findings motivate selective preservation of useful transfer. AMSP studies the narrower decision of whether to replace a particular source update under a source-retention constraint.

Gradient combination and low-rank transitions. PCGrad, CAGrad, GradNorm, and Pareto LoRA provide the gradient-based and balancing baselines evaluated here (Yu et al., 2020; Liu et al., 2021; Chen et al., 2018; Wei et al., 2026). Nash-MTL formulates gradient combination as bargaining, and FAMO balances relative loss decreases efficiently (Navon et al., 2022; Liu et al., 2023). MAdam identifies mismatches between a multi-objective solver's intended geometry and the update realized by Adam (Liu et al., 2026a). Low-rank adaptation introduces an additional distinction

between factors and effective weight updates: Riemannion addresses parameterization ambiguity, and Janus-LoRA analyzes interference introduced by interacting factor updates (Bogachev et al., 2025; Chen et al., 2026). These works support accounting for the actual optimizer transition. They are related methods, and only the explicitly listed baseline implementations enter our experimental comparison.

Experiments for decisions and selective commitment. AMSP's action-equivalence restriction relates to equivalence-class determination in noisy Bayesian active learning (Golovin et al., 2010). Bayesian algorithm execution chooses queries to infer an algorithm's output, while Action-BED designs experiments for downstream action loss (Neiswanger et al., 2021; Rossa et al., 2026). Robust decision-aware design further accounts for uncertain effects that can destabilize the preferred decision (Harikumar et al., 2026). AMSP instantiates a finite, source-constrained probe/action problem on disposable optimizer branches; these connections do not confer the other methods' theoretical guarantees. Selective VQA similarly links acceptance to evidence quality (Rodriguez & Rohrbach, 2026), whereas formal risk control requires its own calibration assumptions (Angelopoulos et al., 2024). Equal-budget lookahead remains essential: the benefit of choosing an informative probe must be assessed against directly testing feasible actions.

What the evaluation establishes. GapEval measures bidirectional capability gaps, Unison evaluates joint understanding–generation synergy, and IMUG-Bench examines multi-turn interleaved interaction (Wang et al., 2026a; Liu et al., 2026b; Meng et al., 2026). Generation-to-understanding inference can also use self-generated visual aids, whose usefulness depends on fidelity and reasoning difficulty (Tong et al., 2026; Zhu et al., 2026). These are complementary to our event-level update consequences. Our frozen object/state classifier follows the broad motivation for object-focused semantic evaluation illustrated by GenEval (Ghosh et al., 2023); it is neither the GenEval protocol itself nor a substitute for realism, diversity, or human evaluation.

B Mathematical Details and First-Order Limits

Finite action difference. Equation 1 gives

$$V_r(a; a_0 \mid x) = L_r(T_{a_0}(x)) - L_r(T_a(x)).$$

Positive intervention value is relative to the ordinary update and need not imply positive absolute receiver gain. If $g_s(a_0 \mid x) > 0$ and $\eta > 0$, skipping has $g_s(a_3 \mid x) = 0$ and violates source feasibility; merely avoiding receiver damage is insufficient. A reduced update qualifies only if it retains the prescribed source progress. All comparisons begin at the same complete state, including optimizer moments and random state.

Low-rank difference. For effective adapter weights $X_1 = B_1A_1$ and $X_2 = B_2A_2$,

$$\operatorname{rank}(X_1 - X_2) \le \operatorname{rank}(X_1) + \operatorname{rank}(X_2) \le 8$$

when each adapter has rank four. Thus, factor-native transitions preserve the adapter parameterization without requiring their effective weight *difference* to have rank four.

When first-order information is insufficient. Fix the receiver evaluation data, stochastic realization, and non-parameter evaluation state. Suppose L_r has a β-Lipschitz gradient on the parameter segments traversed by the compared actions. With $\delta\phi_a = \phi(T_a(x)) - \phi(x)$, the standard Taylor remainder bound gives

$$\Delta_r(a \mid x) = \nabla_\phi L_r(x)^\top \delta\phi_a + R_a, \qquad |R_a| \le \frac{\beta}{2}\|\delta\phi_a\|^2. \tag{17}$$

Subtracting the two action responses yields

$$\left|V_r(a; a_0 \mid x) - \nabla_\phi L_r(x)^\top(\delta\phi_{a_0} - \delta\phi_a)\right| \le \frac{\beta}{2}\left(\|\delta\phi_{a_0}\|^2 + \|\delta\phi_a\|^2\right). \tag{18}$$

The bound follows by applying Taylor's theorem to each action and the triangle inequality to $R_{a_0} - R_a$. When the first-order advantage lies within this remainder bound, it alone cannot certify which action improves the receiver more. This conditional argument does not imply that passive prediction always fails or provide a calibration guarantee for AMSP. The matched-pair and linear-response ablation results provide complementary empirical tests of the information used by the controller.

C Optimizer-State Isolation and Model-Specific Training

Snapshot, measure, discard, recompute. Algorithm 1 requires a snapshot of the trainable LoRA factors, AdamW moments, optimizer step, random state, relevant buffers, and any data-loader position affected by probing. Probe losses and predeclared response summaries may enter the decision; modified factors, optimizer buffers, and random draws must not enter the committed trajectory. After discarding the temporary branch, the selected transition is recomputed from the original snapshot exactly once. Restoring parameters alone would leave a different optimizer transition and would invalidate same-state action comparisons.

Decision-time information. Source data and receiver proxy-A support probe selection and hypothesis filtering. Calibration compositions determine response thresholds before final evaluation. Held-out B is used only after commitment or for retrospective event scoring. Candidate skip and inference fallback remain distinct: skip belongs to the action family, whereas fallback handles insufficient or inconsistent evidence. Their transition semantics must be fixed before evaluation rather than selected using B-side outcomes.

Differentiable generation path. The generation loss must remain differentiable with respect to the shared adapters. For InternVL-U, this requires gradients through the generation decoder's input hidden states into the shared contextual computation. Freezing decoder parameters does not remove this input-gradient requirement; an inference-only path under `no_grad` cannot implement the required training transition. The comparison concerns the common low-rank adaptation channel and does not establish a measured wall-clock or memory advantage for a particular sharding or probe-execution schedule.

D Dataset, Evaluation, and Training Settings

Chop & Learn contains 20 objects, seven cutting styles plus whole, 112 object–state compositions, and 1,338 images (Saini et al., 2023). Our composition-level partition uses 70 training-side, 17 calibration, and 25 final-test compositions (838/204/296 images). The training side is further separated into 54 source-learning and 16 proxy-A compositions; proxy-A does not enter ordinary source gradients. Final-test compositions form held-out B.

Table 4: Shared settings for the controlled comparisons unless a sensitivity configuration explicitly changes the parameter.

Setting	Value
Backbones	Show-o2-7B; InternVL-U-4B
Trainable channel	Rank-4 LoRA on q, k, v, o; base and heads frozen
Optimizer and learning rate	AdamW; 1.5×10^{-5}
Training duration	600 source-learning steps
Batch and precision	Effective batch 8; microbatch 1 with accumulation; BF16 base
Active comparison budget	Two additional response evaluations
Source retention / constructive retention	$\eta = 0.6$ / $\kappa = 0.8$
Receiver neutral band	$\epsilon = 0.003$ in normalized-loss units
Generation evaluation	16 noise-conditioned samples per composition and seed
Main training seeds	Five for Show-o2; three for InternVL-U
Hardware	Two NVIDIA RTX 4090 GPUs

Fair comparison and evaluation. All adapted baselines share the trainable channel and training dose. The U-only and G-only controls match the batch count of one capability; this is not a wall-clock match. Random probing retains the downstream controller, and lookahead receives the same additional response-evaluation budget as AMSP. Understanding predicts object and state in text, and generation uses a frozen object/state classifier. Scores are macro-averaged over final-test compositions. Generation accuracy therefore measures semantic correctness rather than realism, diversity, or human preference.

E Additional Ablations and Sensitivity

Table 5 includes the no-receiver-condition and one-probe variants omitted from the main ablation table. Removing receiver conditioning lowers H by 2.85 percentage points; using one probe lowers it by 0.71. The complete table also preserves the trade-off from removing constructive protection: generation slightly improves while understanding and the joint score decrease. These rows reuse the five-seed configurations reported in the per-seed appendix and are not independent replications.

Table 5: Complete Show-o2 ablations over five matched seeds. ΔH is relative to full AMSP (percentage points).

Variant	U-joint	G-joint	H	ΔH
AMSP	87.62 ± 2.50	77.70 ± 1.01	82.36 ± 1.56	+0.00
Passive replacement	86.55 ± 2.42	73.15 ± 1.23	79.28 ± 1.50	-3.08
Random probe + controller	84.95 ± 1.30	72.30 ± 2.15	78.10 ± 1.45	-4.26
No receiver condition	85.32 ± 2.22	74.45 ± 1.98	79.50 ± 1.81	-2.85
Same action for all states	86.06 ± 2.49	74.30 ± 2.30	79.73 ± 1.85	-2.63
No abstention	87.36 ± 2.25	76.15 ± 0.84	81.36 ± 1.13	-1.00
No constructive protection	85.33 ± 2.56	77.85 ± 1.24	81.40 ± 1.58	-0.95
One probe	87.36 ± 3.21	76.65 ± 0.84	81.64 ± 1.84	-0.71
Linear-response replacement	86.34 ± 2.42	75.80 ± 0.91	80.72 ± 1.45	-1.64

Sensitivity and repeated configurations. The sensitivity study uses three seeds and reuses the default and single-probe configurations. Its values must not be pooled with the five-seed main results as additional independent runs. The rank-4, two-probe default is retained across controlled comparisons; the best sensitivity setting is not substituted into the main comparison after evaluation.

F Additional Response and Misspecification Analyses

Prediction error and selective coverage. Figure 6 supplements the receiver-sign results with normalized-loss MAE over the 224 eligible Show-o2 events. AMSP and lookahead have nearly identical errors (0.00751 and 0.00755). Figure 7 reports accepted-set sign error as residual thresholds vary; the thresholds are fixed independently of held-out B. The curves concern response prediction, not the effect of retraining a control policy at each coverage level. The non-monotone empirical pattern does not establish a general risk guarantee.

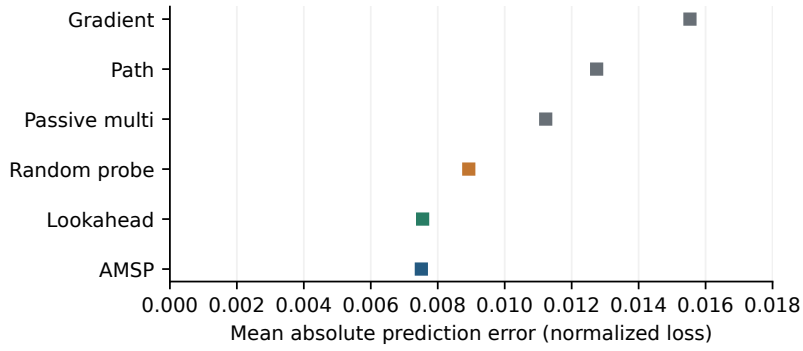


Figure 6: Mean absolute receiver-loss prediction error over 224 eligible Show-o2 events. Values are point estimates in normalized-loss units; smaller is better.

Matched pairs and model misspecification. Among 60 eligible Show-o2 matched pairs, lookahead orders 51 correctly and AMSP orders 45. Thus, AMSP does not dominate every response-prediction metric. Misspecification checks detect 16/32 injected Show-o2 cases and 5/20 InternVL-U cases, leaving substantial undetected errors. The undetected cases limit the reliability of the inferred action values.

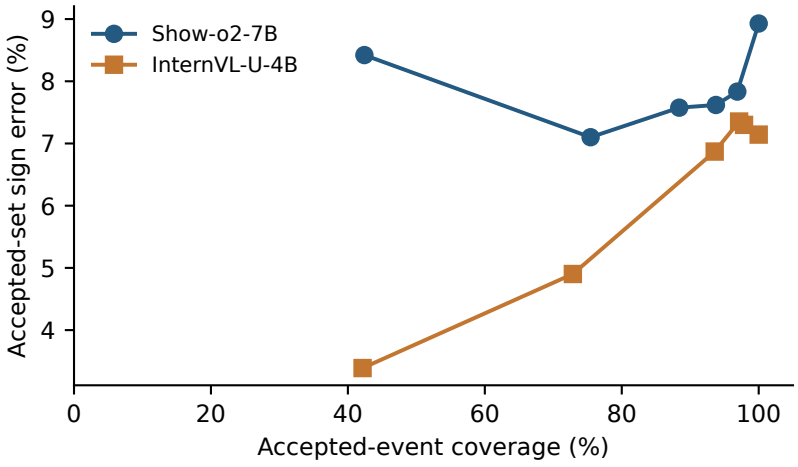


Figure 7: Receiver-sign error versus accepted-event coverage under fixed residual thresholds. These response-model diagnostics do not establish safety of the selected actions or a distribution-free risk guarantee.

G Full Per-Seed and Statistical Results

Each row reports macro-averaged understanding and generation accuracies over 25 held-out compositions and $H_s = 2U_sG_s/(U_s + G_s)$, computed within the same seed. Main-table H values average these per-seed scores; they are not harmonic means of the already averaged U and G columns. Frozen-reference rows occur once and have no sample SD. Sample SD across training seeds and paired seed-by-composition confidence intervals quantify different variation and are reported separately.

The configuration identifiers preserve the correspondence between main comparisons, ablations, and sensitivity records. Sensitivity records use three seeds; reused configurations are not additional independent repetitions.

Table 6: Per-seed joint understanding and generation accuracies and within-seed harmonic scores (%).

Model	Config	Seed	U-joint	G-joint	H
InternVL-U-4B	AMSP	17	88.64	82.25	85.32
InternVL-U-4B	AMSP	29	87.76	79.50	83.42
InternVL-U-4B	AMSP	43	88.94	80.50	84.51
InternVL-U-4B	FROZEN	0	80.91	69.00	74.48
InternVL-U-4B	JOINT	17	87.27	76.50	81.53
InternVL-U-4B	JOINT	29	86.42	72.50	78.85
InternVL-U-4B	JOINT	43	84.21	76.25	80.03
InternVL-U-4B	LOOKAHEAD	17	90.97	80.75	85.56
InternVL-U-4B	LOOKAHEAD	29	87.79	78.75	83.02
InternVL-U-4B	LOOKAHEAD	43	86.55	79.00	82.60
InternVL-U-4B	PARETO	17	89.97	81.50	85.53
InternVL-U-4B	PARETO	29	89.15	77.25	82.78
InternVL-U-4B	PARETO	43	86.55	78.75	82.46
InternVL-U-4B	PCGRAD	17	87.94	77.25	82.25
InternVL-U-4B	PCGRAD	29	86.76	75.75	80.88
InternVL-U-4B	PCGRAD	43	85.88	77.50	81.47
InternVL-U-4B	RANDOM	17	88.30	78.75	83.25
InternVL-U-4B	RANDOM	29	88.15	74.50	80.75
InternVL-U-4B	RANDOM	43	86.21	78.50	82.18
Show-o2-7B	ABL_NO_ABSTAIN	17	88.61	75.25	81.38
Show-o2-7B	ABL_NO_ABSTAIN	29	89.27	76.75	82.54
Show-o2-7B	ABL_NO_ABSTAIN	43	84.94	75.25	79.80
Show-o2-7B	ABL_NO_ABSTAIN	71	84.88	77.00	80.75
Show-o2-7B	ABL_NO_ABSTAIN	103	89.09	76.50	82.32
Show-o2-7B	ABL_NO_ACTIVE	17	88.27	72.50	79.61
Show-o2-7B	ABL_NO_ACTIVE	29	88.55	74.75	81.06
Show-o2-7B	ABL_NO_ACTIVE	43	83.61	71.50	77.08
Show-o2-7B	ABL_NO_ACTIVE	71	84.21	73.75	78.63
Show-o2-7B	ABL_NO_ACTIVE	103	88.09	73.25	79.99
Show-o2-7B	ABL_NO_NONLINEAR	17	87.94	75.75	81.39
Show-o2-7B	ABL_NO_NONLINEAR	29	87.88	76.50	81.80
Show-o2-7B	ABL_NO_NONLINEAR	43	83.24	74.25	78.49
Show-o2-7B	ABL_NO_NONLINEAR	71	84.21	76.25	80.03
Show-o2-7B	ABL_NO_NONLINEAR	103	88.42	76.25	81.89
Show-o2-7B	ABL_NO_POS_PROTECT	17	87.24	76.75	81.66
Show-o2-7B	ABL_NO_POS_PROTECT	29	87.55	79.50	83.33
Show-o2-7B	ABL_NO_POS_PROTECT	43	82.94	76.50	79.59
Show-o2-7B	ABL_NO_POS_PROTECT	71	82.18	78.00	80.04
Show-o2-7B	ABL_NO_POS_PROTECT	103	86.73	78.50	82.41
Show-o2-7B	ABL_NO_RECEIVER	17	85.88	74.75	79.93
Show-o2-7B	ABL_NO_RECEIVER	29	87.61	75.25	80.96
Show-o2-7B	ABL_NO_RECEIVER	43	83.58	71.00	76.78
Show-o2-7B	ABL_NO_RECEIVER	71	82.48	75.25	78.70
Show-o2-7B	ABL_NO_RECEIVER	103	87.06	76.00	81.16
Show-o2-7B	ABL_ONE_PROBE	17	87.94	77.00	82.11
Show-o2-7B	ABL_ONE_PROBE	29	89.61	77.25	82.97
Show-o2-7B	ABL_ONE_PROBE	43	83.61	75.25	79.21
Show-o2-7B	ABL_ONE_PROBE	71	84.55	76.50	80.32
Show-o2-7B	ABL_ONE_PROBE	103	91.09	77.25	83.60
Show-o2-7B	ABL_SAME_ACTION	17	87.94	75.00	80.96
Show-o2-7B	ABL_SAME_ACTION	29	88.24	72.50	79.60
Show-o2-7B	ABL_SAME_ACTION	43	83.94	71.50	77.22
Show-o2-7B	ABL_SAME_ACTION	71	82.85	75.25	78.87
Show-o2-7B	ABL_SAME_ACTION	103	87.33	77.25	81.98
Show-o2-7B	AMSP	17	88.94	77.25	82.68
Show-o2-7B	AMSP	29	89.94	78.75	83.97
Show-o2-7B	AMSP	43	84.94	76.25	80.36
Show-o2-7B	AMSP	71	84.88	77.75	81.16

Model	Config	Seed	U-joint	G-joint	H
Show-o2-7B	AMSP	103	89.42	78.50	83.61
Show-o2-7B	CAGRAD	17	84.55	70.75	77.03
Show-o2-7B	CAGRAD	29	88.21	72.25	79.44
Show-o2-7B	CAGRAD	43	82.58	70.25	75.92
Show-o2-7B	CAGRAD	71	83.21	74.75	78.75
Show-o2-7B	CAGRAD	103	86.64	74.25	79.97
Show-o2-7B	FROZEN	0	78.48	56.75	65.87
Show-o2-7B	GRADNORM	17	84.88	70.75	77.17
Show-o2-7B	GRADNORM	29	86.12	73.75	79.46
Show-o2-7B	GRADNORM	43	82.58	70.75	76.21
Show-o2-7B	GRADNORM	71	83.21	74.25	78.48
Show-o2-7B	GRADNORM	103	86.30	75.00	80.26
Show-o2-7B	G_ONLY	17	77.45	77.00	77.23
Show-o2-7B	G_ONLY	29	77.91	79.00	78.45
Show-o2-7B	G_ONLY	43	75.76	75.75	75.75
Show-o2-7B	G_ONLY	71	74.09	78.00	76.00
Show-o2-7B	G_ONLY	103	80.21	80.00	80.11
Show-o2-7B	JOINT	17	85.55	69.25	76.54
Show-o2-7B	JOINT	29	87.18	70.25	77.81
Show-o2-7B	JOINT	43	82.24	65.75	73.08
Show-o2-7B	JOINT	71	82.52	72.00	76.90
Show-o2-7B	JOINT	103	86.30	71.00	77.91
Show-o2-7B	LOOKAHEAD	17	86.88	73.75	79.78
Show-o2-7B	LOOKAHEAD	29	89.61	75.75	82.10
Show-o2-7B	LOOKAHEAD	43	85.64	73.75	79.25
Show-o2-7B	LOOKAHEAD	71	83.88	75.25	79.33
Show-o2-7B	LOOKAHEAD	103	86.97	77.75	82.10
Show-o2-7B	PARETO	17	87.61	74.00	80.23
Show-o2-7B	PARETO	29	89.94	75.25	81.94
Show-o2-7B	PARETO	43	83.94	73.50	78.37
Show-o2-7B	PARETO	71	85.55	76.25	80.63
Show-o2-7B	PARETO	103	86.30	78.50	82.22
Show-o2-7B	PCGRAD	17	84.55	70.50	76.89
Show-o2-7B	PCGRAD	29	87.18	72.00	78.87
Show-o2-7B	PCGRAD	43	82.58	67.75	74.43
Show-o2-7B	PCGRAD	71	82.85	73.00	77.61
Show-o2-7B	PCGRAD	103	85.97	73.50	79.25
Show-o2-7B	RANDOM	17	84.88	71.50	77.62
Show-o2-7B	RANDOM	29	86.12	72.75	78.87
Show-o2-7B	RANDOM	43	84.27	69.00	75.88
Show-o2-7B	RANDOM	71	83.18	74.25	78.46
Show-o2-7B	RANDOM	103	86.30	74.00	79.68
Show-o2-7B	SENS_01	17	88.24	74.50	80.79
Show-o2-7B	SENS_01	29	89.27	77.00	82.68
Show-o2-7B	SENS_01	43	83.88	75.25	79.33
Show-o2-7B	SENS_02	17	89.27	78.00	83.26
Show-o2-7B	SENS_02	29	90.94	79.00	84.55
Show-o2-7B	SENS_02	43	86.03	76.50	80.99
Show-o2-7B	SENS_03	17	88.61	77.25	82.54
Show-o2-7B	SENS_03	29	89.94	79.00	84.12
Show-o2-7B	SENS_03	43	84.94	76.50	80.50
Show-o2-7B	SENS_04	17	88.61	76.75	82.25
Show-o2-7B	SENS_04	29	88.94	77.50	82.83
Show-o2-7B	SENS_04	43	84.61	75.00	79.51
Show-o2-7B	SENS_05	17	88.94	77.00	82.54
Show-o2-7B	SENS_05	29	89.94	78.75	83.97
Show-o2-7B	SENS_05	43	84.94	75.25	79.80
Show-o2-7B	SENS_06	17	88.94	76.25	82.11
Show-o2-7B	SENS_06	29	89.94	78.25	83.69
Show-o2-7B	SENS_06	43	84.61	75.25	79.65
Show-o2-7B	SENS_07	17	88.61	76.75	82.25
Show-o2-7B	SENS_07	29	88.27	78.25	82.96
Show-o2-7B	SENS_07	43	83.94	75.50	79.50
Show-o2-7B	SENS_08	17	88.27	76.75	82.11
Show-o2-7B	SENS_08	29	88.94	78.50	83.39
Show-o2-7B	SENS_08	43	84.94	74.25	79.24
Show-o2-7B	SENS_09	17	88.24	77.25	82.38
Show-o2-7B	SENS_09	29	89.61	78.50	83.69
Show-o2-7B	SENS_09	43	84.61	76.25	80.21
Show-o2-7B	U_ONLY	17	89.61	54.75	67.97
Show-o2-7B	U_ONLY	29	90.94	53.00	66.97
Show-o2-7B	U_ONLY	43	86.33	50.50	63.72
Show-o2-7B	U_ONLY	71	88.24	54.25	67.19
Show-o2-7B	U_ONLY	103	89.42	52.75	66.36