SELECT2ADAPT: SEMANTICALLY CONSTRAINED PREFERENCE TRANSFER FOR SINGLE-SAMPLE TEXT-TO-IMAGE GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Selecting a preferred image from multiple candidates does not itself teach a generator to produce preferred outputs directly. Select2Adapt formulates this transfer through an alignment-anchored teacher and a bounded residual text-conditioning adapter in an otherwise frozen diffusion model. The teacher establishes score-based semantic eligibility before applying quality preferences: it promotes an alternative only when semantic thresholds and quality gains across fixed evaluation contexts pass, and otherwise retains the alignment reference. Selected and preservation targets define a denoising objective with base-prediction consistency and residual regularization, with intervention confined to the conditional branch. We derive conditional score-error and sampling-sensitivity bounds, and separate evaluation of this supervision from evaluation of single-sample generation. In an audited development study with 48 prompts and 384 shared candidates, a calibrated PickScore teacher changes eight selections and improves independently assessed MUSIQ by 0.3973 (95% interval [0.1036, 0.7700]). However, its BLIP interval fails the predefined semantic noninferiority criterion, and no shape prompt is promoted. These results identify a quality signal and a remaining supervision bottleneck; they do not establish semantic preservation or successful transfer. The framework and evaluation protocol make explicit what must be verified before multi-candidate preferences can support a trained generator that returns one sample without reranking.

1 INTRODUCTION

Latent diffusion models and diffusion transformers provide scalable image synthesis (Rombach et al., 2022; Peebles & Xie, 2023), but stochastic outputs still vary in object presence, attribute binding, and visual quality. Selecting from several candidates exploits this variation at the cost of repeated generation and scoring. A useful selector is therefore only one part of an efficient generation system: transferring its preferences to a generator is a distinct learning problem.

Alignment and appearance can favor different images. A visually attractive candidate may omit an object or attach an attribute to the wrong entity. A scalar combination of alignment and quality makes the resulting tradeoff depend on score scales and weights. Select2Adapt instead establishes semantic eligibility before applying a quality preference. This constrains measured scores; it does not certify that the prompt is satisfied.

The framework connects three stages. An offline teacher selects targets relative to an alignment anchor. Promoted targets and source-sample preservation targets supervise a small residual adapter on frozen text features. The resulting model is intended to generate one image per prompt and seed, using the adapter only on the conditional branch. Denoising, preservation, and residual constraints specify the learning mechanism, while separate teacher and student assessments prevent reranking gains from being mistaken for transfer gains.

We study the first empirical question on shared candidate pools. The resulting teacher has a positive development quality signal: across all 48 prompts, its MUSIQ improvement over the alignment anchor has an interval above zero. Its semantic noninferiority and facet-coverage criteria nevertheless fail. This is a useful distinction: numerically reliable scoring and improved aggregate quality are

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.

insufficient to qualify training supervision. The study uses exposed development material, not a held-out test, and no adapted-generator efficacy is claimed.

Our contributions are threefold. First, we formulate an eligibility-first teacher with deterministic fallback and characterize its finite-pool score guarantees. Second, we specify bounded conditioning adaptation, its selected-target objective, and conditional bounds on the propagation of conditioning changes. Third, we provide an audited selection study and a transfer-evaluation protocol that separate perceptual quality, semantic preservation, intervention coverage, and single-sample generation. Historical selection results and diagnostic image pairs provide supplementary context; they are not adapted-generator outputs.

2 Related Work

Selection and compositional assessment. CLIP learns transferable image–text representations (Radford et al., 2021), and CLIPScore adapts their similarity to reference-free compatibility assessment (Hessel et al., 2021). PickScore, ImageReward, and HPS v2 learn signals from human preferences, while VQAScore uses image-conditioned text probabilities to assess compatibility (Kirstain et al., 2023; Xu et al., 2023; Wu et al., 2023; Lin et al., 2025). These signals support ranking but need not agree on visual quality and prompt faithfulness. Rich human feedback further distinguishes overall ratings from localized defects and text–image misalignment (Liang et al., 2024). GenAI-Bench and GenAI-Rank examine compositional assessment and black-box ranking (Li et al., 2024a); GenEval probes objects, counts, colors, positions, and attribute binding (Ghosh et al., 2023), while T2I-CompBench evaluates attribute binding, relationships, and complex compositions (Huang et al., 2023). Feynman–Kac steering uses multiple diffusion trajectories and intermediate resampling at inference (Singhal et al., 2025). Select2Adapt instead targets one sampled image after offline adaptation.

Preference transfer to diffusion models. RAFT fine-tunes generative models on reward-ranked samples, and GORS selects training samples to improve compositional generation (Dong et al., 2023; Huang et al., 2023). DDPO applies reinforcement learning to diffusion sampling; DRaFT backpropagates differentiable rewards through sampling; ReFL uses ImageReward feedback for model adaptation (Black et al., 2024; Clark et al., 2024; Xu et al., 2023). Direct Preference Optimization learns from preference pairs in language modeling, and Diffusion-DPO extends preference optimization to diffusion models (Rafailov et al., 2023; Wallace et al., 2024). These approaches establish several routes from evaluation signals to a trained generator. Select2Adapt studies eligibility-filtered targets and a restricted adaptation boundary. Its selected-target denoising objective is filtered supervised learning; rejected candidates do not enter a pairwise likelihood loss.

Constrained decisions and selective prediction. Constrained optimization distinguishes feasibility from objective value (Boyd & Vandenberghe, 2004). Our feasible set is a finite candidate pool; no convexity of scores or images is assumed. Selective classification studies a risk–coverage tradeoff by allowing a predictor to abstain (Geifman & El-Yaniv, 2017). Here the teacher can decline a promotion but still returns its anchor. Promotion coverage therefore measures intervention frequency, not prediction coverage, and selective-risk guarantees do not transfer to this rule. Both connections motivate explicit constraints and coverage reporting without turning a proxy threshold into a semantic certificate.

Parameter-efficient and conditioning adaptation. Adapters and LoRA reduce trainable capacity through added modules or low-rank weight updates (Houlsby et al., 2019; Hu et al., 2022). In text-to-image generation, textual inversion learns concept embeddings for a frozen model (Gal et al., 2023), whereas ControlNet and T2I-Adapter introduce modules that incorporate additional spatial conditions (Zhang et al., 2023; Mou et al., 2024). These settings differ from transferring prompt-level preferences without extra inference inputs. TextCraftor fine-tunes the text encoder using quality rewards and a soft CLIP alignment term; TexForce applies reinforcement learning to that encoder (Li et al., 2024b; Chen et al., 2025). Select2Adapt freezes both the text encoder and denoiser, learning a bounded residual on conditional text features from teacher-selected targets. The combination of target selection and conditioning adaptation requires comparison with both families of methods; a feature bound alone provides no output-level semantic guarantee.

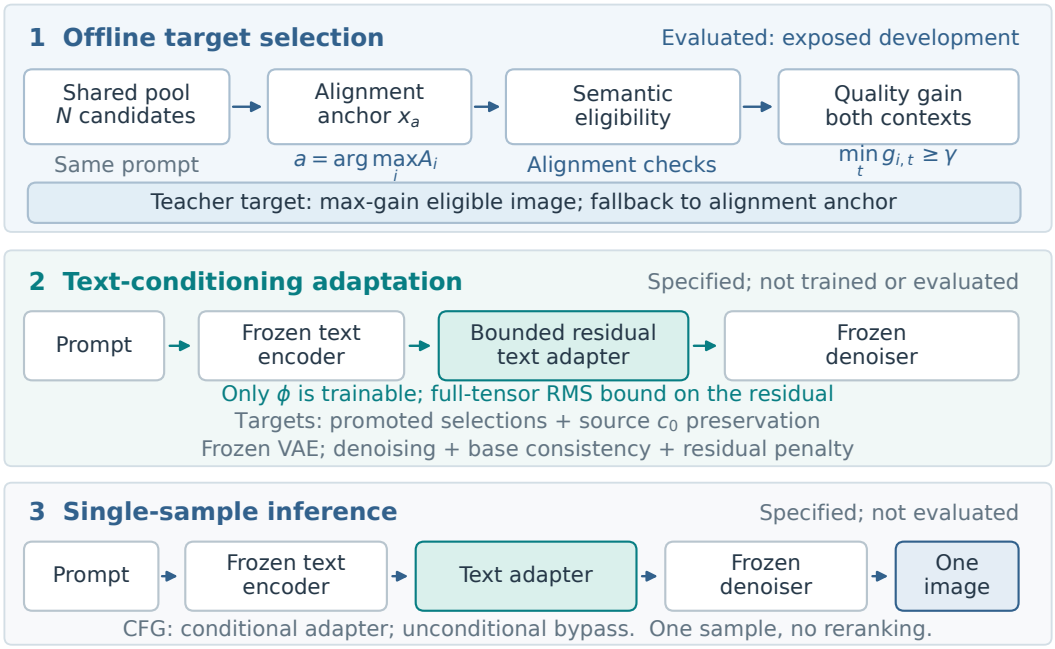

Figure 1: Select2Adapt computation and evidence boundary. Offline selection supplies promoted targets and identifies non-promoted prompts. Adaptation uses those promotions and candidate-zero preservation targets, with a bounded text-conditioning residual and frozen backbone. Deployment uses one sampled trajectory and no reranking. The selection stage is evaluated here; adapter efficacy and single-sample transfer remain unestablished.

3 SELECT2ADAPT

Figure 1 separates the evaluated selection stage from the adaptation and deployment computation. The text encoder, autoencoder, and denoiser remain frozen; the only trainable component is the text-conditioning adapter. The formulation couples an explicit target-selection rule to a restricted adaptation boundary.

3.1 SEMANTIC ELIGIBILITY BEFORE QUALITY PREFERENCE

For prompt p, generate $\mathcal{X}(p) = \{x_i = G_0(p; z_i)\}_{i=0}^{N-1}$ with recorded seeds and a fixed sampler. Let A_i be a whole-prompt alignment score and $A_i^{(r)}$ an alignment score for semantic fragment r. Higher scores indicate stronger compatibility. The anchor is

$$a = \arg\max_i (A_i, -i), \tag{1}$$

with the smaller candidate index resolving ties. This best-of-N reference must not be confused with a single-sample base output.

A fixed semantic predicate can combine absolute and relative score thresholds:

$$S_{i,a} = \mathbf{1}\{A_i \geq \max(\tau_A, \eta_A A_a)\} \prod_{r=1}^{R} \mathbf{1}\{A_i^{(r)} \geq \max(\tau_r, \eta_r A_a^{(r)})\}. \tag{2}$$

The numerical study uses two object–attribute fragments. Other semantic decompositions define different teachers and require separate validation. Relative thresholds permit score loss; neither these thresholds nor high score values imply calibrated correctness probabilities.

Let $q_t(p, x_i)$ be a quality-preference score under fixed evaluation context $t \in \mathcal{T}$. Define paired gains against the same anchor and their conservative aggregate:

$$g_{i,t} = q_t(p, x_i) - q_t(p, x_a), \qquad g_i = \min_{t \in \mathcal{T}} g_{i,t}, \qquad \mathcal{E}(p) = \{i \neq a : S_{i,a} = 1,\ g_i \geq \gamma\}, \tag{3}$$

where $\gamma > 0$. The minimum of paired gains is generally different from subtracting two minimum scores. Requiring $g_i \geq \gamma$ imposes a positive margin in every context. Contexts from the same model can share biases; their agreement tests stability, not independent validity. Quality scores must also be characterized operationally: a prompt-conditioned preference score can respond to both relevance and appearance.

The teacher selects deterministically:

$$j = \arg\max_{i \in \mathcal{E}(p)} (g_i, A_i, -i), \qquad y_T(p) = \begin{cases} x_j, & \mathcal{E}(p) \neq \varnothing, \\ x_a, & \mathcal{E}(p) = \varnothing. \end{cases} \tag{4}$$

The conditional definition avoids evaluating j for an empty set. Invalid alignment prevents reliable anchor construction. The present experiment requires complete, finite scores and stops selection on an unusable candidate; it never drops an inconvenient candidate or prompt. Empty eligibility with valid scores retains the reference. Pool identities, all failed checks, scorer versions, and chosen targets are recorded before assessment.

3.2 Bounded Residual Conditioning

For frozen text features $H = E_0(p) \in \mathbb{R}^{L \times d}$, a small network produces $R_\phi = f_\phi(H)$. With residual scale one, a per-sample bound is

$$B(p) = \rho \|H\|_F, \qquad \Delta_\phi(H) = R_\phi \min\left\{1, \frac{B(p)}{\|R_\phi\|_F}\right\}, \qquad H_\phi = H + \Delta_\phi(H), \tag{5}$$

taking $\Delta_\phi = 0$ when $R_\phi = 0$ or $B(p) = 0$. Because input and residual have the same shape, this is equivalent to an RMS-ratio bound over the full conditioning tensor. It is not a bound restricted to non-padding tokens. The PixArt implementation uses 120×4096 features and a $4096 \to 64 \to 4096$ SiLU MLP (528,448 trainable parameters), with no dropout, zero-initialized output layer, and $\rho = 0.10$. It rechecks the effective output-minus-input residual after low-precision addition. These are implementation properties, not evidence of improved images.

Only ϕ is optimized. Frozen denoiser weights still transmit gradients through their conditioning input, and zero residual recovers base conditioning. For classifier-free guidance (Ho & Salimans, 2022),

$$\hat{\epsilon}_\phi = \epsilon_0(z_t, t, H_\varnothing) + w\left[\epsilon_0(z_t, t, H_\phi) - \epsilon_0(z_t, t, H_\varnothing)\right]. \tag{6}$$

The unconditional branch bypasses the adapter. The configured empty negative prompt and any conditioning-dropout behavior must retain this branch distinction. A feature bound provides no output-level semantic certificate: nonlinear denoising and guidance can amplify a small conditioning change.

3.3 Selector-Supervised Objective and Deployment

Promoted prompt–target pairs form $\mathcal{D}_+$. For non-promoted prompts, the training implementation uses the recorded source candidate x_0 as the preservation target, forming $\mathcal{D}_0$. This source-preservation rule differs from the teacher's alignment-anchor fallback x_a used in selection assessment; replacing x_0 by x_a would change the training distribution. A frozen VAE encodes target x into z_x; the fixed scheduler samples $z_t = \alpha_t z_x + \sigma_t \epsilon$. For the noise-prediction implementation (Ho et al., 2020),

$$\ell_{\text{den}}(\phi; p, x) = \mathbb{E}_{t,\epsilon}\,\text{MSE}\left(\epsilon_0(z_t, t, H_\phi), \epsilon\right). \tag{7}$$

The specified sampler balances promoted and preservation roles 1:1. Base consistency applies only to preservation examples, using identical noisy latents, timesteps, and noise for both predictions:

$$\mathcal{L}_{\text{den}} = \tfrac{1}{2}\mathbb{E}_{\mathcal{D}_+}\ell_{\text{den}} + \tfrac{1}{2}\mathbb{E}_{\mathcal{D}_0}\ell_{\text{den}}, \tag{8}$$

$$\mathcal{L}_{\text{bc}} = \mathbb{E}_{\mathcal{D}_0,t,\epsilon}\,\text{MSE}\left(\epsilon_0(z_t, t, H_\phi), \text{sg}[\epsilon_0(z_t, t, H)]\right), \tag{9}$$

$$\mathcal{L}_{\text{res}} = \mathbb{E}_{\mathcal{B}}\frac{\sum_{b\in\mathcal{B},\ell} M_{b\ell}\|\Delta_\phi(H_b)_\ell\|_2^2}{d\sum_{b\in\mathcal{B},\ell} M_{b\ell}}, \qquad \mathcal{L} = \mathcal{L}_{\text{den}} + \lambda_{\text{bc}}\mathcal{L}_{\text{bc}} + \lambda_{\text{res}}\mathcal{L}_{\text{res}}. \tag{10}$$

Here M marks valid text tokens, $\mathcal{B}$ is a minibatch, and sg denotes stop-gradient. MSE averages over the latent elements; the residual penalty averages jointly over valid tokens in the minibatch.

The residual penalty is a masked mean squared update, not a relative norm penalty; its mask differs from the full-tensor bound in Equation 5. The documented configuration uses $\lambda_{\mathrm{bc}} = 1$ and $\lambda_{\mathrm{res}} = 0.01$. Both target roles and nonempty valid-token masks are required for this configuration. The implemented denoiser emits noise and variance channels; these losses use only the noise-prediction half. Alternative prediction parameterizations require corresponding targets.

Target-distribution interpretation. Let $\pi \in (0, 1)$ be the promoted fraction in a prospective training cohort. Uniform prompt sampling weights the two roles by $(\pi, 1 - \pi)$, whereas Equation 8 intentionally weights them equally. Relative to uniform sampling, promoted and preservation examples receive factors $1/(2\pi)$ and $1/[2(1 - \pi)]$. This changes the learning distribution; it is not an unbiased estimate of the uniform-cohort denoising risk. If either role is empty, the stated balanced objective is undefined and requires a separately specified training rule. Coverage therefore affects both the diversity of supervision and its effective weighting. A matched-control experiment must hold these weights fixed to distinguish target quality from oversampling.

This objective is selected-target supervised denoising, not pairwise preference optimization. Rejected images do not enter a likelihood-ratio loss, and minimizing the objective does not guarantee that fresh samples pass the teacher's gates. After training, the intended deployment returns $G_\phi(p; z)$ from one fresh latent trajectory without scoring other candidates. One sample means one trajectory, not one denoising step. The present teacher study does not train this adapter or establish its latency, quality, or transfer performance.

3.4 What the Constraints Guarantee

The following elementary bounds separate properties of selection from properties of adaptation. They assume finite scores and constants; they are not semantic certificates. Proofs appear in Appendix A.

Proposition 1 (finite-pool selection and score error). Fix a nonempty finite candidate pool, a nonempty finite context set, the alignment scores, their anchor a, and the semantic predicate in Equation 2. Let $\hat{q}_{i,t}$ be the scores used for selection, $\hat{g}_i = \min_t(\hat{q}_{i,t} - \hat{q}_{a,t})$, and $\hat{\mathcal{E}}$ the resulting eligible set with margin $\gamma > 0$. Write s for the selected index, including $s = a$ on fallback. On $\mathcal{F} = \{a\} \cup \hat{\mathcal{E}}$, the teacher maximizes $\hat{g}_i$, with $\hat{g}_a = 0$. Suppose reference scores $q^*_{i,t}$ satisfy $|\hat{q}_{i,t} - q^*_{i,t}| \le \varepsilon_q$ for every candidate and context, where $\varepsilon_q \ge 0$. Defining $g^*_i = \min_t(q^*_{i,t} - q^*_{a,t})$, we have

$$g^*_s \ge \begin{cases} \gamma - 2\varepsilon_q, & s \ne a, \\ 0, & s = a, \end{cases} \qquad \max_{i \in \mathcal{F}} g^*_i - g^*_s \le 4\varepsilon_q. \tag{11}$$

Thus $\gamma > 2\varepsilon_q$ guarantees a positive reference-score gain in every context on a promotion. The regret bound is only over the *observed* feasible set $\mathcal{F}$; it does not recover candidates excluded by erroneous semantic scores. The fallback anchor need not satisfy the absolute semantic floors. No error bound relative to human judgments is established by the blur calibration.

Corollary (semantic-score perturbations at a fixed anchor). Keep the observed alignment anchor a fixed. Suppose the measured whole-prompt scores obey $|A_j - A^*_j| \le \varepsilon_A$ for $j \in \{i, a\}$, where A^* denotes reference scores and $\varepsilon_A \ge 0$. For finite scores, thresholds, and $\eta_A \ge 0$, a candidate passing the whole-prompt component of Equation 2 satisfies

$$A^*_i \ge \max\{\tau_A - \varepsilon_A,\ \eta_A A^*_a - (1 + \eta_A)\varepsilon_A\}. \tag{12}$$

For each fragment, the same statement holds with its own scores, threshold, nonnegative relative multiplier and error bound. This quantifies how candidate and anchor errors weaken each measured constraint. The comparison remains relative to the *same observed anchor*, which need not maximize the reference alignment scores. The result applies only when the corresponding measured gate passes; fallback does not imply that the anchor meets an absolute floor. Without an empirically justified error bound to a meaningful semantic reference, neither the corollary nor high measured scores certify human semantic correctness. Thus the result characterizes conditional robustness rather than validating the chosen thresholds.

Proposition 2 (conditioning and finite-step sensitivity). Fix p and let $h = H$, $h' = H_\phi$, with $\|h' - h\|_F \le B = \rho\|H\|_F$ and finite $\rho \ge 0$ (here $\rho = 0.10$). At step k, write $U_k(z) = \epsilon_0(z, t_k, H_\varnothing)$ and $V_k(z, h) = \epsilon_0(z, t_k, h)$. Assume finite nonnegative constants L_k^u, L_k^z, L_k^h such that U_k is L_k^u-Lipschitz in the latent, $V_k(\cdot, h')$ is L_k^z-Lipschitz, and $\|V_k(z, h') - V_k(z, h)\|_2 \le L_k^h\|h' - h\|_F$ at the compared states. Define guided predictions $F_k^v = (1 - w_k)U_k + w_k V_k(\cdot, v)$ for any real w_k. Then

$$\|F_k^{h'}(z) - F_k^h(z)\|_2 \le |w_k| L_k^h B. \tag{13}$$

For the explicit recursion $z_{k+1}^v = \alpha_k z_k^v + \beta_k F_k^v(z_k^v) + \sigma_k \xi_k$, couple both runs with identical initial latent and identical noise ξ_k. For a nonnegative integer K and finite real coefficients, set

$$C_k = |\alpha_k| + |\beta_k|\big(|1 - w_k| L_k^u + |w_k| L_k^z\big),$$
$$D_k = |\beta_k w_k| L_k^h B, \qquad \|z_K^{h'} - z_K^h\|_2 \le \sum_{i=0}^{K-1} D_i \prod_{j=i+1}^{K-1} C_j. \tag{14}$$

Empty products equal one; the bound is zero for $K = 0$ or $B = 0$. This explicit one-step abstraction does not certify the current multistep sampler; a multistep extension requires an augmented state and corresponding bounds. It does not assume contraction, so products can amplify the perturbation. The constants have not been measured for our model, and an image-level metric bound would require additional decoder and metric regularity.

Selection does not imply transfer. A teacher can prefer an eligible candidate with a positive score gain while a zero-residual student still equals the base generator for every seed. The adapter bound and successful selection then hold, but student improvement is zero. A transfer claim therefore requires evidence linking training to independently assessed generated outputs, beyond either proposition.

4 Audited Teacher Development

4.1 Shared Pools, Frozen Rules, and Independent Assessment

Cohort and generation. We use 48 previously exposed development prompts, with 16 each for color, shape, and texture binding. Each prompt has two object–attribute fragments and eight frozen candidates, giving 384 images. The pools were expanded from four to eight before this study; no new images were generated for this comparison. Generation uses PixArt-XL-2 at 512×512 (Chen et al., 2024), 20 inference steps, guidance 4.0, and an empty negative prompt. All six selectors see identical images. Historical assessments and failures had already been inspected, so freezing the new rules does not make this a held-out or blind experiment.

Selection models and calibration. Whole-prompt and fragment scores are the previously recorded Qwen3-VL compatibility scores (Bai et al., 2025). We retain $\tau_A = \tau_r = 0.5$, $\eta_A = 0.25$, and $\eta_r = 0.9$. PickScore (Kirstain et al., 2023) supplies native scaled cosine logits under two contexts: the original prompt and `An image of {prompt}.` No cross-prompt absolute quality floor is imposed. Twelve fixed clean/blur controls, four per facet, use Gaussian blur with radius three. Before selecting targets, set

$$\gamma = \max\left\{0, \max_c \min_{t \in \mathcal{T}} \left[q_t(x_c^{\text{blur}}) - q_t(x_c^{\text{clean}})\right]\right\} + 10^{-4}. \tag{15}$$

Only canonical clean-first control logits enter calibration. The resulting $\gamma = 0.2882565094$ excludes every consistently preferred blur among these controls. The added 10^{-4} is an engineering margin; this calibration is neither a human preference threshold nor an out-of-sample guarantee.

Six comparison arms. We compare fixed candidate zero, the alignment anchor, preference-only ranking by g_i, a scalar rule $0.5A_i + 0.5\,\mathrm{softmax}_{\mathcal{X}}(g)_i$, the new constrained teacher, and the preceding all-eligible quality teacher. Thus the preference-only arm uses the same anchor-relative quality statistic but omits semantic eligibility; the scalar arm uses an explicitly fixed, scale-dependent normalization. Ties use the smaller index, except the teacher's additional alignment tie-break. The preceding teacher's decisions are copied unchanged, not reoptimized against the new assessments.

Table 1: Complete-cohort selection changes relative to the alignment anchor. All rows use the same 48 prompts; 95% intervals retain unchanged prompts. Changes are numbers of selected candidate identities that differ from the anchor. These are development selection outcomes, not generated-student results.

Selector	Changes	Δ MUSIQ [95% interval]	Δ BLIP [95% interval]
Fixed candidate zero	41	−0.6711 [−1.8029, 0.3947]	−0.2424 [−0.3437, −0.1426]
Alignment anchor	0	0 (reference)	0 (reference)
Preference-only	40	+0.2700 [−0.3967, 0.9409]	−0.2382 [−0.3261, −0.1536]
Scalar mixture	22	+0.2089 [−0.2834, 0.7209]	−0.0836 [−0.1463, −0.0263]
Preceding teacher	2	+0.0869 [0.0000, 0.2407]	+0.0003 [−0.0008, 0.0016]
Constrained teacher	8	+0.3973 [0.1036, 0.7700]	−0.0032 [−0.0266, 0.0124]

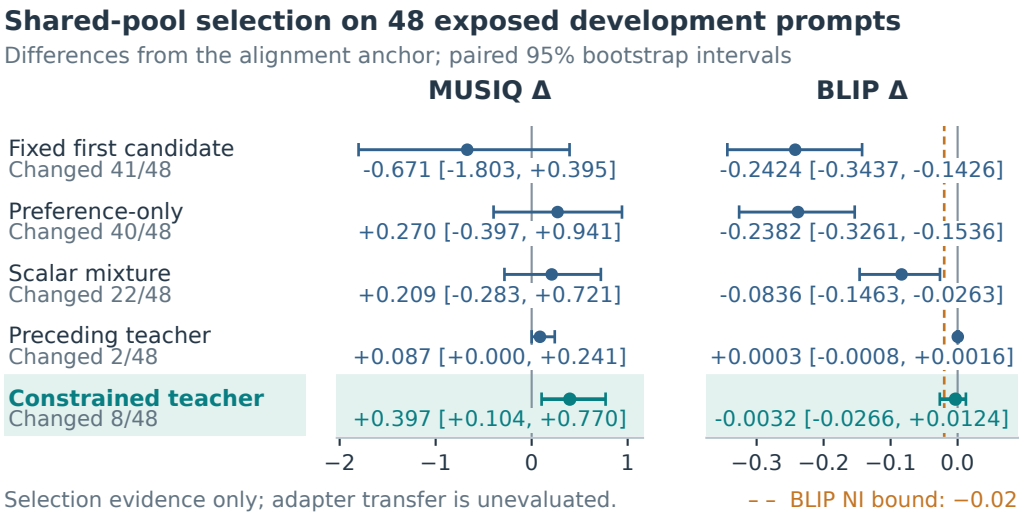

Figure 2: Paired development effects on the same full cohort. Points and intervals use the frozen 48-prompt bootstrap; the semantic margin is −0.02. The constrained teacher improves MUSIQ, but its BLIP lower bound crosses the margin. The two metrics have different units and must be interpreted separately.

Assessment and uncertainty. MUSIQ-KonIQ assesses perceptual quality (Ke et al., 2021); the frozen BLIP-based attribute assessment supplies a separate semantic proxy (Li et al., 2022; Huang et al., 2023). Neither model selects targets. Model separation does not imply independence of model training data or unbiased measurement. Reused assessment files cover all 384 candidates, including one previously documented object-name parsing truncation. We retain that case. The primary paired unit is the prompt, including all unchanged selections. We compute 20,000 stratified bootstrap replicates with 16 prompts per facet, PCG64 seed 1234, and linear 95% quantiles. The 16 reported contrasts comprise five arms versus alignment and teacher versus preference-only, scalar, and preceding teacher, for two metrics. Intervals are descriptive, without multiplicity correction.

4.2 Quality Signal and Failed Joint Qualification

Technical validity. The six-GPU scoring run comprises 768 candidate/context calls, 48 control/context/order calls, and 20 actual repeats: 836 calls and 892 image inputs. All outputs are finite and untruncated. Feature-formula and native-forward logits agree exactly in all 836 records; all 24 reversed-order checks and 20 repeats also agree exactly. Ten of 12 controls prefer the clean image in both contexts and both orders, meeting the fixed technical minimum. Numerical stability is an expected property of direct scoring and does not establish preference validity.

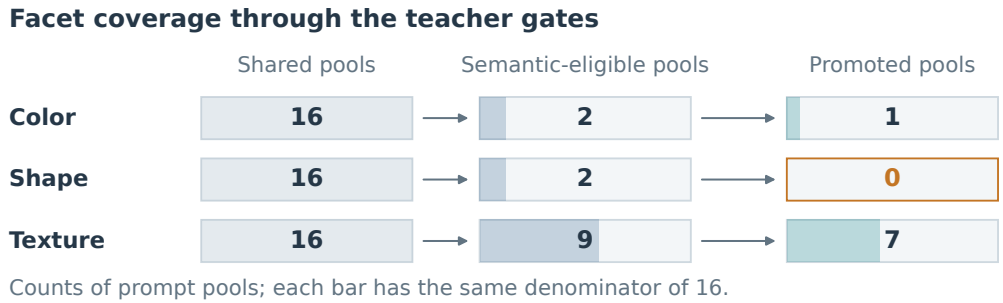

Figure 3: Facet coverage through the frozen teacher gates on exposed development data. Each facet contains 16 prompt pools with eight candidates per pool. A semantic-eligible pool contains at least one non-anchor candidate satisfying all whole-prompt and atomic alignment thresholds, before the preference margin is applied. Color, shape, and texture retain 2, 2, and 9 such pools; the joint rule promotes 1, 0, and 7 pools, respectively. Counts are recomputed from the original candidate scores and frozen decisions; they diagnose selection coverage, not adapted-generator performance.

Coverage and complete-cohort effects. For any endpoint S, let $d_i = S(p_i, y_T(p_i)) - S(p_i, x_{a_i})$, where a_i is prompt i's anchor. Let $\mathcal{P}$ index the m promoted prompts among n prompts. Since fallback gives $d_i = 0$, the complete-cohort mean decomposes exactly as

$$\frac{1}{n}\sum_{i=1}^{n} d_i = \frac{m}{n}\left(\frac{1}{m}\sum_{i\in\mathcal{P}} d_i\right), \qquad m > 0. \tag{16}$$

Here $m/n = 8/48$. Reporting only changed prompts would remove this intervention-frequency factor and answer a different question. When $m = 0$, the full-cohort effect is zero and the changed-prompt average is undefined. We therefore retain all prompts in both point estimates and bootstrap replicates, including facets with no interventions.

Teacher quality. The teacher changes eight pools: one color, zero shape, and seven texture. Its full-cohort MUSIQ gain is 0.3973, with interval $[0.1036, 0.7700]$ (Table 1). Relative to the preceding teacher, the gain is 0.3105 with interval $[0.0762, 0.5732]$. The corresponding comparisons with preference-only and scalar selection include zero (Appendix C), so superiority over those quality endpoints is not established. Seven texture promotions improve MUSIQ, while the color promotion decreases it; the complete-cohort interval includes all eight changes.

Semantic and coverage limits. The fixed development screen requires MUSIQ's lower bound above zero, BLIP's lower bound at least -0.02, at least six promotions, and at least one promotion per facet. BLIP changes by -0.00324 with interval $[-0.02658, 0.01236]$: its small mean loss does not satisfy noninferiority. Zero shape promotions also fail coverage. The joint screen therefore fails, even though technical and quality checks pass. The teacher has better BLIP outcomes than preference-only and scalar selection in this cohort, but that comparison cannot substitute for preservation relative to the alignment anchor.

Failure mechanisms. Figure 3 localizes the coverage bottleneck. Only 30 alternatives in 13 pools meet semantic eligibility before the quality check. Of the seven shape-eligible alternatives, six have negative conservative preference gain; the only positive one has gain 0.01932, below γ. The color promotion for "a yellow bicycle and a purple hairbrush" passes all score-based gates but loses 0.4821 BLIP and 1.6207 MUSIQ. It shows why score eligibility cannot certify correct object identity or attribute binding. We retain this case and the original thresholds; excluding it or relaxing the quality margin would define a different experiment.

Audit and scope. An independent CPU implementation checks all 836 numerical records, 48 six-arm choices, 336 alternative-candidate checks, and all 16 paired endpoints against the frozen artifacts. The audit agrees with the failed joint screen. Historical GenEval results in Appendix D and qualitative semantic contrasts in Appendix E provide additional context. Neither those records nor the present selection outcomes establish trained one-sample improvement.

5 Evaluating Preference Transfer

The next empirical stage must freeze training, development, and test prompts before fitting thresholds or choosing checkpoints. Exposed development pools cannot establish held-out student generalization. For an independent endpoint S, teacher improvement and student improvement are different quantities:

$$\Delta_T = \mathbb{E}[S(p, y_T) - S(p, x_a)], \qquad \Delta_S = \mathbb{E}[S(p, G_\phi(p; z)) - S(p, G_0(p; z))]. \tag{17}$$

Their reference distributions differ, so dividing them into a "transfer efficiency" is unjustified without additional assumptions.

A student comparison should pair base and adapted generation on the same unseen prompt and initial noise, holding resolution, sampler, guidance, and step count fixed. Each trial returns one image per system; predefined repeated trials estimate that distribution without best-of-N selection. Prompt-level resampling must retain all seeds together. Matched-capacity controls should train on anchor-only, preference-only, scalar-selected, and constrained targets; pairwise preference training is a separate objective baseline. Ablations must distinguish changes in data volume from the effects of semantic gates, preservation, base consistency, and the residual bound.

Assess quality and fidelity separately, including a blinded human comparison with randomized order and ties when appropriate study governance is in place. Report training-run variability, diversity, failures, and the costs of candidate generation, scoring, training, and deployment separately. A small feature bound, agreement across templates, or failure to detect a semantic difference cannot replace a prespecified noninferiority assessment. These evaluations remain necessary to test transfer from the teacher studied here.

6 Limitations and Conclusion

The current evidence concerns one backbone, small shared pools, and a previously exposed development cohort. Its quality interval supports a development signal, not a confirmatory population claim. Automatic scorers can share biases, and one attribute-parser limitation remains in the assessment. The teacher's limited facet coverage and failed semantic screen preclude treating these targets as qualified transfer supervision. Human preference validation, cross-backbone evidence, trained-adapter comparisons, and single-sample deployment measurements remain absent.

Select2Adapt makes the supervision boundary explicit: semantic eligibility precedes preference, fallback retains the alignment reference, and only conditional text features are adapted. The audited study demonstrates that reliable scoring can yield a positive quality signal while still failing semantic and coverage requirements. The central transfer question remains open: whether qualified selection preferences can improve independently assessed single-sample outputs without unacceptable loss of prompt fidelity.

AI USE STATEMENT

Generative AI tools assisted with manuscript editing, literature checks, figure preparation, methodological and experimental-design refinement, research code and experiment orchestration, data processing and numerical verification, interpretation of results and qualitative examples, and drafting and checking mathematical statements and proofs. Reported measurements derive from retained computational outputs; experimental images were not generatively edited. Human authors are responsible for the final content and submission.

ETHICS STATEMENT

Text-to-image generators and preference scorers can reproduce stereotypes, harmful content, and biases in whose preferences they reflect. Score-based eligibility does not remove these risks or establish a safety guarantee. Any future human evaluation requires appropriate consent and governance, and releases must respect model and data licenses. The qualitative examples are diagnostic illustrations, not human-study labels. No completed human-evaluation outcome is claimed.

REPRODUCIBILITY STATEMENT

The selection rule, calibration, adaptation boundary, and assessment protocol are specified in the main text and Appendix B; assumptions and complete proofs accompany the conditional bounds in Appendix A. The accompanying source package provides portable per-candidate scores, per-prompt choices and metrics, and a CPU endpoint-recalculation script for the new development study. It distinguishes historical aggregate tables from the audited per-prompt measurements. These records support selection-result checks; they do not constitute a trained-adapter experiment or establish complete reproducibility of the older aggregate-only study.

REFERENCES

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Jie Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Mingsheng Li, Mei Li, Kaixin Li, Zicheng Lin, Junyang Lin, Xuejing Liu, Jiawei Liu, Chenglong Liu, Yang Liu, Dayiheng Liu, Shixuan Liu, Dunjie Lu, Ruilin Luo, Chenxu Lv, Rui Men, Lingchen Meng, Xuancheng Ren, Xingzhang Ren, Sibo Song, Yuchong Sun, Jun Tang, Jianhong Tu, Jianqiang Wan, Peng Wang, Pengfei Wang, Qiuyue Wang, Yuxuan Wang, Tianbao Xie, Yiheng Xu, Haiyang Xu, Jin Xu, Zhibo Yang, Mingkun Yang, Jianxin Yang, An Yang, Bowen Yu, Fei Zhang, Hang Zhang, Xi Zhang, Bo Zheng, Humen Zhong, Jingren Zhou, Fan Zhou, Jing Zhou, Yuanzhi Zhu, and Ke Zhu. Qwen3-VL technical report. *arXiv preprint arXiv:2511.21631*, 2025. doi: 10.48550/arXiv.2511.21631.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/14f75513f0f1ca01de1e826b52e6b840-Abstract-Conference.html.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. ISBN 9780521833783. doi: 10.1017/CBO9780511804441.

Chaofeng Chen, Annan Wang, Haoning Wu, Liang Liao, Wenxiu Sun, Qiong Yan, and Weisi Lin. Enhancing diffusion models with text-encoder reinforcement learning. In *Computer Vision – ECCV 2024*, volume 15083 of *Lecture Notes in Computer Science*, pp. 182–198, Cham, 2025. Springer. doi: 10.1007/978-3-031-72698-9_11.

Junsong Chen, Jincheng Yu, Chongjian Ge, Lewei Yao, Enze Xie, Yue Wu, Zhongdao Wang, James Kwok, Ping Luo, Huchuan Lu, and Zhenguo Li. PixArt-α: Fast training of diffusion transformer for photorealistic text-to-image synthesis. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/fe989bb038b5dcc44181255dd6913e43-Abstract-Conference.html.

Kevin Clark, Paul Vicol, Kevin Swersky, and David J. Fleet. Directly fine-tuning diffusion models on differentiable rewards. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/13e8be77982beb73d7ed0bbf122f9f3c-Abstract-Conference.html.

Hanze Dong, Wei Xiong, Deepanshu Goyal, Yihan Zhang, Winnie Chow, Rui Pan, Shizhe Diao, Jipeng Zhang, Kashun Shum, and Tong Zhang. RAFT: Reward ranked finetuning for generative foundation model alignment. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=m7p5O7zblY.

Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H. Bermano, Gal Chechik, and Daniel Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=NAQvF08TcyG.

Yonatan Geifman and Ran El-Yaniv. Selective classification for deep neural networks. In *Advances in Neural Information Processing Systems*, volume 30, pp. 4878–4887. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/4a8423d5e91fda00bb7e46540e2b0cf1-Abstract.html.

Dhruba Ghosh, Hannaneh Hajishirzi, and Ludwig Schmidt. GenEval: An object-focused framework for evaluating text-to-image alignment. In *Advances in Neural Information Processing Systems*, volume 36, pp. 52132–52152, 2023. doi: 10.52202/075280-2270.

Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. CLIPScore: A reference-free evaluation metric for image captioning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 7514–7528. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.emnlp-main.595.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. doi: 10.48550/arXiv.2207.12598.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for NLP. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2790–2799. PMLR, 2019. URL https://proceedings.mlr.press/v97/houlsby19a.html.

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. In *The Tenth International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Kaiyi Huang, Kaiyue Sun, Enze Xie, Zhenguo Li, and Xihui Liu. T2I-CompBench: A comprehensive benchmark for open-world compositional text-to-image generation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 78723–78747, 2023. doi: 10.52202/075280-3443.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. MUSIQ: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5148–5157, October 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Ke_MUSIQ_Multi-Scale_Image_Quality_Transformer_ICCV_2021_paper.html.

Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-Pic: An open dataset of user preferences for text-to-image generation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 36652–36663, 2023. doi: 10.52202/075280-1594.

Baiqi Li, Zhiqiu Lin, Deepak Pathak, Jiayao Li, Yixin Fei, Kewen Wu, Tiffany Ling, Xide Xia, Pengchuan Zhang, Graham Neubig, and Deva Ramanan. GenAI-Bench: Evaluating and improving compositional text-to-visual generation. *arXiv preprint arXiv:2406.13743*, 2024a. doi: 10.48550/arXiv.2406.13743.

Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. BLIP: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 12888–12900. PMLR, 2022. URL https://proceedings.mlr.press/v162/li22n.html.

Yanyu Li, Xian Liu, Anil Kag, Ju Hu, Yerlan Idelbayev, Dhritiman Sagar, Yanzhi Wang, Sergey Tulyakov, and Jian Ren. TextCraftor: Your text encoder can be image quality controller. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7985–7995, June 2024b. URL https://openaccess.thecvf.com/content/CVPR2024/html/Li_TextCraftor_Your_Text_Encoder_Can_be_Image_Quality_Controller_CVPR_2024_paper.html.

Youwei Liang, Junfeng He, Gang Li, Peizhao Li, Arseniy Klimovskiy, Nicholas Carolan, Jiao Sun, Jordi Pont-Tuset, Sarah Young, Feng Yang, Junjie Ke, Krishnamurthy Dj Dvijotham, Katherine M. Collins, Yiwen Luo, Yang Li, Kai J Kohlhoff, Deepak Ramachandran, and Vidhya Navalpakkam. Rich human feedback for text-to-image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19401–19411, June 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Liang_Rich_Human_Feedback_for_Text-to-Image_Generation_CVPR_2024_paper.html.

Zhiqiu Lin, Deepak Pathak, Baiqi Li, Jiayao Li, Xide Xia, Graham Neubig, Pengchuan Zhang, and Deva Ramanan. Evaluating text-to-visual generation with image-to-text generation. In *Computer Vision – ECCV 2024*, volume 15067 of *Lecture Notes in Computer Science*, pp. 366–384. Springer, Cham, 2025. doi: 10.1007/978-3-031-72673-6_20.

Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, and Ying Shan. T2I-Adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion

models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(5):4296–4304, March 2024. doi: 10.1609/aaai.v38i5.28226.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4195–4205, October 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Peebles_Scalable_Diffusion_Models_with_Transformers_ICCV_2023_paper.html.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems*, volume 36, pp. 53728–53741, 2023. doi: 10.52202/075280-2338.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, June 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html.

Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A general framework for inference-time scaling and steering of diffusion models. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 55810–55827. PMLR, 2025. URL https://proceedings.mlr.press/v267/singhal25b.html.

Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. Diffusion model alignment using direct preference optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8228–8238, June 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Wallace_Diffusion_Model_Alignment_Using_Direct_Preference_Optimization_CVPR_2024_paper.html.

Xiaoshi Wu, Yiming Hao, Keqiang Sun, Yixiong Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. *arXiv preprint arXiv:2306.09341*, 2023. URL https://arxiv.org/abs/2306.09341.

Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. ImageReward: Learning and evaluating human preferences for text-to-image generation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 15903–15935, 2023. doi: 10.52202/075280-0700.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3836–3847, October 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Zhang_Adding_Conditional_Control_to_Text-to-Image_Diffusion_Models_ICCV_2023_paper.html.

A Proofs and Boundary Conditions

All claims below are conditional mathematical statements about the stated finite pool and sampling recursion. Reference scores need not represent human preference. Alignment scores, their anchor and the semantic predicate are fixed; semantic correctness is not assumed. Only a promotion must pass that predicate. The fallback anchor need not satisfy its absolute floors.

Proof of Proposition 1. Finiteness, nonempty contexts and finite scores make every minimum well defined. If $\hat{\mathcal{E}}$ is empty, $\mathcal{F} = \{a\}$ and $s = a$; otherwise every eligible candidate has $\hat{g}_i \geq \gamma > 0 = \hat{g}_a$, so the teacher's lexicographic rule selects a maximizer of $\hat{g}$ on $\mathcal{F}$. Its additional tie-breaks do not change this value. For every i, t, the triangle inequality gives

$$\left|(\hat{q}_{i,t} - \hat{q}_{a,t}) - (q^*_{i,t} - q^*_{a,t})\right| \leq |\hat{q}_{i,t} - q^*_{i,t}| + |\hat{q}_{a,t} - q^*_{a,t}| \leq 2\varepsilon_q. \tag{18}$$

If two finite vectors differ coordinatewise by at most r, their minima differ by at most r: every first-vector coordinate is at least the second-vector minimum minus r, and the reverse inequality follows by exchange. Consequently $|\hat{g}_i - g^*_i| \leq 2\varepsilon_q$. A promotion satisfies $g^*_s \geq \hat{g}_s - 2\varepsilon_q \geq \gamma - 2\varepsilon_q$; on fallback $g^*_a = 0$ exactly. If i^* maximizes g^*_i over $\mathcal{F}$, then

$$g^*_{i^*} - g^*_s \leq (g^*_{i^*} - \hat{g}_{i^*}) + (\hat{g}_s - g^*_s) \leq 4\varepsilon_q, \tag{19}$$

using $\hat{g}_{i^*} \leq \hat{g}_s$. A negative $\gamma - 2\varepsilon_q$ is a lower bound, not an improvement claim. A singleton pool therefore falls back, and $\varepsilon_q = 0$ gives exact reference-score optimality on $\mathcal{F}$. The $2\varepsilon_q$ promotion correction cannot generally be reduced: with one context and two semantically eligible images, take observed anchor/candidate scores $0, \gamma$ and reference scores $\varepsilon_q, \gamma - \varepsilon_q$. Each score error is ε_q, but the reference gain is exactly $\gamma - 2\varepsilon_q$. At $\gamma = 2\varepsilon_q$, only a nonnegative gain follows.

Proof of the semantic-score corollary. Passing the observed gate gives $A_i \geq \tau_A$ and $A_i \geq \eta_A A_a$. The candidate error bound implies $A^*_i \geq A_i - \varepsilon_A \geq \tau_A - \varepsilon_A$. For the relative part, the anchor error gives $A_a \geq A^*_a - \varepsilon_A$; since $\eta_A \geq 0$, multiplication preserves this inequality. Hence $A^*_i \geq A_i - \varepsilon_A \geq \eta_A A_a - \varepsilon_A \geq \eta_A A^*_a - (1 + \eta_A)\varepsilon_A$. Both lower bounds hold simultaneously, proving Equation 12. Applying the identical argument separately to each fragment proves its version. The derivation includes $\eta_A = 0$ and $\varepsilon_A = 0$, and never replaces the observed anchor by a reference-score maximizer. Negative multipliers are outside the stated assumptions.

Proof of Proposition 2. Equation 5 yields $\|\Delta_\phi(H)\|_F = \min(\|R_\phi\|_F, B) \leq B$ when both norms are nonzero. Its specified zero cases give the same inequality. This calculation uses real arithmetic; finite-precision use requires the implementation's effective-residual recheck to enforce the assumed bound. Subtracting the two guided predictions at identical z cancels their unconditional terms: $F^{h'}_k(z) - F^h_k(z) = w_k[V_k(z, h') - V_k(z, h)]$. Taking norms proves Equation 13 for positive, zero or negative guidance. The latent Lipschitz constant of $F^{h'}_k$ is at most $\ell_k = |1 - w_k|L^u_k + |w_k|L^z_k$. Let $e_k = \|z^{h'}_k - z^h_k\|_2$. Shared additive noise cancels on subtracting the updates. Add and subtract $F^{h'}_k(z^h_k)$ to obtain

$$\begin{aligned} e_{k+1} &\leq |\alpha_k| e_k + |\beta_k| \left(\|F^{h'}_k(z^{h'}_k) - F^{h'}_k(z^h_k)\|_2 + \|F^{h'}_k(z^h_k) - F^h_k(z^h_k)\|_2 \right) \\ &\leq (|\alpha_k| + |\beta_k|\ell_k) e_k + |\beta_k w_k| L^h_k B = C_k e_k + D_k. \end{aligned} \tag{20}$$

Induction over integers $K \geq 0$ gives $e_K \leq e_0 \prod_{j=0}^{K-1} C_j + \sum_{i=0}^{K-1} D_i \prod_{j=i+1}^{K-1} C_j$; $e_0 = 0$ proves Equation 14. This division-free form remains valid when any $C_j = 0$, with an empty sum zero and empty product one. Zero conditioning norm, zero residual ratio, or zero guidance at every step gives identical coupled trajectories. If a deterministic decoder is L_D-Lipschitz on the compared latents, the decoded-image distance is at most $L_D e_K$; converting this into a semantic or quality-score guarantee requires further regularity. Neither those constants nor such a semantic guarantee are established here.

B Implementation and Development Record

Numerical identity. The PickScore checkpoint is pinned to revision a4e4367c6dfa7288a00c550414478f865b875800, and its CLIP processor to 1c2b8495b28150b8a4922ee1c8edee224c

Table 2: Realized selection study and specified adaptation configuration. Training settings describe the implementation; no efficacy run using the present teacher is reported.

Component	Configuration
Cohort	48 prompts, 16 per color/shape/texture; two fragments per prompt; eight candidates per prompt; full cohort retained.
Generation	PixArt-XL-2, 512×512, 20 steps, guidance 4.0; empty negative prompt; bf16 pipeline and fp32 VAE decoding.
Semantic selection	Frozen whole/fragment Qwen3-VL scores; whole and fragment floors 0.5; whole relative floor 0.25; fragment relative floor 0.9.
Preference selection	PickScore native scaled cosine logits; original prompt and "An image of {prompt}."; fp32 eager attention; all model parameters frozen.
Calibration	Twelve fixed clean/blur controls, Gaussian radius 3; paired-gain margin 0.288256509399414; no assessment-driven threshold adjustment.
Adapter	Two linear layers with SiLU, hidden 64, dropout 0, zero-initialized final layer, residual scale 1, full-tensor RMS ratio at most 0.10.
Objective	Promoted/source-x_0 preservation roles 1:1; noise-prediction MSE; preservation-only base consistency (weight 1); valid-token residual MSE (weight 0.01).
Assessment	Reused independent MUSIQ and BLIP-based attribute outputs for all 384 images; all 48 prompt differences enter each endpoint.
Uncertainty	20,000 facet-stratified bootstrap replicates; PCG64 seed 1234; linear 2.5/97.5 percentiles; no multiplicity adjustment.

284c0c. Each call normalizes image and text embeddings and multiplies their dot product by the learned exponential logit scale. A separate native forward supplies the numerical comparison. The maximum native discrepancy is zero; independent CPU reconstruction from stored vectors differs by at most 4.60×10^{-6}. Single-image softmax is identically one and is not used as quality confidence. Runtime records include inputs, normalized vectors, preprocessing, and actual repeated forward outputs.

Screen and retention. Numerical tolerances are 10^{-4} for logits and 10^{-5} for probabilities. Both context/order comparisons must prefer clean for a control to count; 10/12 pass. The fixed screen does not require all 12 controls to be correct. The two shape-control errors remain in calibration and diagnostics. Model assets, source, schedule, calibration, selection, and assessment bindings are recorded. The complete 836-call audit agrees with the production results, and 169 focused software tests plus 38 independent-auditor selftests pass. These checks concern the recorded computation, not human preference validity or the unperformed adaptation experiment.

Assessment implementation. MUSIQ uses the KonIQ-10k-trained checkpoint through `pyiqa 0.1.13`; its weight SHA256 begins `e95806b9`. BLIP uses `model_base_vqa_capfilt_large.pth` and the T2I-CompBench implementation at commit `1b7094991a57f3c22abdd4f6e8ba6c1a15517073`. For each parsed noun phrase, the evaluator takes first-token decoder probabilities at BOS, forms $P(\text{yes})/[P(\text{yes}) + P(\text{no})]$, and rounds to four decimal places. It multiplies the nonempty phrase scores in float32 and rounds the product to four decimals. Empty phrase slots contribute one. This operational score is neither a full-sequence likelihood nor a human semantic label. The native products for all 384 candidates were independently reconstructed from retained slot outputs.

Semantic assessment limitation. For one retained color prompt, "a red accordion and a green watering can," the frozen assessment parser yields "a green watering" as the second object phrase. We retain its recorded value in every full-cohort comparison. This limits interpretation of the semantic proxy; it does not justify deleting the prompt after observing outcomes.

Preceding quality teacher. The comparator retains the original all-eligible decisions using its original quality scale and checks, including a quality floor of 4.00, gain margin 0.10, and maximum between-template discrepancy 0.50 for each image on its 1–5 scale. Its two promotions both lie in texture. It is a historical selector on these same candidates, not an adapter baseline. No threshold from that 1–5 quality scale is transferred to PickScore logits.

Implementation checks and future training record. A zero adapter residual should reproduce base predictions with matched inputs. The optimizer must update only adapter parameters, with gradients propagated through the frozen denoiser. The unconditional branch must bypass adaptation and the effective residual bound must hold after arithmetic in the chosen dtype. The existing training cache accepts four-candidate records and stores source candidate zero for preservation even when selection falls back to a different alignment anchor. Integration of the present eight-candidate teacher with that cache has not been validated; this manuscript does not claim an end-to-end training run for it. Future efficacy runs additionally need frozen prompt splits, target and preservation manifests, optimizer settings, training steps and seeds, checkpoint-selection rules, and matched single-sample outputs. None of these missing efficacy results is inferred from successful software checks.

C Additional Paired Contrasts

All intervals below use the same 48 prompts and bootstrap draws as the main comparison. They complete the 16 declared endpoints: ten non-anchor-versus-anchor metric contrasts and the six contrasts in Table 3.

Table 3: Constrained teacher minus alternative selectors. Positive effects favor the constrained teacher on the given metric.

Reference selector	Δ MUSIQ [95% interval]	Δ BLIP [95% interval]
Preference-only	$+0.1273$ $[-0.4752,\ 0.7256]$	$+0.2350$ $[0.1510,\ 0.3235]$
Scalar mixture	$+0.1885$ $[-0.2246,\ 0.6017]$	$+0.0804$ $[0.0269,\ 0.1416]$
Preceding teacher	$+0.3105$ $[0.0762,\ 0.5732]$	-0.0035 $[-0.0270,\ 0.0124]$

Table 4: Facet outcomes for the constrained teacher. Means include all 16 prompts in each facet, including non-promotions.

Facet	Changes / prompts	Mean Δ MUSIQ	Mean Δ BLIP
Color	1/16	-0.101294	-0.030131
Shape	0/16	0	0
Texture	7/16	$+1.293341$	$+0.020419$

The sole color promotion is retained despite decreasing both endpoints. No shape promotion means there is no observed teacher intervention on that facet; zero change is not evidence that an intervention would preserve semantics. Facet analyses explain the coverage failure but do not replace the complete-cohort endpoint.

Table 5: Previously reported GenEval selection scores (%). All best-of-four selectors share a candidate pool; Macro averages the six tasks equally. The table evaluates selection, not the adapted generator.

Method	Macro	Single	Two	Count	Color	Pos.	Attr.
Base, $N = 1$	46.58	96.88	47.47	44.06	76.06	8.25	6.75
Expected random-of-4	47.23	97.97	46.78	45.78	76.99	8.25	7.63
CLIPScore-of-4	51.23	98.44	57.32	53.13	78.99	10.25	9.25
Single-Qwen-of-4	51.20	97.50	57.07	55.00	77.13	11.00	9.50
VQAScore-form Qwen	55.13	98.44	62.63	59.69	80.05	15.00	15.00
Raw Gate4	51.35	97.50	57.07	55.63	77.13	11.00	9.75

D Previously Reported GenEval Selection Context

Setting and scope. The preceding selection study uses PixArt-XL-2 at 512×512 (Chen et al., 2024) and 553 GenEval prompts, with four output positions per prompt and four candidates per output: 2,212 selection units and 8,848 images. We analyze previously reported aggregate scores; raw outputs and the exact checkpoint revision were unavailable for recomputation. These experiments evaluate candidate selection rather than trained Select2Adapt outputs.

Gate4 uses a seed-0 fallback, facet-level Qwen3-VL evidence (Bai et al., 2025), preservation booleans, and a CLIP tie-break. It has no naturalness gate and is not the alignment-anchored teacher of Section 3. The VQAScore-form baseline uses teacher-forced $P(\text{Yes})$ with Qwen3-VL-4B, following the scoring form of Lin et al. (2025). It is not a reproduction of the original CLIP-FlanT5 implementation. GenAI-Bench is a related ranking reference (Li et al., 2024a), not the sole source of that formula.

Sampling headroom exceeds realized selection gain. Table 5 reports the six-task macro scores. Oracle-of-four reaches 62.61%, 16.03 points above the 46.58% base, whereas Gate4 realizes 4.77 points. The source reports a prompt-cluster 95% interval of $[3.74, 5.85]$ for the latter difference. The shared-pool comparison isolates selection from the availability created by extra sampling. It does not show that a conditioning adapter can recover the same gain.

Stronger baselines narrow the claim. Gate4 differs from CLIPScore by 0.12 points with reported interval $[-1.31, 1.52]$, and from single-Qwen by 0.15 points with interval $[-0.33, 0.64]$. These intervals do not resolve a difference; they do not establish equivalence. VQAScore-form Qwen has a higher macro score. A percentile-normalized weighted sum reaches 53.30%, and removing the fallback reaches 52.54%. Consequently, these data do not support either selector superiority or the claim that hard constraints inherently outperform scalar selection. That comparison is central to the new teacher evaluation.

Effect of the legacy checks. Removing the old preservation checks leaves the macro score at 51.35% and changes only two decisions. The source reports that those checks block only 1 of 130 target-positive calibration comparisons. This is evidence of redundancy in that setting, not evidence that preservation constraints improve outcomes. An attempted pairwise route also failed its prespecified order-consistency criterion; it cannot be retrospectively described as successful cross-template consistency for Select2Adapt.

Among 2,212 units, raw Gate4 makes 112 benchmark-improving and eight benchmark-degrading selections. Their net change is $100 \times 104/2212 \approx 4.70$ image-weighted percentage points, distinct from the six-task macro gain. The eight degrading events are automatic benchmark transitions, not human harms. The source also reports 871 candidate-pool failures and 244 missed opportunities. These diagnostics identify selection limitations without making a human-quality or transfer claim.

Evidence for selection and transfer. The reported human study was incomplete, with no validated calibration lock, human utility estimate, or risk certificate. Its proposed risk-control procedure concerns a frozen selector and does not extend automatically to an adapted generator. Adapter effi-

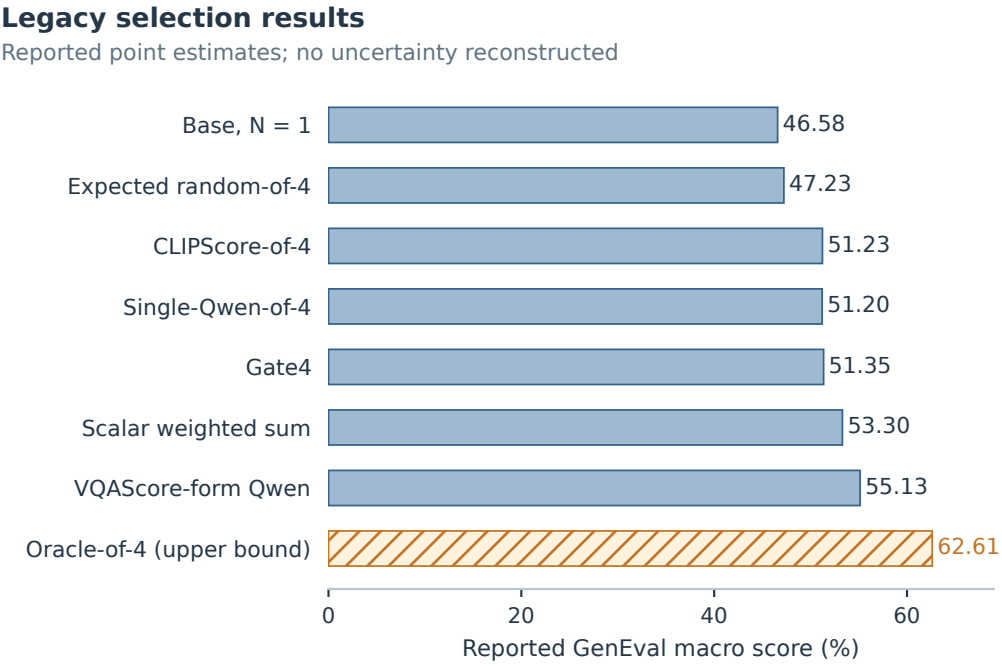

Figure 4: Selection performance on shared candidate pools. Weighted-sum and VQAScore-form selection exceed raw Gate4. Oracle uses benchmark labels and provides an analysis-only bound. Bars reproduce reported point estimates; selected pairwise confidence intervals are reported in the text.

cacy, training cost, single-sample latency, and student preference outcomes remain to be established under the protocol in Section 5.

Aggregation and uncertainty. The preceding record contains 320 single-object, 396 two-object, 320 counting, 376 color, 400 position, and 400 attribute-binding output units, totaling 2,212. Macro scores weight the six tasks equally. Expected random-of-four averages all candidate correctness labels; oracle-of-four uses benchmark labels and is analysis-only. The reported intervals resample prompts within tasks while retaining all output positions. They are reproduced as reported, without the original draws; no new significance test is inferred from rounded values.

E Qualitative Semantic Contrasts and Provenance

The exploratory image collection is independent of the GenEval results. It contains 80 presentations, 91 unique image contents, 53 unordered pairs, and 38 prompts; repeated images and reversed A/B presentations were identified by SHA256. The four selected pairs are distinct. The collection contains no returned human answers or method mapping, and its presentation count is not an independent-trial count. A/B and the lookup identifiers record display positions, not experimental treatments.

Table 6: Prompt and source lookup for Figure 5. Full IDs, filenames, hashes, and observations accompany the figure source.

Panel	Full source prompt	Presentation ID prefix
A	a black chair and a brown table	`a8e37d6a`
B	a woman on the left of a sheep	`25d73080`
C	The warm, inviting glow of the fireplace flickered in the cozy, dimly lit living room, as the family gathered around for a night in.	`1bbe67fb`
D	The pyramidal terrarium with its spherical cactus plants and triangular succulents made for a beautiful centerpiece on the square coffee table.	`796b6f01`

Image handling and selection limits. The source PNGs are copied byte-for-byte and embedded without cropping, retouching, recoloring, sharpening, or generative editing. Each image is shown at the same square size. All callouts sit outside the images. Cases were chosen for interpretable differences, not to measure prevalence; visual notes are editorial observations, not human-study labels. In particular, the terrarium pair is not a clean semantic pass: B contains a columnar cactus, and the complete tabletop outline is not visible. Prompt excerpts improve readability without changing the original wording or generating conditions. Publication of a Base/Select2Adapt comparison would require a verified treatment mapping and generated-image provenance beyond these blinded files.

F Architecture and Evidence Boundaries

The framework in Figure 1 adapts only conditional text features. The text encoder, autoencoder, unconditional conditioning, and denoiser parameters remain fixed. Candidate selection is an offline supervision operation; the intended deployment performs one trajectory without an inference-time quality scorer. The diagram describes computation and does not depict generated before/after outcomes.

Three evidence levels remain distinct. The historical GenEval table is an aggregate selection record whose raw experiment is not reconstructed here. The PickScore study has retained per-call and per-prompt records with independent recalculation, but uses exposed development data and fails joint qualification. The adaptation component has an implementation-aligned definition and conditional mathematical analysis; neither establishes trained-student efficacy or single-sample transfer on an independent cohort.

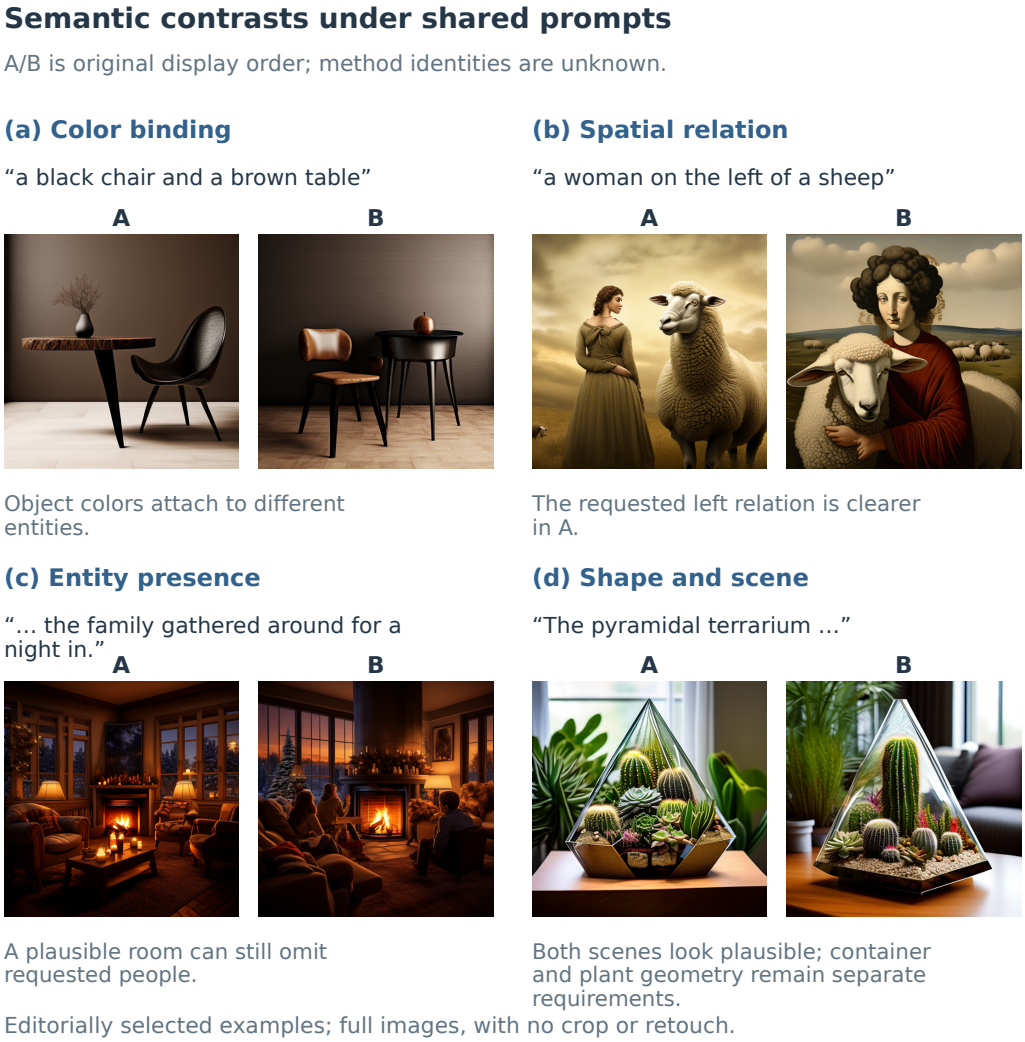

Figure 5: Four diagnostic A/B pairs: color binding, spatial relations, entity presence, and geometric details. Images retain original content and order. A/B labels are display positions, not Base/Select2Adapt identities. Prompts are excerpted where indicated; full prompts appear in the provenance table. These editorially chosen examples do not estimate prevalence or demonstrate model improvement.