CAPA: A CONTEXT-AWARE AGENT FOR CONTINUAL PERSONALIZED PHOTO RETOUCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

A continual photo-retouching agent must use personal feedback across images while preserving scene-specific preferences and temporary intent. We introduce Context-Aware Preference Agent (CAPA), which connects feedback use, acquisition, and retention through hierarchical preference states and an axis-level evidence ledger. Candidate contrasts represent each choice relative to the alternatives shown. Responses adapt the current session immediately, while the ledger tracks image, session, and scope information for persistent promotion. Remaining-axis masks let eligible coordinates enter long-term memory while preserving the others for later promotion; each coordinate is consumed once. CAPA with Decision Value (CAPA-DV) uses personal response probabilities and semantic candidate partitions to select questions and follow-ups within an interaction budget. In chronological replay of data from 31 participants, hierarchical memory achieves 44.23% candidate-selection Top-1 accuracy versus 40.26% for flat memory using the same history. In fixed-prefix replay with choice-consistent answers, one-step CAPA-DV reaches 80.14% versus 67.75% for fixed coverage, using 326 versus 452 question exposures. Synthetic tests further characterize answer-noise sensitivity and retention under temporary perturbations. CAPA links current editing intent to comparative evidence and its persistent reuse in a continual personalization workflow.

1 INTRODUCTION

Personalized photo retouching aligns color and tone with individual taste, using signals such as comparative judgments and reference edits (Kim et al., 2026; Weldengus et al., 2025). Agentic systems connect visual analysis and user instructions to editing operations, including Lightroom parameter control (Chen et al., 2025). As these systems are used across images and sessions, they must also determine which feedback should guide future edits and which expresses the current intent.

Continual personalization addresses this challenge through explicit memory that adapts from interaction and distinguishes persistent preferences from contextual changes (Liang et al., 2026; Hossain et al., 2026). In retouching, a user who generally prefers natural colors may choose vivid tones for travel photographs or request a warmer appearance for one session. The scope of feedback—global, scene-specific, or session-specific—therefore matters alongside its content. For candidate choices, that content also depends on the alternatives shown: choosing one edit over others supplies comparative evidence on the axes that vary within the set. Reusing such feedback therefore requires preserving both its comparative meaning and its contextual scope.

These properties connect three decisions: how to use feedback, what to ask next, and what to retain. Session adaptation and persistent reuse require different levels of support. Clarification should target candidate distinctions that affect selection, accounting for interaction cost and follow-up questions (Howard, 1966; Boutilier, 2002). All three decisions depend on what each comparison reveals and where that evidence applies.

To connect these decisions, we introduce the Context-Aware Preference Agent (CAPA). Each choice is represented by its contrast with the displayed alternatives, the axes that varied, and its image, session, and scope. The same comparison supports immediate session adaptation and separately tracked eligibility for persistent memory. Global, scene-specific, and session-specific summaries condition the current decision, with explicit instructions taking priority. A scope-tagged ledger tracks

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.

confidence, temporary status, cross-image and cross-session support, and response-type conflicts. Remaining-axis masks let eligible coordinates enter persistent memory while retaining the others for later promotion. This ties reuse to the originating comparisons and consumes each promoted coordinate once. In the online editing prototype, retrieved preferences also condition structured plans and constrained Lightroom adjustments.

CAPA-DV uses this preference information to decide what to ask next among fixed rendered candidates. A personal response model distinguishes specific candidate choices from set-level responses, and semantic editing axes define the available answer partitions. The budgeted policy evaluates questions together with their possible follow-ups and stops when further discrimination does not justify its cost. Figure 1 shows how the editing path, preference updates, retention, and querying fit into the overall workflow.

Our contributions are:

- A comparison-driven memory mechanism that links rendered-candidate contrasts and their contextual provenance to immediate session adaptation and persistent promotion. Its remaining-axis ledger tracks eligibility across images and sessions, retains coordinates awaiting support, and consumes each promoted coordinate once.
- CAPA-DV, a personalized clarification policy that combines candidate-choice and set-level response forecasts with semantic partitions of rendered candidates. It evaluates decision gains, question costs, and the remaining budget to select questions, plan follow-ups, and stop.

2 Related Work

Personalized and agentic retouching. Recent work personalizes visual editing through reference conditioning, preference learning, and agentic control. PPS learns photographic color and tone preferences from pairwise judgments; its PPSD dataset contains responses from 767 users (Kim et al., 2026). RefRetouch encodes paired retouching examples into style latents and retrieves content-relevant references to adapt without test-time fine-tuning (Weldengus et al., 2025). C-DPO uses graph-based user embeddings and collaborative preference optimization to align diffusion edits with individual and neighboring users' tastes (Dunlop et al., 2025). PALATE keeps the editor fixed and personalizes candidate selection through a shared global reward, category residuals, and user adapters (Wang et al., 2026b). For agentic execution, PhotoArtAgent connects artistic analysis and iterative evaluation to Lightroom parameter control (Chen et al., 2025). PerTouch combines semantic-region retouching with a vision-language model (VLM) agent, feedback-driven refinement, and scene-aware memory (Chang et al., 2025). These approaches motivate connecting personal preferences to controllable edits. CAPA focuses on how comparative feedback enters a personalized retouching workflow: which parameter axes were observable, which context a response applies to, and when accumulated evidence becomes eligible for persistent reuse.

Continual personalization and selective memory. PAHF couples pre-action clarification, memory-conditioned actions, and post-action feedback to update explicit user memory (Liang et al., 2026). PGMem connects dialogue events, transient persona states, and persistent traits through provenance and support, contradiction, and temporal-shift relations (Choi et al., 2026). QUMem uses typed memories and staged retrieval to infer user states applicable to the current query and context (Wang et al., 2026a). Personalize-then-Store uses session-level gating to skip memory operations for transient interactions (In et al., 2026). CAPTURE combines temporal belief tracking, a preference ledger, and clarification to distinguish preference changes from contextual shifts and memory poisoning (Hossain et al., 2026). CAPA links memory eligibility to the comparisons that produced the evidence: candidate variation determines the recorded axes, and the ledger promotes each axis after cross-image and cross-session support. Session adaptation and persistent promotion therefore operate on the same records with distinct eligibility rules.

Interactive visual preference acquisition. PASTA learns latent preference and choice models from sequential feedback and uses value-based reinforcement learning to select prompt-expansion slates for interactive image generation (Nabati et al., 2024). Hahn et al. (2024) represent uncertain visual intent through editable belief graphs and investigate proactive clarification, including

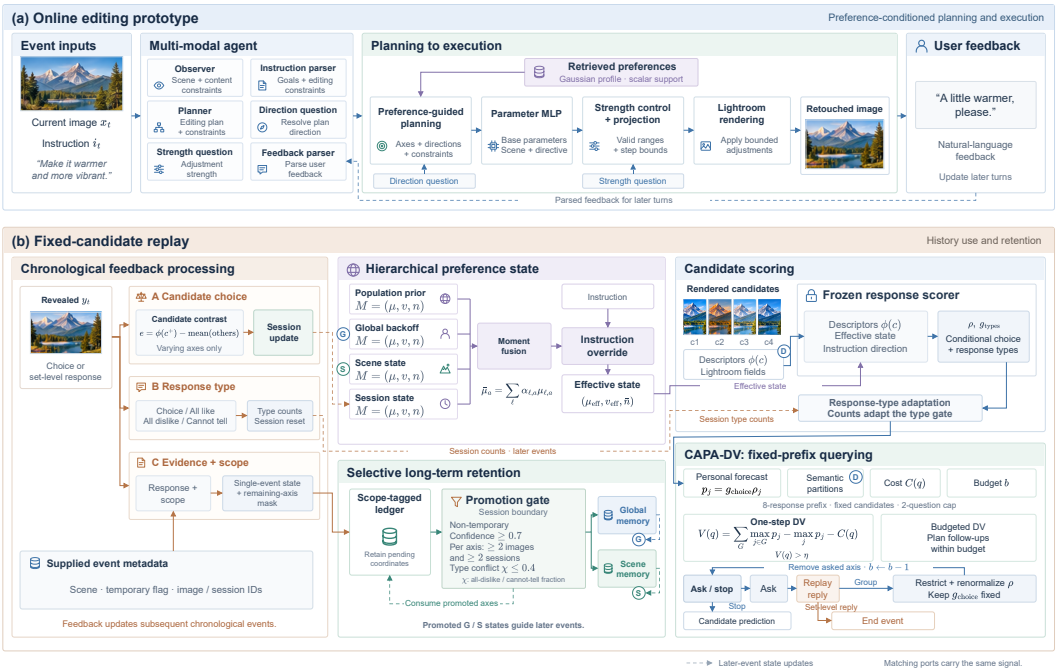


Figure 1: CAPA overview. (a) Image context, instructions, and retrieved preferences guide constrained Lightroom editing. (b) Fixed-candidate replay integrates chronological feedback, hierarchical preferences, selective memory promotion, and CAPA-DV querying from an eight-response prefix. Feedback informs later events, while current instructions override retrieved preferences.

importance-weighted entropy scoring. These systems elicit preferences while refining the content to be generated. CAPA-DV addresses the complementary stage after retouching candidates have been rendered, using binary questions about semantic editing axes to distinguish among a fixed set of rendered alternatives.

Decision-directed elicitation. Value-of-information analysis evaluates observations through their effect on decisions (Howard, 1966); sequential preference elicitation extends this perspective to costly questions, uncertain answers, and future interactions (Boutilier, 2002). SAGE maintains structured beliefs over tool calls and scores questions by expected gains in maximum candidate probability, with interaction costs and threshold-based stopping (Suri et al., 2025). Building on decision-value elicitation, CAPA-DV combines personalized seven-outcome response forecasts with semantic partitions of rendered candidates and budgeted follow-up selection. The candidate set and its editing axes define the available questions, while the response model accounts for both candidate-specific and set-level feedback.

3 METHOD

Figure 1 summarizes CAPA's workflow. At event t, image x_t and instruction i_t are processed using user u's previously revealed history $H_{u,<t}$. The online prototype (panel a) converts preference-conditioned plans into constrained Lightroom adjustments. Fixed-candidate replay (panel b) combines hierarchical preference states and instruction priority for response scoring, which guides CAPA-DV's cost-aware clarification within the remaining budget. In chronological replay, revealed feedback updates session-level information and the scope-tagged ledger; at session boundaries, eligible evidence is promoted axis by axis to scene and global memory before the next session's first prediction.

3.1 Agent Workflow and Evaluation Configurations

The online prototype combines image observations, instructions, and retrieved preferences into editing plans. Plan-direction and strength questions refine directives; a parameter multilayer perceptron (MLP) produces Lightroom adjustments projected onto valid ranges and per-step bounds. Six Qwen3-VL interfaces (Bai et al., 2025) share one low-rank adaptation (LoRA) adapter (Hu et al., 2021); the parameter predictor and response scorer are separate networks (Appendix A).

Figure 1(b) uses axis-masked states and a frozen response scorer for replay. Chronological replay measures history use and retention; fixed-prefix replay evaluates CAPA-DV from separately initialized target histories. The editing prototype in panel (a) uses a Gaussian memory profile with one scalar support value per source and the variance-utility router in Appendix A.5. The synthetic tests generate their own probabilities and evidence to examine the decision and retention rules under controlled perturbations.

3.2 Scope-aware Preference Modeling

The replay configuration combines a population prior with global, scene-specific, and current-session user states. Each source stores three nine-dimensional vectors: $\mathcal{M} = (\mu, v, n) \in \mathbb{R}^9 \times \mathbb{R}^9_{>0} \times \mathbb{R}^9_{\geq 0}$, representing signed comparative evidence, uncertainty, and fusion support scores, respectively. The axes are warmth, tint, brightness, contrast, saturation, detail, grain, vignette, and matte. Descriptors $\phi(c) \in [-1, 1]^9$ summarize normalized Lightroom parameters (Appendix A.1). They encode candidate adjustments; the state summarizes comparisons among those adjustments.

Replay conditions on supplied scene labels and event-level temporary-scope flags. A new session resets the session state and response-type counts; image changes within a session do not. Temporary scope and explicit instructions are separate inputs: scope controls retention, while an instruction controls the current decision. The global state pools promoted scene evidence for use as a backoff across scenes. The ledger retains observations awaiting promotion, and retrieval uses the promoted states and current-session state.

Retrieval combines the population, global, matching-scene, and session states by matching their first and second moments. Let $\mathcal{L}$ index these sources, with axis-dependent weights $\alpha_{\ell,a} \geq 0$ and $\sum_{\ell \in \mathcal{L}} \alpha_{\ell,a} = 1$. Fusion computes

$$\begin{aligned} \bar{\mu}_a &= \sum_{\ell \in \mathcal{L}} \alpha_{\ell,a} \mu_{\ell,a}, \\ \bar{v}_a &= \max\left(10^{-4}, \sum_{\ell \in \mathcal{L}} \alpha_{\ell,a}(v_{\ell,a} + \mu^2_{\ell,a}) - \bar{\mu}^2_a\right), \\ \bar{n}_a &= \sum_{\ell \in \mathcal{L}} \alpha_{\ell,a} n_{\ell,a}. \end{aligned} \tag{1}$$

The second moment captures within-source uncertainty and between-source disagreement. The support score determines source availability and fusion weights: event updates add confidence, and summary fusion averages the scores. Cross-image and cross-session counts come from ledger identifiers. The instruction vector $z \in \{-1, 0, +1\}^9$ specifies decrease, no direction, or increase on each axis. A nonzero z_a overrides the current mean and sets its variance to 10^{-3} without overwriting memory. The parameter controller resolves adjustment strength separately. Appendix A.6 specifies the fusion order and overrides.

Observable-axis update. For candidates $\mathcal{C}_t = \{c_{t,1}, \ldots, c_{t,K}\}$ with $K = 4$, the response y_t either selects a candidate or is a set-level response in $\mathcal{R} = \{\text{all-like}, \text{all-dislike}, \text{cannot-tell}\}$. If c_t^+ is selected, the contrast evidence and observable-axis mask are

$$e_{t,a} = \phi_a(c_t^+) - \frac{1}{K-1} \sum_{c \in \mathcal{C}_t \setminus \{c_t^+\}} \phi_a(c), \tag{2}$$

$$o_{t,a} = \mathbb{I}[y_t \in \mathcal{C}_t]\, \mathbb{I}\left[\max_{1 \leq j \leq K} \phi_a(c_{t,j}) - \min_{1 \leq j \leq K} \phi_a(c_{t,j}) > 10^{-8}\right]. \tag{3}$$

Here $\mathbb{I}[\cdot]$ is the indicator function. For confidence $\kappa_t \in [0,1]$, set $r_t = 0.25 + 0.75(1-\kappa_t)$, $w_t = \max(\kappa_t, 10^{-3})$, and $\tau_a = v_a^{-1}$. For session-state axes with $o_{t,a} = 1$, update

$$\begin{aligned} \tau'_a &= \tau_a + w_t/r_t, & \mu'_a &= (\tau_a\mu_a + w_t e_{t,a}/r_t)/\tau'_a, \\ v'_a &= \mathrm{clip}((\tau'_a)^{-1}, 10^{-4}, 4), & n'_a &= \min(n_a + w_t, 10^4). \end{aligned} \tag{4}$$

For $o_{t,a} = 0$, $(\mu'_a, v'_a, n'_a) = (\mu_a, v_a, n_a)$. This update summarizes the chosen candidate's coordinate-wise contrast with the displayed alternatives. The mask tests numerical variation; co-varying axes receive evidence together, preserving their joint contrast without attributing the choice uniquely to one axis. Recorded human responses use $\kappa_t = 1$. Set-level responses update the session response-type counts and ledger, with no directional state update.

Long-term promotion. The ledger stores scope metadata, remaining-axis masks, and each choice's single-event state, obtained by updating an empty source. At session close, eligible records must be non-temporary with confidence at least 0.7. An attempt proceeds when at most 40% of eligible records are all-dislike or cannot-tell. Each promoted axis additionally requires choice evidence from two distinct images and two sessions. Confidence-weighted single-event summaries update the matching-scene memory and pooled global backoff; consumed axis flags prevent reuse. Each scene has at most three slots. One successful scene-level attempt counts as one promotion, and the final flush affects write counts only. The conflict screen measures response types, while image/session counts measure repeated support. Global pooling uses the scene-storage eligibility rule. Appendix A.2 specifies slot assignment, fusion, and ledger consumption.

3.3 Response Scoring

The frozen structured-feature scorer conditions on candidate descriptors, effective state, and instruction direction. History enters through the state and response-type counts. The candidate head estimates the choice-conditional probabilities $\rho_j = P(y = c_j \mid y \in \mathcal{C}_t)$. The type head estimates $g_{\text{choice}} = P(y \in \mathcal{C}_t)$ and $g_r = P(y = r)$ for $r \in \mathcal{R}$, giving

$$p_j = g_{\text{choice}}\rho_j, \qquad p_r = g_r \quad (r \in \mathcal{R}). \tag{5}$$

Three networks' seven-class probabilities are averaged before gate calibration on synthetic development records and session-count adaptation. The latter smooths the four response-type counts with four prior observations distributed according to training frequencies. Explicit directives use the raw ensemble distribution. Counts update after response revelation; changing counts alone preserves ρ, while changing the preference state can affect both heads. Neural weights remain fixed. Appendix A.7 gives the objectives and adaptation formulas.

3.4 Decision-value Querying

CAPA-DV scores clarification by expected improvement in the best candidate decision, following the maximum-probability criterion used by SAGE (Suri et al., 2025). Each varying editing axis defines a binary partition $\mathcal{G}(q)$ at its initial minimum–maximum midpoint. Asked axes are removed. For axis spread $d_a = \max_j \phi_a(c_j) - \min_j \phi_a(c_j)$, the cost proxy is $C(q_a) = \mathrm{round}_4(0.05 + 0.1/(1+d_a))$, where round_4 rounds to four decimal places. Costs and the stopping margin are expressed on the unit-reward scale. Let $g = g_{\text{choice}}$ and $p_j = g\rho_j$. With unit reward for selecting the preferred candidate and zero reward for set-level outcomes, the one-step value under deterministic group answers is

$$V(q) = \sum_{G \in \mathcal{G}(q)} \max_{j \in G} p_j - \max_j p_j - C(q). \tag{6}$$

The policy asks the highest-value question that splits the remaining candidate set when $V(q) > \eta = 0.05$ and budget remains. Group replies restrict and renormalize ρ while retaining the initial response-type gate. This update separates candidate discrimination from the forecast of whether the event yields a specific choice.

Remaining-budget planning. Sequential elicitation evaluates future questions when choosing the current one (Boutilier, 2002). For a remaining candidate set S with positive total mass, unasked

questions $\mathcal{Q}$, and remaining budget b, define

$$\rho_j^S = \frac{\rho_j \mathbb{I}[j \in S]}{\sum_{k \in S} \rho_k}, \qquad \Phi(S) = g \max_j \rho_j^S,$$
$$P(G \mid S, \text{choice}) = \sum_{j \in S \cap G} \rho_j^S, \qquad F(S, \mathcal{Q}, 0) = \Phi(S). \tag{7}$$

The budgeted policy optimizes a fixed-g surrogate that compares stopping with choice-conditional continuation:

$$F(S, \mathcal{Q}, b) = \max \Bigg\{ \Phi(S), \max_{q \in \mathcal{Q}(S)} \Bigg[-C(q) - \eta + \sum_{\substack{G \in \mathcal{G}(q) \\ P(G|S,\text{choice}) > 0}} P(G \mid S, \text{choice}) F(S \cap G, \mathcal{Q} \setminus \{q\}, b-1) \Bigg] \Bigg\}, \tag{8}$$

where $\mathcal{Q}(S) = \{q \in \mathcal{Q} : q \text{ splits } S \text{ into two nonempty groups}\}$. Terminal selection reward is weighted by the initial choice mass g, while question costs accumulate along choice-conditional paths. Appendix A.8 details this surrogate's treatment of future costs. With no feasible question, no choice mass, or no remaining budget, the policy stops. Ties favor stopping; question ties use a fixed axis order. For $b = 1$, continuation improves on stopping by $V(q) - \eta$, recovering the one-step rule. Both policies use the same probabilities, costs, question pool, and threshold, with different planning depths.

In the original replay, answers identify the group containing the recorded choice and are revealed only after question selection. Set-level responses terminate after the first exposed question, whose cost still counts, and leave the prediction unchanged. Hard conditioning can exclude the target after an incorrect answer; Appendix D specifies the synthetic answer-noise tests. Appendix A.8 gives the fallback rules; Appendix A.5 describes the online variance-utility router.

4 Experiments

We evaluate historical conditioning, fixed-candidate clarification, and evidence retention using a 31-participant cohort and two synthetic tests. The cohort provides chronological replay and an independent-target, fixed-prefix query replay. The simulations evaluate query decisions and stored scene summaries under controlled perturbations. Qualitative examples in Appendix E (Figure 2) illustrate personalized retouching across four user profiles and the effects of explicit temporary instructions.

4.1 Data and Experimental Setup

Training. Source photographs come from RAISE (Dang-Nguyen et al., 2015), PARA (Yang et al., 2022), and the HDR+ Burst Photography Dataset (Hasinoff et al., 2016). Scene analysis, preset retrieval, style/strength variation, and quality filtering yield 28,404 Lightroom-rendered candidates. A simulator instantiates 10,000 users, each with three sessions of ten interactions and four candidates per interaction, producing 300,000 supervision records. Latent preferences generate targets and are excluded from model inputs. Training/development counts are 60,000/1,000 for the response scorer, 15,400/534 for the six structured interfaces, and 4,858/1,142 for parameter prediction. These module-specific samples can overlap; known source groups remain together within each module's split. Human-evaluation photographs are separated from synthetic training and development photographs by source image. The interfaces use Qwen3-VL-8B-Instruct (Bai et al., 2025) with rank-16 LoRA for two epochs. The response scorer averages three independently seeded networks; the parameter module averages projected deltas from three independently seeded MLPs, each with two 256-unit hidden layers. These three response networks form one ensemble for replay. Calibration uses synthetic development records. During human replay, neural parameters remain fixed; feedback updates preference states and response statistics. Appendix B describes synthetic supervision and source grouping; Appendix A gives the component architectures and training objectives.

Table 1: Participant-macro replay metrics. States P/H/F/M denote population/hierarchical/flat/mismatched preferences; types P/U denote population/personal response statistics. Conditional NLL and Top-1 use choices. Bold marks column optima, including ties.

State / type	Joint NLL ↓	Type NLL ↓	W. cand. NLL ↓	Cond. NLL ↓	Top-1 ↑
P/P	2.746872	1.734154	1.012718	1.289912	36.73%
P/U	2.080695	1.067976	1.012718	1.289912	36.73%
F/U	2.044240	1.036397	1.007843	1.283703	40.26%
M/U	2.149770	1.136236	1.013534	1.279146	35.78%
H/P	2.657362	1.656747	**1.000614**	**1.274495**	**44.23%**
H/U (CAPA)	**1.923290**	**0.922675**	**1.000614**	**1.274495**	**44.23%**

Replay and metrics. The cohort contributes 2,016 state events across three sessions, with 31, 27, and 26 participants, respectively. Tables 1 and 3 report metrics for 1,416 scored events, including 1,094 choices; 496 calibration and 104 quality-control events update state without contributing predictive metrics. Predictions precede feedback. Cross-session first-event evaluation requires two completed preceding sessions, yielding 26 events, including 19 choices. Joint negative log-likelihood (NLL) uses natural logarithms over seven outcomes. Losses are first averaged within each participant and then averaged equally across participants. Conditional candidate NLL and Top-1 use eligible choices; weighted candidate NLL includes zero contribution on set-level events. Type plus weighted candidate NLL equals joint NLL before rounding. The human replay tables report participant-macro averages. Appendix C specifies the denominators.

Separate fixed-prefix query replay. Table 2 uses 248 first-session targets: 204 choices and 44 set-level responses. For each target, the personal state is initialized from eight calibration responses, without target-to-target feedback accumulation. After a question is selected, the replay assigns a *choice-consistent replay answer*: the group containing the recorded choice. The answer is derived from the recorded final selection, so this protocol measures candidate identification under consistent group feedback. A set-level response terminates the event after the first exposure. The cap is two questions. Total query cost sums $C(q)$ in Eq. (30) over exposed questions. It measures target-time engineering cost; acquiring the calibration prefix is outside this total.

4.2 Historical Preference Utilization

All conditions share candidates, networks, instruction priority, and event order. Flat states aggregate the same history, observable-axis masks, and confidences in one state; mismatched states are taken from a different participant. Table 1 compares preference-state configurations and response statistics. Hierarchical states reach 44.23% Top-1 versus 40.26% for flat states and 36.73% for population states. With population states fixed, personal response counts lower type NLL from 1.734154 to 1.067976 without changing conditional candidate predictions. The preference state also conditions the response head, so its effect can extend to type NLL. With history and the response network held fixed, the hierarchical configuration improves candidate-selection accuracy by 3.97 percentage points over flat aggregation.

4.3 Query Selection on Fixed Candidates

Policies share candidates, the initial semantic question pool, and a two-question cap. Variance utility evaluates Eq. (22) on this fixed semantic-axis question pool. The fixed-coverage and random-coverage policies select unasked axes, permitting identical partitions from different axes. DV considers only questions that split the remaining candidate set and improve its objective. The uniform-probability control retains the same choice mass g but uses uniform conditional probabilities for selection; answer updates and final scoring retain the common personal scorer. These comparisons evaluate complete policy configurations, including their question filtering, ranking, and stopping rules.

In Table 2, personal one-step DV reaches 80.14% Top-1 versus 67.75% for fixed coverage, with 326 versus 452 exposures and 33.64% lower proxy cost. Uniform-probability selection reaches 75.91%.

Table 2: Fixed-prefix replay with choice-consistent answers: 248 targets (204 choices), eight calibration responses per target, and a two-question cap. NLL and choice Top-1 are participant-macro; exposures, group answers, and costs are totals. Set-level replies consume exposures. Cost excludes prefix acquisition. Bold marks best NLL, Top-1, and cost.

Strategy	Joint NLL ↓	Top-1 ↑	Exposures / answers	Proxy cost ↓
No query	1.9115	27.00%	-/-	-
Variance utility	1.6323	42.38%	185/155	**21.1620**
Fixed coverage	1.2479	67.75%	452/408	61.3163
Random coverage	1.1746	69.34%	452/408	60.1862
CAPA-DV				
One-step DV / personal	1.0868	80.14%	326/308	40.6897
One-step DV / uniform	1.1361	75.91%	344/324	41.5212
Budgeted DV / personal	**0.9329**	**90.38%**	385/356	48.4024

Table 3: Participant-macro retention replay. First NLL uses 26 cross-session first events; First Top-1 uses 19 choices. Screened promotion is Table 1's H/U condition. Bold marks column optima.

Strategy	Replay NLL ↓	First NLL ↓	First Top-1 ↑
No long-term writes	1.9339	2.9414	31.58%
Immediate	1.9456	2.5396	31.58%
Screened promotion	**1.9233**	**2.2025**	**42.11%**

Budgeted DV reaches 90.38% using 385 exposures, acquiring more information at greater cost than one-step DV. Personal one-step DV exceeds uniform selection by 4.23 percentage points under the same DV rules. Relative to fixed coverage, the complete CAPA-DV configuration improves identification with fewer exposures; budgeted planning uses additional questions to raise accuracy further.

4.4 Long-Term Promotion

The event stream and within-session updates are fixed. No long-term writes retains session state; Immediate persists available directional evidence; Screened promotion applies ledger eligibility. Table 3 reports first-event NLL of 2.2025 for Screened promotion, 2.5396 for Immediate, and 2.9414 without long-term writes. Screened promotion predicts eight of 19 first-event choices, versus six for either control. The 26 promotion operations respect the rule excluding temporary records. The comparison includes the policies' commit timing and batching; the controlled test below examines cross-session eligibility separately.

4.5 Synthetic Query Stress Test

We generate 4,096 four-candidate cases from 256 virtual profiles, using nine numeric axes and supplied probabilities. The sample contains 3,154 choice cases and 942 set-level cases. Policies share midpoint partitions, costs, and the two-question cap. Binary answers flip with probability $\epsilon \in \{0, 0.1, 0.2\}$; targets remain fixed. The same pre-sampled random numbers are reused across policies and noise levels. The entropy control maximizes choice-mass-weighted expected conditional entropy reduction minus question cost, with the same stopping threshold. All query policies use unasked axes that split the remaining candidate set and select questions before seeing the target. Top-1 is micro-averaged over choices; question exposures and proxy costs are averaged over all cases. Appendix D specifies the generator and execution rules.

Table 4 shows that noiseless budgeted DV achieves 94.61% accuracy at cost 0.0943 per case, versus 98.38% at 0.1222 for entropy gain. At 20% flips, one-step and budgeted DV achieve 73.15% and 73.24%, respectively, versus 70.77% for entropy gain, indicating little benefit from additional lookahead. Under hard conditioning, incorrect answers can permanently exclude the target, exposing an accuracy–cost trade-off as noise increases.

Table 4: Synthetic querying. Top-1 uses 3,154 choices; Q_0 and C_0 are mean exposures and proxy cost per case at $\epsilon = 0$. Table 7 gives all noise levels. Bold marks column optima.

	Top-1 (%) ↑				
Policy	$\epsilon = 0$	0.1	0.2	$Q_0 \downarrow$	$C_0 \downarrow$
No query	59.48	59.48	59.48	-	-
Variance utility	95.28	77.96	64.14	1.651	0.1582
Entropy gain	**98.38**	**84.12**	70.77	1.367	0.1222
One-step DV	91.63	82.44	73.15	**0.982**	**0.0858**
Budgeted DV	94.61	83.70	**73.24**	1.067	0.0943

Table 5: Synthetic retention: mean stored-scene MSE over 128 streams per regime after session-3 and session-6 promotion (Appendix D.2). Bold marks column minima within each regime.

Regime	Policy	Session 3 ↓	Session 6 ↓
Stationary	Immediate	0.0643	0.0594
	Filtered boundary	**0.0156**	0.0156
	Full promotion	0.0221	**0.0091**
Flagged perturbation	Immediate	0.0790	0.0595
	Filtered boundary	**0.0221**	0.0157
	Full promotion	**0.0221**	**0.0113**
Unmarked perturbation	Immediate	0.0790	0.0595
	Filtered boundary	0.0459	0.0163
	Full promotion	**0.0221**	**0.0108**

4.6 Synthetic Retention Test

Each regime contains 128 streams with six sessions and eight choices per scene in two scenes per session. Candidates vary on warmth and brightness. Stored-scene mean squared error (MSE) compares these axes against the expected single-event summary under the unperturbed distributions, fixed across regimes. Regimes are stationary or reverse warmth in session three only, with or without a temporary flag. Immediate writes every observation; Filtered boundary batches non-temporary evidence per session; Full promotion also requires cross-session support. Policies share streams and slot operations. All responses are unit-confidence choices, so the active contrast concerns temporary scope and cross-session eligibility.

In Table 5, both filtering policies exclude flagged evidence, giving MSE 0.0221 after session three. For unmarked perturbations, Full promotion gives 0.0221 versus 0.0459 for Filtered boundary. Under stationary preferences, it initially trails Filtered boundary (0.0221 versus 0.0156), then achieves lower error by session six (0.0091 versus 0.0156). Cross-session eligibility therefore limits transient perturbations while slowing initial accumulation. Appendix D.2 specifies the evaluation.

5 Discussion and Conclusion

CAPA connects comparative feedback to its immediate use and persistent reuse through the provenance of rendered-candidate comparisons. Its remaining-axis ledger tracks which coordinates can be promoted now and which retain evidence for later sessions, while CAPA-DV directs questions toward candidate differences that affect selection. Hierarchical memory improves replay Top-1 by 3.97 percentage points over same-history flat aggregation; personal decision-value querying improves identification with fewer exposures than fixed coverage. The synthetic tests reveal the benefits and adaptation delay of cross-session promotion, as well as the accuracy–cost tradeoff under noisy answers.

AI USE STATEMENT

Gemini 3.5 Flash generated synthetic supervision for the six structured interfaces described in Appendix B; ChatGPT 5.5 checked the generated outputs against the task requirements. Generative AI also assisted with manuscript review and revision, including feedback on methodology and evaluation reporting, interpretation of the reported results, literature and reference checks, language editing, and LaTeX preparation. AI-assisted checks compared mathematical definitions, cross-references, and result descriptions with the manuscript and tables. The authors take responsibility for the final content, including text, claims, and artifacts produced with generative AI.

ETHICS STATEMENT

The study involved 31 participants who took part voluntarily and provided informed consent before the experiment. We collected only photo-retouching choices and preferences and did not collect sensitive personal information. Participants are identified in the paper only by anonymous identifiers (User 1–31). Ethics approval was not required for this study.

REPRODUCIBILITY STATEMENT

Section 3 describes the state updates, promotion rules, and clarification policies. Appendix A documents the semantic descriptors, model architectures and features, training objectives, and calibration procedures. Appendix B describes synthetic supervision and source grouping. Appendix C gives cohort counts, metric denominators, and replay chronology; Appendix D specifies the controlled query and retention generators and their evaluation metrics.

REFERENCES

Shuai Bai, Yuxuan Cai, Rui-Zhe Chen, Ke qin Chen, Xiong-Hui Chen, Zesen Cheng, Liang-Hao Deng, Wei Ding, Rongyao Fang, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Qidong Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Ming-Sheng Li, Mei Li, Kaixin Li, Zicheng Lin, Junyang Lin, Xue-Jing Liu, Jiawei Liu, Cheng-Long Liu, Yang Liu, Dayiheng Liu, Shixuan Liu, Dunjie Lu, Ruilin Luo, Chen-Xu Lv, Rui Men, Lingchen Meng, Lingchen Meng, Xin yi Ren, Sibo Song, Yu chen Sun, Jun Tang, Jianhong Tu, Jian-Qiang Wan, Peng Wang, Pengfei Wang, Qiu-Yue Wang, Yuxuan Wang, Tian-Bao Xie, Yihe Xu, Haiyang Xu, Jin Xu, Zhibo Yang, Mingkun Yang, Jianxin Yang, An Yang, Bo-Wen Yu, Fei Zhang, Hang Zhang, Xi Zhang, Botao Zheng, Humen Zhong, Jingren Zhou, Fanxi Zhou, Jingren Zhou, Yuanzhi Zhu, and Keming Zhu. Qwen3-vl technical report. *ArXiv*, abs/2511.21631, 2025. URL https://api.semanticscholar.org/CorpusID:283262018.

Craig Boutilier. A pomdp formulation of preference elicitation problems. In *AAAI/IAAI*, 2002. URL https://api.semanticscholar.org/CorpusID:2285985.

Zewei Chang, Zheng-Peng Duan, Jianxing Zhang, Chun-Le Guo, Siyu Liu, Hyung-Ju Chun, Hyunhee Park, Zi-Kun Liu, and Chong-Yi Li. Pertouch: Vlm-driven agent for personalized and semantic image retouching. In *AAAI Conference on Artificial Intelligence*, 2025. URL https://api.semanticscholar.org/CorpusID:283073337.

Haoyu Chen, Keda Tao, Yizao Wang, Xinlei Wang, Lei Zhu, and Jinjin Gu. Photoartagent: Intelligent photo retouching with language model-based artist agents. *ArXiv*, abs/2505.23130, 2025. URL https://api.semanticscholar.org/CorpusID:278996941.

Wonjun Choi, Yerim Kim, Yukyung Lee, and Susik Yoon. Pgmem: Tightly coupled persona-memory graph for lifelong personalized agents. 2026. URL https://api.semanticscholar.org/CorpusID:290830432.

Duc-Tien Dang-Nguyen, Cecilia Pasquini, Valentina Conotter, and Giulia Boato. Raise: a raw images dataset for digital image forensics. *Proceedings of the 6th ACM Multimedia Systems Conference*, 2015. URL https://api.semanticscholar.org/CorpusID:15349589.

Connor Dunlop, Matthew Zheng, Kavana Venkatesh, and Pinar Yanardag. Personalized image editing in text-to-image diffusion models via collaborative direct preference optimization. *ArXiv*, abs/2511.05616, 2025. URL https://api.semanticscholar.org/CorpusID:282912824.

Meera Hahn, Wenjun Zeng, Nithish Kannen, Rich Galt, Kartikeya Badola, Been Kim, and Zi Wang. Proactive agents for multi-turn text-to-image generation under uncertainty. *ArXiv*, abs/2412.06771, 2024. URL https://api.semanticscholar.org/CorpusID:274610848.

Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (TOG)*, 35:1 – 12, 2016. URL https://api.semanticscholar.org/CorpusID:2344617.

Asif Hossain, Ruksat Khan, Shayoni, and Kishor Morol. Capture: Disentangling preference drift from memory poisoning in personalized llm agents. 2026. URL https://api.semanticscholar.org/CorpusID:291611258.

Ronald A. Howard. Information value theory. *IEEE Trans. Syst. Sci. Cybern.*, 2:22–26, 1966. URL https://api.semanticscholar.org/CorpusID:206799032.

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. *ArXiv*, abs/2106.09685, 2021. URL https://api.semanticscholar.org/CorpusID:235458009.

Yeonjun In, Wonjoong Kim, Sang Yoon Park, Kanghoon Yoon, and Chanyoung Park. Personalize-then-store: Benchmarking and learning personalized memory for long-horizon agents. *ArXiv*, abs/2605.25535, 2026. URL https://api.semanticscholar.org/CorpusID:288671292.

Jinwoo Kim, Jihye Yoo, and Seon Joo Kim. Learning personalized photographic style from pairwise user preferences. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1134–1144, June 2026.

Kaiqu Liang, Julia Kruk, Shengyi Qian, Xianjun Yang, Shengjie Bi, Yuanshun Yao, Shaoliang Nie, Mingyang Zhang, Lijuan Liu, Jaime Fernández Fisac, Shu-Yan Zhou, and Saghar Hosseini. Learning personalized agents from human feedback. *ArXiv*, abs/2602.16173, 2026. URL https://api.semanticscholar.org/CorpusID:285725843.

Ofir Nabati, Guy Tennenholtz, Chih-Wei Hsu, Moonkyung Ryu, Deepak Ramachandran, Yinlam Chow, Xiang Li, and Craig Boutilier. Preference adaptive and sequential text-to-image generation. In *International Conference on Machine Learning*, 2024. URL https://api.semanticscholar.org/CorpusID:274776736.

Manan Suri, Puneet Mathur, Nedim Lipka, Franck Dernoncourt, Ryan A. Rossi, and Dinesh Manocha. Structured uncertainty guided clarification for llm agents. In *Annual Meeting of the Association for Computational Linguistics*, 2025. URL https://api.semanticscholar.org/CorpusID:282939794.

Heng Wang, Yifei Li, Lingling Zhang, Peng-Yu Li, Xin-Yu Che, Xinyu Zhang, and Ze-Sheng Yang. Qumem: Personalized memory for query-conditioned user-state inference in llm agents. 2026a. URL https://api.semanticscholar.org/CorpusID:291183670.

Jing-Xuan Wang, Yi-Fan Mei, Yu-Xia Niu, Chaowan Jiao, and Qijin Shen. Palate: Personalized aesthetic learning through adaptive taste evolution for multi-user portrait retouching. 2026b. URL https://api.semanticscholar.org/CorpusID:291217491.

Temesgen Muruts Weldengus, Binnan Liu, Fei Kou, Youwei Lyu, Jinwei Chen, Qi Fan, and Changqing Zou. Refretouch: Personalized image retouching without test-time fine-tuning. 2025. URL https://api.semanticscholar.org/CorpusID:285787463.

Yuzhe Yang, Liwu Xu, Leida Li, Nan Qie, Yaqian Li, Peng Zhang, and Yandong Guo. Personalized image aesthetics assessment with rich attributes. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19829–19837, 2022. URL https://api.semanticscholar.org/CorpusID:247839464.

Table 6: Semantic axes constructed from normalized Lightroom schema fields. An axis containing multiple fields uses their mean, except where stated.

Axis	Lightroom schema fields
Warmth	`IncrementalTemperature`
Tint	`IncrementalTint`
Brightness	`Exposure2012`, `Shadows2012`, `Whites2012`
Contrast	`Contrast2012`, `Clarity2012`, `Dehaze`
Saturation	`Saturation`, `Vibrance`
Detail	`Texture`, `Clarity2012`, `Sharpness`
Grain	`GrainAmount`
Vignette	Negated `PostCropVignetteAmount`
Matte	`Shadows2012`, `Blacks2012`

A Implementation Details

A.1 Semantic Axes and State Initialization

For a Lightroom parameter with schema bounds $[L_p, U_p]$, normalization uses $\widetilde{p} = 2(p - L_p)/(U_p - L_p) - 1$. The semantic representation is constructed using the schema mapping in Table 6. Each mean uses available members only; an empty group returns zero. Every resulting coordinate is clipped to $[-1, 1]$. The common scale is dimensionless. `Clarity2012` contributes to contrast and detail, and `Shadows2012` contributes to brightness and matte.

Given N preference-state summaries from synthetic training records, the population source is initialized by

$$\begin{aligned} \mu_a^{pop} &= \frac{1}{N}\sum_i \mu_{i,a}, \\ v_a^{pop} &= \max\left(10^{-4}, \frac{1}{N}\sum_i (v_{i,a} + \mu_{i,a}^2) - (\mu_a^{pop})^2\right), \qquad n_a^{pop} = N. \end{aligned} \tag{9}$$

The same N summaries contribute to every axis, and the prior is fixed before human replay. Individual state updates accumulate confidence in n; subsequent fusion uses Eq. (1) for n as well as the first two moments.

A.2 Promotion, Slot Assignment, and Ledger Consumption

The ledger records image, scene, session, temporary scope, confidence, response type, and a remaining-axis mask. Each choice also stores the single-event state obtained by applying Eq. (4) to an empty source. Promotion is evaluated for every scene with pending records at a session boundary, before the next session's first prediction.

Let U_s contain the scene's pending, non-temporary records with confidence at least 0.7. If $U_s = \varnothing$, no promotion is attempted. Otherwise, define the response-type conflict score

$$\chi_s = \frac{1}{|U_s|}\sum_{t \in U_s} \mathbb{I}[y_t \in \{\text{all-dislike}, \text{cannot-tell}\}]. \tag{10}$$

An attempt proceeds when $\chi_s \le 0.4$. An axis qualifies if its choice records with a remaining mask on that axis cover at least two images and two sessions. Confidence-normalized weights fuse those records' single-event states into a promoted summary. Unqualified axes remain untouched. Set-level records affect χ_s only; they contribute neither directional evidence nor image/session support. The promoted summary updates the matching-scene memory and global backoff using the rules below.

If no slot exists for the scene, the first promotion creates one. Otherwise, let J be the axes supported by both an existing slot and the promoted summary. For $J \neq \varnothing$, their distance is

$$d_{\text{slot}} = \sqrt{\frac{1}{|J|}\sum_{a \in J} (\mu_a^{slot} - \mu_a^{new})^2}. \tag{11}$$

Among slots sharing supported axes, the closest is selected. If its distance exceeds 0.75 and fewer than three slots exist for the scene, a new slot is created; otherwise, the selected slot and promoted summary are fused with equal weights. If no slot shares a supported axis, the slot with the smallest total support is updated. Global memory combines the old and promoted summaries using axis-wise support-proportional weights. Scene retrieval combines the available slots by the same support-proportional rule.

All these operations use Eq. (1), including its weighted average of support. Thus, n records effective support rather than a sum of observation counts. Recursive support-weighted fusion depends on the order and batching of promotions; the specified session-boundary schedule is part of the evaluated policy.

Consumed axis flags are cleared in their source records. A choice record is removed once all its axes are consumed; records with remaining axes stay pending. The scene's ledger is cleared when no record retains an observable axis. Consequently, one event can contribute different axes at different promotion attempts, but an axis is consumed only once. One successful scene-level attempt counts as one promotion, irrespective of the number of axes updated. The final flush after the last observed session affects operation counts without changing earlier predictions. The global state pools promoted evidence, including evidence from a single scene; promotion does not test cross-scene invariance or agreement between evidence signs.

A.3 Response Scorer Features and Architecture

For candidate set $\mathcal{C}_t = \{c_{t,1}, \dots, c_{t,K}\}$, the full response space is

$$\mathcal{Y}_t = \mathcal{C}_t \cup \{\text{all-like}, \text{all-dislike}, \text{cannot-tell}\}. \tag{12}$$

The response scorer operates on structured descriptors rather than image pixels. Write $x_j = \phi(c_j)$ and use (μ, v) for the effective state. Let m_X and s_X be the coordinate-wise mean and uncorrected standard deviation across the current candidates, $\delta_j = x_j - \mu$, $r_j = x_j - m_X$, and

$$a_j = \frac{1}{9} \sum_a \frac{\delta_{j,a}^2}{\max(v_a, 10^{-4})}. \tag{13}$$

The candidate ranker receives the 87-dimensional feature vector

$$\begin{aligned} f_j = [&x_j, \mu, v, \delta_j, m_X, s_X, r_j, a_j, \cos(x_j, \mu), \\ &\|x_j\|/3, \|r_j\|/3, x_j^\top \mu/3, \text{mean}(v), z, x_j \odot z]. \end{aligned} \tag{14}$$

The vector contains nine nine-dimensional blocks and six scalars: $9 \times 9 + 6 = 87$. Cosine similarity uses a norm floor of 10^{-6}. The ranker has two 384-unit hidden layers. Its utilities are

$$\begin{aligned} u_j &= R_\theta(f_j) + (1 + \text{softplus}(\beta)) x_j^\top z/3, \\ u'_j &= u_j + R_\theta^h([0_9, 0_9, |x_j|, |x_j|, 0, 0]), \\ \rho_j^\theta &= \text{softmax}_j(u'_j). \end{aligned} \tag{15}$$

Here β is the scalar controlling the additional instruction-alignment term, and $|x_j|$ denotes element-wise absolute value. The $38 \to 192 \to 1$ residual R_θ^h receives, in order, two zero nine-vectors, two copies of $|x_j|$, and two zero scalars. The stated input therefore contains 18 candidate-feature entries and 20 zeroed history-summary entries.

The raw set summary s_{raw} concatenates five coordinate-wise candidate statistics (mean, maximum, minimum, standard deviation, and range), $(\mu, v, z, \|z\|_\infty)$, and four statistics of a_j (mean, standard deviation, minimum, and maximum). It has $5 \times 9 + 3 \times 9 + 1 + 4 = 77$ dimensions. An $87 \to 384 \to 192$ encoder acts on the candidate features f_j. Pooling its outputs by mean, maximum, minimum, and standard deviation gives $s_{\text{pool}} \in \mathbb{R}^{768}$. The eight base-utility statistics s_{util} are mean, standard deviation, minimum, maximum, range, top-two gap, normalized softmax entropy, and maximum softmax probability; gradients are stopped through these statistics. The response-head input is

$$h_{\text{resp}} = [s_{\text{raw}}, s_{\text{pool}}, s_{\text{util}}] \in \mathbb{R}^{77+768+8} = \mathbb{R}^{853}. \tag{16}$$

Two response heads, $853 \to 384 \to 192 \to 2$ and $853 \to 384 \to 192 \to 3$, predict choice versus set-level feedback and the conditional set-level distribution. Their softmax outputs $b \in \mathbb{R}^2$ and $s \in \mathbb{R}^3$ define

$$g^\theta = (b_{\text{choice}}, b_{\text{set}} s_{\text{all-like}}, b_{\text{set}} s_{\text{all-dislike}}, b_{\text{set}} s_{\text{cannot-tell}}). \tag{17}$$

Zero-input history residuals contribute learned fixed logit offsets. The complete response distribution is $(g^\theta_{\text{choice}} \rho^\theta_{1:4}, g^\theta_{\text{all-like}}, g^\theta_{\text{all-dislike}}, g^\theta_{\text{cannot-tell}})$.

Appendix A.7 gives the fitting objective, ensemble calibration, and session-count adaptation.

A.4 Structured Interfaces and Parameter Prediction

The six structured interfaces—`observer`, `instruction_parser`, `feedback_parser`, `planner`, `question_plan`, and `question_parameter`—share one Qwen3-VL-8B-Instruct backbone and one rank-16 LoRA adapter. The shared adapter is trained for two epochs. Its objective is completion-token cross-entropy:

$$\mathcal{L}_{\text{interface}} = -\frac{1}{\sum_i |O_i|} \sum_i \sum_{k \in O_i} \log P_\theta(o_{i,k} \mid x_i, o_{i,<k}), \tag{18}$$

where O_i is the set of supervised assistant-output token positions for example i, and x_i denotes its conditioning input. Prompt and padding positions are masked. At deployment, all six interfaces call this adapted backbone.

Each parameter network has two 256-unit hidden layers. It receives base Lightroom parameters, scene, a sparse nine-axis direction vector, and a strength input fixed to one by the online wrapper. Strength adjustment is handled by the surrounding controller. Runtime averages the projected deltas of three independently seeded parameter models. The interface, response-scoring, and parameter networks have separate parameter tensors and training objectives. Within each component family, neural weights are shared across users and remain fixed during interaction.

For predicted and target deltas δ, δ^*, base parameters p^0, projected predictions p^+, positive delta bounds b_k, and an active parameter mask M_{act}, parameter training uses

$$\begin{aligned} \mathcal{L}_{\text{parameter}} = \; & \langle h(\delta_k/b_k, \delta_k^*/b_k) \rangle_{M_{\text{act}}} \\ & + 0.2 \, \langle [0.01 - (\delta_k/b_k)\,\text{sign}(\delta_k^*)]_+ \rangle_{M_{\text{act}}} \\ & + 0.01 \, \langle |\delta_k/b_k| \rangle_{\neg M_{\text{act}}} \\ & + 0.5 \, \left\langle h(p_k^+/b_k, (p_k^0 + \delta_k^*)/b_k) \right\rangle_{M_{\text{act}}}. \end{aligned} \tag{19}$$

Here h is smooth-L1; brackets denote masked means over active or inactive parameter entries, respectively. The terms fit normalized deltas, encourage the target direction, penalize inactive adjustments, and supervise projected parameters. The last term uses the delta scale b_k because its residual is $(p_k^+ - p_k^0 - \delta_k^*)/b_k$; a common base-parameter offset cancels in the smooth-L1 difference.

A.5 Online Question Routing and Execution Profile

The structured question interface supplies $\psi(q)$ without a unit-norm constraint. Plan-direction questions refer to the alternative plans' semantic axes; parameter questions resolve strength on one axis. For previously exposed questions $\mathcal{Q}_t^{\text{asked}}$ in the current session, redundancy is

$$R_t(q) = \begin{cases} 0, & \mathcal{Q}_t^{\text{asked}} = \varnothing, \\ \max\limits_{q' \in \mathcal{Q}_t^{\text{asked}}} \dfrac{|\psi(q)^\top \psi(q')|}{\max(\|\psi(q)\|\|\psi(q')\|, 10^{-8})}, & \mathcal{Q}_t^{\text{asked}} \neq \varnothing. \end{cases} \tag{20}$$

For plan effects $b_{k,a}$ and their union of axes $\mathcal{A}$, planning sensitivity is

$$D_t = \begin{cases} 0, & \mathcal{A} = \varnothing, \\ \max\left(10^{-3}, \dfrac{1}{|\mathcal{A}|} \displaystyle\sum_{a \in \mathcal{A}} \text{Var}_k \, b_{k,a}\right), & \mathcal{A} \neq \varnothing. \end{cases} \tag{21}$$

The cost-aware online router multiplies plan sensitivity by 1.4. Parameter sensitivity is one. Plan-question cost is the nonnegative `interaction_cost` provided by the question interface; parameter-question cost follows Eq. (30). Optional forced plan-to-parameter follow-ups can bypass the utility threshold while respecting the hard budget. CAPA-DV uses the fixed-candidate decision rule independently of those online follow-up routes.

Query selection for generative editing. Let $m_{t,a} = \mathbb{I}[z_{t,a} \neq 0]$ and $v_{t,a}^{eff} = m_{t,a}10^{-3} + (1 - m_{t,a})\bar{v}_{t,a}$ be the instruction-adjusted variance; Eq. (23) gives the corresponding mean. For question direction $\psi(q)$, the editing router uses the utility

$$U_t(q) = \log\left(1 + \sum_{a=1}^{9} v_{t,a}^{eff}\psi_a(q)^2\right) D_t - C(q) - \lambda R_t(q). \tag{22}$$

Here $D_t \geq 0$ measures variation in plan effects, $C(q)$ is a cost proxy, and $R_t(q)$ is the maximum absolute cosine similarity to an earlier question. The base policy chooses the highest-utility question when $U_t(q) > \eta = 0.05$, using $\lambda = 0.2$. Exposing a question consumes budget.

Historical replay uses the axis-masked state configuration and the structured-feature response scorer defined above. The original editing prototype retains its scalar-support Gaussian memory profile: `all_like` and `all_dislike` make weak zero-centered updates, and its `ignore` policy skips state updates for `cannot_tell`. A set-level reply returns a feedback receipt and ends the online turn. The structured response scorer supplies offline replay predictions and is not an additional neural reranker in that online prototype.

A.6 State Fusion and Instruction Priority

Preference fusion. Empty personal sources initialize to $(\mathbf{0}, \mathbf{1}, \mathbf{0})$. The population prior is estimated from the synthetic-training state summaries. Define $\mathcal{F}_{\mathbf{w}}(\mathcal{M}^{(1)}, \mathcal{M}^{(2)})$, $\mathbf{w} \in [0,1]^9$, by Eq. (1) with nominal axis-wise weights $(1 - w_a, w_a)$. Sources with zero support are excluded and the remaining weights are normalized. If the second source has no support, that axis is copied from the first; if both are empty, it remains empty. The matching scene's slots are first combined with support-proportional weights. Retrieval fuses population and global states with nominal weight $w_a = 0.8$ on the global source for every axis, fuses that result with the matching scene summary using the same weight, and incorporates the session state with $w_a = n_a^{\text{session}}/(1 + n_a^{\text{session}})$. This order fixes the effective source weights $\alpha_{\ell,a}$ in Eq. (1).

Instruction priority. The instruction vector $z_t \in \{-1, 0, +1\}^9$ encodes resolved editing directions: $+1$ for an increase, -1 for a decrease, and 0 when no direction is specified. With $m_{t,a} = \mathbb{I}[z_{t,a} \neq 0]$, the effective state is

$$\begin{aligned} \mu_{t,a}^{eff} &= m_{t,a}z_{t,a} + (1 - m_{t,a})\bar{\mu}_{t,a}, \\ v_{t,a}^{eff} &= m_{t,a}10^{-3} + (1 - m_{t,a})\bar{v}_{t,a}. \end{aligned} \tag{23}$$

The override applies to the current decision and leaves stored summaries unchanged. Thus, *warmer* establishes a positive warmth direction; the parameter controller and optional strength question determine its magnitude. Direction priority and strength resolution play distinct roles.

A.7 Scorer Fitting and Response Calibration

Training and ensemble calibration. The scorer's objective combines type prediction, conditional ranking, pairwise margin, probability accuracy, and permutation consistency:

$$\mathcal{L} = \mathcal{L}_{\text{type}}^{w} + 1.5\mathcal{L}_{\text{cond}} + 0.35\mathcal{L}_{\text{pair}} + 0.1\mathcal{L}_{\text{Brier}} + 0.05\mathcal{L}_{\text{perm}}. \tag{24}$$

The type term $\mathcal{L}_{\text{type}}^{w}$ is four-type negative log-likelihood (NLL), normalized by the sum of example weights. Type weights are proportional to training frequencies raised to -0.35, with an additional

factor of two for choice. Conditional NLL is averaged over specific-choice examples. The same example-wise average, denoted by $\langle\cdot\rangle_{\text{choice}}$, defines the margin loss

$$\mathcal{L}_{\text{pair}} = \left\langle [0.2 - \log \rho_y^\theta + \max_{j\in\{1,\dots,4\}\setminus\{y\}} \log \rho_j^\theta]_+ \right\rangle_{\text{choice}}. \tag{25}$$

The Brier loss sums squared probability errors over seven classes and averages over examples. The permutation-consistency loss is the mean-squared difference between probability vectors after randomly permuting candidates and undoing the permutation of the four candidate outputs. All seven entries are included, and the original prediction is treated as a stop-gradient target. Ensemble averaging, gate calibration, and count adaptation follow the order below.

The first five epochs optimize the conditional and margin terms; the complete objective is used thereafter. Three independently seeded models produce seven-class distributions p^{θ_k}, which are averaged before further calibration:

$$\begin{aligned} \bar{p} &= \frac{1}{3}\sum_{k=1}^{3} p^{\theta_k}, \qquad \rho_j = \frac{\bar{p}_j}{\bar{g}_{\text{choice}}}, \\ \bar{g} &= (\textstyle\sum_{j=1}^{4} \bar{p}_j, \bar{p}_5, \bar{p}_6, \bar{p}_7). \end{aligned} \tag{26}$$

The gate order is choice, all-like, all-dislike, and cannot-tell. Before taking logarithms, each gate probability is lower-bounded by 10^{-8}: $h_r = \log \max(\bar{g}_r, 10^{-8})$. Calibration then uses

$$\widetilde{g} = \text{softmax}\left(W_{\text{cal}}[h - \text{mean}_r\, h_r] + b_{\text{cal}}\right), \tag{27}$$

where $W_{\text{cal}} \in \mathbb{R}^{4\times4}$ and $b_{\text{cal}} \in \mathbb{R}^4$ are learned on synthetic development records. Class-balanced multinomial logistic calibration uses the regularization parameter $C_{\text{reg}} = 1$ and five-fold user-grouped out-of-fold checks before fitting the full 1,000-record development set.

Session-specific response adaptation. Four counts $N_{t,r}$ accumulate revealed response types in the current session and reset at the next session boundary. Every response contributes once, including temporary and explicit-instruction events. Types follow the gate order above. With training frequencies $\pi = (0.50, 0.05, 0.01, 0.44)$, the adapted probabilities are

$$\begin{aligned} q_{t,r} &= \frac{N_{t,r} + 4\pi_r}{\sum_k N_{t,k} + 4}, \\ g_{t,r} &= \frac{\widetilde{g}_{t,r}(q_{t,r}/0.25)}{\sum_k \widetilde{g}_{t,k}(q_{t,k}/0.25)}, \\ p_{t,j} &= g_{t,\text{choice}}\rho_{t,j}, \qquad p_{t,r} = g_{t,r} \quad (r \in \mathcal{R}). \end{aligned} \tag{28}$$

The factor $q_{t,r}/0.25$ reweights the class-balanced gate relative to a uniform four-type reference prior. The sums over k run over these four types. The population-count control substitutes $q_t = \pi$. For a nonzero explicit directive, the scorer uses $g_t = \bar{g}_t$ and $p_t = \bar{p}_t$, bypassing post-hoc calibration and count adaptation; the response still updates the counts afterward. Changing counts alone preserves ρ, whereas changing preference state can affect both heads. Define $\text{type}(y) = \text{choice}$ for $y \in \mathcal{C}_t$, and $\text{type}(y) = y$ for $y \in \mathcal{R}$. The joint response likelihood decomposes as

$$-\log p_t(y) = -\log g_{t,\text{type}(y)} + \begin{cases} -\log \rho_{t,y}, & y \in \mathcal{C}_t, \\ 0, & y \in \mathcal{R}. \end{cases} \tag{29}$$

A.8 Fixed-Candidate Clarification

Fixed-candidate partitions and conditioning. Each nonconstant semantic axis defines a binary question whose threshold is the midpoint between that axis's minimum and maximum candidate values; the lower group includes the threshold. Partitions remain fixed throughout the event, and an exposed axis is removed from the question pool. For spread $d_a = \max_j \phi_a(c_j) - \min_j \phi_a(c_j)$, the cost is

$$C(q_a) = \text{round}_4\left(0.05 + \frac{0.1}{1 + d_a}\right). \tag{30}$$

Here round_4 rounds to four decimal places. With $\phi_a \in [-1, 1]$, costs lie between approximately 0.0833 and 0.15; they are dimensionless proxies, not measured response times. After a nonempty group answer G, candidate probabilities are restricted to that group while the choice mass and set-level probabilities remain fixed:

$$p'_j = \begin{cases} \dfrac{g\, p_j \mathbb{I}[j \in G]}{\sum_{k \in G} p_k}, & \sum_{k \in G} p_k > 0, \\ \dfrac{g}{|G|} \mathbb{I}[j \in G], & \sum_{k \in G} p_k = 0, \end{cases} \qquad (31)$$
$$p'_r = p_r \quad (r \in \mathcal{R}).$$

Thus a zero-mass answer allocates g uniformly among candidates in the answer group G, using the original question partition. On this fallback branch, candidates excluded by an earlier answer can receive mass again. This update refines candidate identity within the choice branch and retains the initial response-type forecast. For $g > 0$, at the next query decision $\rho_j = p'_j/g$ denotes the current conditional distribution, including any uniform fallback. The lookahead tree conditions this distribution on the candidate subsets retained under each hypothetical answer.

One-step value identity. The one-step value follows from cancellation of the within-group normalizer. For $m_G = \sum_{j \in G} \rho_j$, sum over positive-mass groups:

$$g \sum_{\substack{G \in \mathcal{G}(q) \\ m_G > 0}} m_G \max_{j \in G} \frac{\rho_j}{m_G} = \sum_{G \in \mathcal{G}(q)} \max_{j \in G} p_j. \qquad (32)$$

Zero-mass groups contribute zero to the right-hand side.

Remaining-budget planning. Eq. (8) specifies the remaining-budget recursion, including its stopping value, conditional branch probabilities, and zero-budget boundary in Section 3.4. After each observed group answer, the next planning call uses the updated conditional distribution defined above, including the uniform fallback when necessary. The choice mass g remains fixed within and across these planning calls. The resulting objective is a fixed-g decision-value surrogate: terminal accuracy is multiplied by g, while future question costs are charged on choice-conditional paths. In an exact event-level utility for the original replay, future costs after the first answer would also be weighted by the probability of reaching the choice branch. The evaluated policy uses the stated surrogate, which assigns a larger relative cost to continuation when $g < 1$.

In the original choice-consistent replay, if the response is set-level, the first question is recorded as exposed and its cost is incurred, and the event terminates without a candidate-group answer. The original seven-class prediction is retained. Candidate Top-1 is evaluated on specific choices, while joint NLL includes set-level responses. In fixed-candidate replay, group answers are obtained from recorded choices after question selection; final target responses update the longitudinal state only under the chronology of the corresponding evaluation protocol.

B Synthetic Supervision

Simulator. Users are balanced across six handcrafted style centers: natural-balanced, warm-bright, cool-cinematic, vivid-crisp, soft-muted, and grain-matte. Global targets add independent Gaussian noise with standard deviation 0.16 to their center and are clipped to $[-1, 1]^9$. Each user has one to three scene labels; each scene has one preference slot, or two with probability 0.35. Slot means perturb the global target with standard deviation 0.22; the second slot additionally reverses one axis with a uniform $[-0.2, 0.2]$ perturbation. Slot selection weights are normalized $\mathrm{Gamma}(1.5, 1)$ draws floored at 0.05. An interaction samples the applicable scene slot, falling back to the global target when absent. Axis sensitivities are independently uniform on $[0.45, 1]$.

For each accepted candidate, utility is the negative square root of a weighted average of target-specific, sensitivity-weighted squared axis errors. Active long-term, short-term, and instruction targets have weights 1, 1.6, and 4, respectively; each target's axis errors and the active context weights are separately normalized. The simulator's short-term target is a confidence-weighted mean of previously chosen candidates' raw signed axis-coordinate vectors and resets at session boundaries. This

simulator target uses candidate coordinates; CAPA's session state uses the comparative contrasts in Eq. (2). On eligible odd-numbered turns in session two, a temporary instruction sets one varying axis among warmth, saturation, contrast, and brightness to -0.8 or 0.8, opposite to the active preference's sign; eligibility requires candidate span above 0.05. For 20% of simulated users, a one-axis shift sampled from $[-0.45, 0.45]$ starts in session two or three.

Set-level labels are assigned before sampling a choice. Let $u_{\max}$, $u_{\min}$, and u_{second} denote utility order statistics. User thresholds are sampled as $\tau_{\text{accept}} \sim U[-0.50, -0.35]$, $\tau_{\text{dislike}} \sim U[-0.90, -0.70]$, and $\epsilon \sim U[0.003, 0.020]$. The label is all_like if $u_{\min} \geq \tau_{\text{accept}}$ and $u_{\max} - u_{\min} \leq \epsilon$; otherwise all_dislike if $u_{\max} \leq \tau_{\text{dislike}}$; otherwise cannot_tell if $u_{\max} - u_{\text{second}} \leq \epsilon$. Remaining responses sample $\text{softmax}(u_j/T)$. Base temperature is uniform on $[0.08, 0.20]$, or $[0.25, 0.45]$ for a 20% high-noise cohort, and increases with interaction index by a user-specific fatigue factor in $[0.002, 0.018]$. Preference-training records are subsequently response-type resampled, so their class balance is not a natural human response frequency.

Source grouping and module construction. Synthetic split construction groups variants by original-file hashes for RAISE, source-session directories for PARA, and burst directories for the HDR+ Burst Photography Dataset (Hasinoff et al., 2016). Within each module's synthetic split, variants of a known original, source session, or burst remain together. User-disjoint and combined user/source-group partitions are constructed separately from the temporal session split. The parameter training package uses its own source-image split. The original question bank is sampled from the combined synthetic test interactions.

Gemini 3.5 Flash generates supervision for the six structured interfaces. ChatGPT 5.5 reviews these outputs against the task requirements. Parameter supervision is checked through rendering for adjustment direction and visible changes. During development, 30,000 query-supervision and 30,000 memory-supervision records, each with 1,000 development records, also support auxiliary learned policies. The final CAPA-DV and promotion rules use the explicit procedures in the method; those auxiliary models are outside the evaluated decision path. Module-specific packages may share source photographs and candidates. Human-evaluation photographs are separated from the synthetic training and development packages by source image. Human response labels are also excluded from neural training and calibration.

C Additional Replay Details

C.1 Cohort, Scoring Events, and Macro Averages

The original study involved 31 participants and produced 2,016 state events. Recruitment targeted individuals with undergraduate-level or higher education, with successive sessions separated by at least 24 hours. Each participant viewed an original and four rendered candidates, then selected one candidate or one of three set-level responses. The three sessions included 31, 27, and 26 participants, respectively. Tables 1 and 3 report metrics for the same 1,416 scored events: 1,094 choices and 322 set-level responses. The 496 calibration and 104 quality-control events update chronological state but are excluded from predictive metrics. Each prediction precedes the corresponding feedback. Scene labels are supplied to the replay; event-level temporary-intent flags come from the questionnaire. The first-event endpoint requires two completed previous sessions, yielding one event from each of 26 participants; 19 are choices.

For participant u, let T_u be the eligible scoring events and $C_u \subseteq T_u$ the choice events. Seven-outcome joint NLL uses natural logarithms. Each reported loss is first averaged within participant and then equally across participants. Weighted candidate NLL sums $-\log \rho_{t,y_t}$ over C_u and divides by $|T_u|$; conditional candidate NLL divides the same sum by $|C_u|$. Choice-only Top-1 follows the same participant-macro rule. Conditional metrics include only participants with at least one eligible choice. Type NLL plus weighted candidate NLL equals joint NLL before rounding. Different participant choice rates mean that weighted and conditional macro NLLs need not have a common ratio across models.

C.2 Fixed-Prefix Query Replay

Table 2 is a separate replay on 248 first-session targets from the same 31 participants: 204 choices and 44 set-level responses. For each target, the personal state is initialized from eight calibration responses; target feedback does not enter later target histories. The No query row evaluates this starting state. After question selection, the replay derives a choice-consistent replay answer from the recorded choice: the candidate group containing that choice. This protocol evaluates target identification under deterministic group membership. Set-level responses terminate the event after the first exposed question. NLL and choice-only Top-1 are macro-averaged over participants.

An *exposure* is a displayed question and consumes one unit of the question budget. A *candidate-group answer* identifies a partition group; a set-level reply consumes an exposure without producing such an answer. Exposure counts, candidate-group answer counts, and proxy costs are cumulative totals over the target events. Every exposure incurs the cost in Eq. (30), including an exposure followed by a set-level response. These target-time costs exclude acquiring the eight-response prefix and use an engineering scale rather than measured time or effort.

Every policy shares the initial semantic question pool and a maximum of two exposures. Variance utility uses Eq. (22). Fixed coverage follows a fixed available-axis order; random coverage samples available axes without replacement. Distinct axes may induce identical partitions for these coverage controls. CAPA-DV asks questions that split the remaining candidate set and improve its objective. The uniform-probability control keeps the same choice mass g and uses a uniform conditional distribution on the remaining candidate set S for selection only. Answer updates and final scoring retain the common personal scorer. Thus, the comparison includes question filtering, ranking, and stopping as parts of each policy configuration.

D Controlled Synthetic Evaluation

D.1 Query Inputs and Answer Noise

For profile u, draw a four-entry base probability vector from $\mathrm{Dirichlet}(2\mathbf{1}_4)$, floor each entry at 0.05, and renormalize to obtain q_u. Separately, draw a concentration α_u log-uniformly on $[2, 12]$. The profile's semantic tilt is $w_u \sim \mathcal{N}(0, 0.8^2 I_9)$, and its axis uncertainty scores $s_{u,a}$ are independently log-uniform on $[0.03, 1.5]$. Draw the profile choice mass $g_u \sim \mathrm{Uniform}[0.6, 0.95]$.

For each case i, sample context $h_i \sim \mathcal{N}(0, 0.25^2 I_9)$ and four descriptors $\phi_{i,j} = \tanh(\xi_{i,j} + h_i)$, with $\xi_{i,j} \sim \mathcal{N}(0, I_9)$. Draw $b \sim \mathrm{Dirichlet}(\alpha_u q_u)$, floor its entries at 10^{-12}, and normalize the following scores to obtain

$$\rho_{i,j} \propto b_j \exp\left(\frac{\phi_{i,j}^\top (w_u + h_i)}{\sqrt{9}}\right). \tag{33}$$

The case choice mass is $g_{u,i} = \mathrm{clip}(g_u + z_{u,i}, 0.5, 0.98)$, where $z_{u,i} \sim \mathcal{N}(0, 0.04^2)$. With probability $g_{u,i}$, the target is a candidate sampled from ρ_i; otherwise one of the three set-level types is sampled uniformly. The uncertainty inputs are $s_{u,i,a} = \mathrm{clip}(s_{u,a}(1 + 0.2|h_{i,a}|), 0.02, 2)$. Below, g and ρ denote the current case's supplied mass and conditional candidate distribution.

Questions split candidates at the axis midpoint. All policies are restricted to unasked axes that split the remaining candidate set into two nonempty groups. Variance utility uses the supplied uncertainty inputs with $\psi(q_a) = e_a$, $D = 1$, $\lambda = 0.2$, and $\eta = 0.05$. Distinct basis vectors have zero redundancy. The entropy control selects the largest positive value of

$$g\left[\mathsf{H}(\rho^S) - \sum_G P(G \mid S, \text{choice})\mathsf{H}(\rho^{S\cap G})\right] - C(q) - \eta, \tag{34}$$

where ρ^S is the normalized restriction in Eq. (7) and H uses natural logarithms. One-step DV uses Eq. (6); budgeted DV uses the fixed-g planning surrogate in Eq. (8). No policy receives the target label when selecting a question. Exact question ties use fixed axis order, and ties between continuation and stopping favor stopping.

For each case–axis pair, draw $u_{i,a} \sim \mathrm{Uniform}(0, 1)$ once before running any policy. A presented binary answer is flipped when $u_{i,a} < \epsilon$; the same values are reused across policies and noise lev-

Table 7: Synthetic query stress test: point estimates on 4,096 abstract cases from 256 virtual profiles, with four numeric candidates and supplied probabilities. Binary answers flip with probability ϵ using common random numbers; targets stay fixed. Top-1 is micro accuracy on the same 3,154 choice cases. Question exposures and engineering proxy costs are per case, including set-level outcomes. Bold marks column optima within each noise level.

ϵ	Policy	Top-1 (%) ↑	Exposures ↓	Proxy cost ↓
0.0	No query	59.48	-	-
0.0	Variance utility	95.28	1.651	0.1582
0.0	Entropy gain	**98.38**	1.367	0.1222
0.0	One-step DV	91.63	**0.982**	**0.0858**
0.0	Budgeted DV	94.61	1.067	0.0943
0.1	No query	59.48	-	-
0.1	Variance utility	77.96	1.627	0.1560
0.1	Entropy gain	**84.12**	1.383	0.1236
0.1	One-step DV	82.44	**0.979**	**0.0855**
0.1	Budgeted DV	83.70	1.074	0.0949
0.2	No query	59.48	-	-
0.2	Variance utility	64.14	1.612	0.1546
0.2	Entropy gain	70.77	1.393	0.1245
0.2	One-step DV	73.15	**0.977**	**0.0853**
0.2	Budgeted DV	**73.24**	1.085	0.0958

els. Set-level cases terminate after their first exposed question. Choice cases retain the original conditional probabilities restricted to the answered candidate set. Every asked partition splits the remaining set, so both answer branches have positive mass and the zero-mass fallback is inactive. Hard conditioning can remove the target after an incorrect answer. Final accuracy compares the predicted identity with the original sampled target, including these failures.

Top-1 is the number of correct predictions divided by the common number of choice cases. Question exposures and proxy costs are averaged over all 4,096 cases, including set-level and no-query cases. These synthetic case averages differ from the participant-macro averages in Tables 1–3. Table 7 reports the point estimates for the shared generated sample at each noise level.

D.2 Comparison-Summary Retention

The memory test uses 128 independent random streams with six sessions each. The four candidates occupy the ordered corners $(-0.6, -0.6)$, $(-0.6, 0.6)$, $(0.6, -0.6)$, and $(0.6, 0.6)$ on warmth and brightness; their other seven coordinates are zero. Scene A selects these candidates with probabilities $(0.1, 0.2, 0.2, 0.5)$; scene B uses $(0.2, 0.5, 0.1, 0.2)$. There are eight choices per scene per session, with distinct synthetic image identifiers and a randomized within-session order. Confidence is one for every observation.

Let e_j be the chosen candidate's contrast with the mean of the other three. The symmetric corners give $e_j = 4\phi_j/3$. Starting from an empty state, Eq. (4) with confidence one yields the single-event mean $0.8e_j$. With the active-axis set $\mathcal{J} = \{\text{warmth}, \text{brightness}\}$, the reference persistent summary is

$$\begin{aligned} \mu^*_s &= 0.8 \sum_{j=1}^{4} \pi_{s,j} e_j, \\ (\mu^*_{s=A})_{\mathcal{J}} &= (0.256, 0.256), \qquad (\mu^*_{s=B})_{\mathcal{J}} = (-0.256, 0.256). \end{aligned} \tag{35}$$

This reference is the expected single-event observational summary: the two scenes have opposite warmth contrasts and equal brightness contrasts. Let $\hat{\mu}_{r,s,a}$ be the support-weighted stored-slot readout for stream r, scene s, and axis a at the checkpoint. The reported error averages axes, scenes,

and streams equally:

$$\mathrm{MSE} = \frac{1}{128 \cdot 2 \cdot |\mathcal{J}|} \sum_{r=1}^{128} \sum_{s \in \{A,B\}} \sum_{a \in \mathcal{J}} (\hat{\mu}_{r,s,a} - \mu^*_{s,a})^2. \tag{36}$$

Slots are combined by their evidence support; the reference is never used to select a slot. The reference uses the unperturbed choice distributions in every regime and at both checkpoints, including the boundary after session three.

In the perturbed regimes, the probabilities in session three are permuted as $(\pi_3, \pi_4, \pi_1, \pi_2)$, reflecting warmth while preserving the brightness marginal. The flagged regime marks all session-three events temporary; the unmarked regime supplies no temporary flag. All other sessions use the original probabilities. Stationary and perturbed regimes share random uniforms and within-session orders, and all policies process identical observations within a regime.

The test executes the precision update, observable-axis mask, single-event states, moment fusion, scene slots, and pending-axis consumption from the method. Immediate commits every single-event state. Filtered boundary commits one batch of the session's non-temporary evidence for each scene. Full promotion retains pending evidence until its axes have support from two distinct images and two sessions, then consumes only promoted axes. All three use the same slot-distance threshold of 0.75, three-slot capacity, equal-weight updates to an existing slot, and support-weighted slot reading. Only Full promotion applies cross-image and cross-session eligibility. Immediate differs from the boundary policies in both commit timing and batching; the Filtered boundary–Full promotion comparison isolates the additional eligibility gate in this generator. Full promotion denotes this synthetic policy; Screened promotion denotes the human replay configuration.

Checkpoints follow promotion at the ends of sessions three and six, before any later feedback. Seven unobserved axes retain their initial mean, variance, and support throughout. All responses are choices with confidence one, so confidence and response-type conflict screening are inactive. Distinct image identifiers make image-count eligibility automatic once an axis has two-session support. The test measures stored scene summaries, with population fusion, neural ranking, and rendering excluded from its execution path.

The resulting mean stored-scene MSE values are reported in Table 5. Both checkpoints and all three regimes use the evaluation rule defined above.

E Qualitative Comparison

As shown in Figure 2, we provide qualitative examples to illustrate how CAPA balances persistent user preferences with explicit temporary instructions. Under broad editing requests, the same input can lead to profile-specific retouching results, whereas explicit temporary instructions override habitual preferences when a specific editing direction is requested.

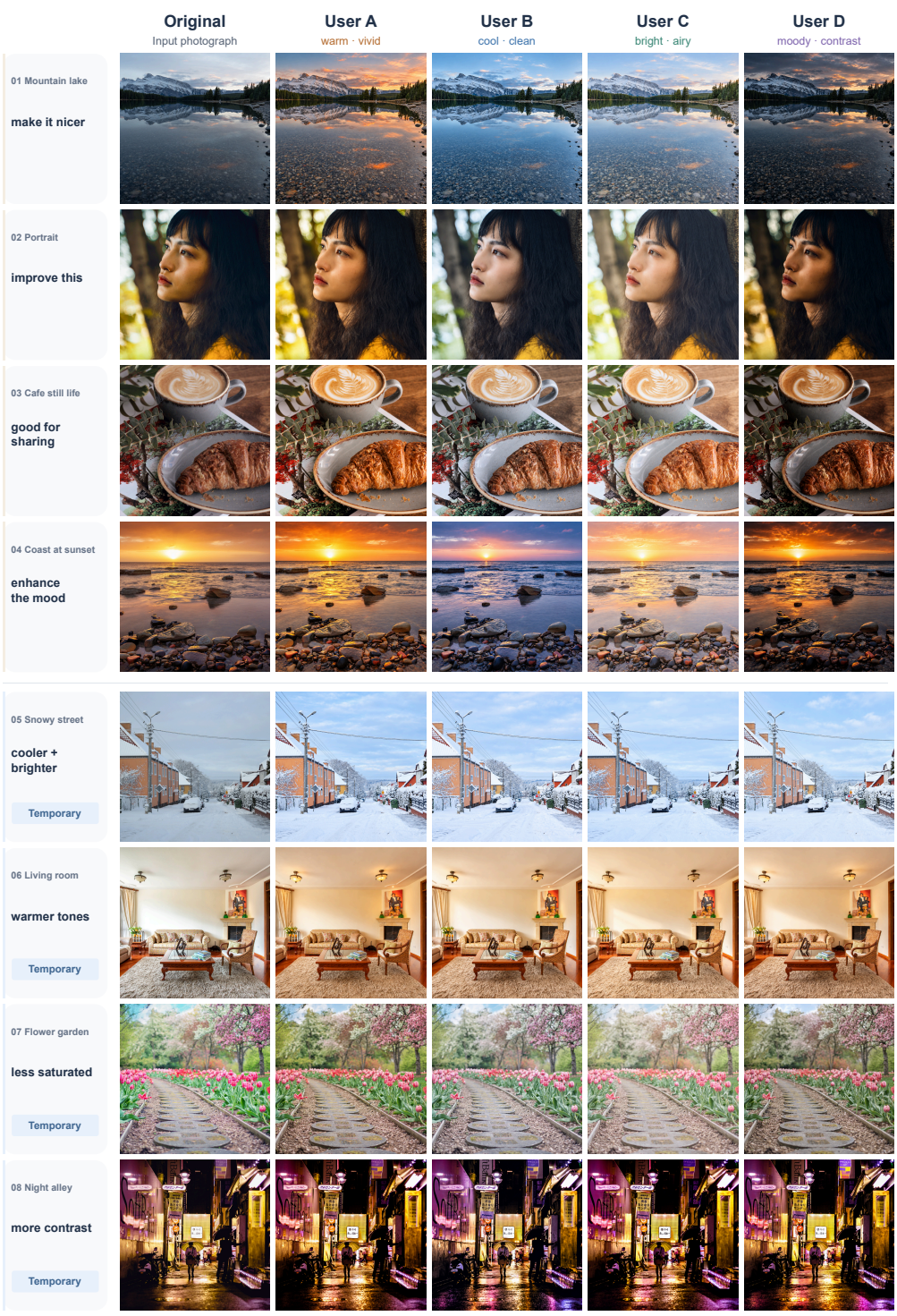


Figure 2: Qualitative comparison across eight input photographs and four user profiles. Columns show the original photograph and results for Users A–D, whose preferences are warm and vivid, cool and clean, bright and airy, and moody with strong contrast, respectively. The top four rows show profile-specific color and tone under broad editing requests. The bottom four rows show results under explicit temporary instructions: cooler and brighter, warmer tones, less saturated, and more contrast. Each row uses the same input photograph and instruction across profiles.