FIFTY REWARDS, ONE UPDATE: AUDITING OBJECTIVE CAPACITY BEYOND MECHANISM ACTIVATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-reward language-model optimizers are commonly described by the number of rewards they expose, but specified rules, aggregation tiers, emitted channels, empirical advantage rank, objective-gradient span, and capacity consumed by an optimizer are different quantities. We introduce an objective-capacity audit that traces these quantities from reward definitions to parameter updates. Pointwise within a shared rollout batch, every fixed linear combination of per-objective policy gradients is exactly equivalent to one scalar advantage. ObjCapBench shows that aligned controls collapse, structural panels activate PCGrad and resist held-out reconstruction by one batch-invariant weighting, and natural activation is rare and model-dependent. A fixed-rank intervention passes 31 of 35 mechanism and prediction gates, while a synthetic construction preserves nominal width 50, rank, mean, contrast energy, and the full singular spectrum yet changes PCGrad under an objective-basis rotation. Crucially, a separately sealed two-model, five-seed downstream test finds no positive advantage over the paired per-cell maximum of three registered static challengers: the estimate is -0.00083, with a 95% interval $[-0.00333, 0]$, and the preregistered primary gate fails closed. Nine of ten blocks lie at the endpoint's zero floor, limiting sensitivity but not changing the registered decision. Mechanism activation and non-reducibility therefore do not imply downstream benefit; no universal scaling law or algorithmic-superiority claim is supported.

INTRODUCTION

Language-model training systems increasingly expose multiple rewards: correctness, format, concision, evidence use, safety, style, and task-specific constraints. It is tempting to describe such systems by a single number, such as "50 rewards." That number can be misleading. Fifty implemented rules may be pooled into nine tiers, emitted as ten trainer channels, span only a few sample-level directions, and ultimately be consumed through one scalar advantage. Conversely, a smaller interface can remain operationally multi-objective when a data-dependent consumer uses conflict geometry that cannot be replaced by one batch-invariant weighting.

This distinction matters for both method claims and scaling claims. If nominal reward count, empirical rank, gradient span, and consumed capacity are conflated, an apparent benefit from "more rewards" can instead be caused by a changed weighting, a larger group, additional computation, or a different regularizer. The problem is not limited to one optimizer. It appears whenever an implementation crosses several contracts: reward source, aggregation graph, trainer emission, advantage construction, gradient formation, and gradient consumption.

We propose an auditable objective-capacity contract for this path. It reports reward specification count K_{spec}, aggregation-tier count K_{tier}, emitted-channel count K_{emit}, empirical advantage rank D_A, objective-gradient rank D_G, and operational consumer capacity K_{consume} separately. We accompany this contract with machine-readable receipts that bind source, tasks, model revisions, rollouts, rank thresholds, update hashes, and validation outcomes.

A key diagnostic lemma is linear collapse. Pointwise within one batch shared by the objectives, any fixed linear combination of per-objective policy gradients is exactly the gradient induced by a scalar advantage; the lemma does not assert that a nonlinear consumer and its linear controls follow

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.

Quantity	Meaning
K_{spec}	implemented reward rules
K_{tier}	aggregation groups or tiers
K_{emit}	channels exposed to the trainer
$D_A = \text{rank}(A)$	realized sample-level advantage rank
$D_G = \text{rank}(G)$	realized objective-gradient span
K_{consume}	operational capacity relative to a registered reconstruction hierarchy and tolerance; current experiments establish non-reducibility to one global fixed scalarization, not a universal integer dimension

Table 1: The six quantities are distinct. None should be reported under the single label "reward dimension."

the same cross-update trajectory. Thus $D_A > 1$ and $D_G > 1$ do not by themselves establish consumption beyond a fixed scalarization class. A nonlinear, data-dependent, objective-conditioned, or credit-map-specific consumer must be shown to activate, and its update must fail held-out reconstruction by one batch-invariant weighting. Existing gradient methods such as PCGrad, MGDA, and CAGrad provide consumers and baselines (Yu et al., 2020; Sener & Koltun, 2018; Liu et al., 2021); we do not claim these algorithms as ours.

We make five contributions:

- an end-to-end accounting contract from reward specification to parameter update, with a public auditor and fail-closed evidence schema;
- a linear-collapse diagnostic separating nominal multi-reward structure from consumption beyond one fixed scalarization class;
- ObjCapBench, which separates aligned, reward-tradeoff, structural-conflict, and natural panels under shared rollout banks; and
- a fixed-rank intervention plus a spectrum-preserving synthetic counterexample showing why rank-only or spectrum-only descriptions are insufficient for coordinate-dependent consumers, while retaining every failed predictive gate; and
- a fully completed, preregistered downstream test showing that controlled activation and non-reducibility did not translate into positive benefit over a strong static envelope in the two fixed model strata.

OBJECTIVE-CAPACITY CONTRACT

Let $r(x, y) \in \mathbb{R}^{K_{\text{spec}}}$ denote implemented rule outputs for prompt x and response y. A directed aggregation graph h maps rules into tiers and emitted trainer channels $q = h(r) \in \mathbb{R}^{K_{\text{emit}}}$. For a shared group of g completions, per-channel normalization constructs an advantage matrix $A \in \mathbb{R}^{g \times K}$, where K is the requested interface width. The optimizer then forms objective gradients $G = [g_1, \ldots, g_K]$ and a consumer maps G to one parameter update.

For group-centered advantages, $\mathbf{1}^\top A = 0$, so

$$D_A \le \min(g - 1, K), \qquad D_G \le \min(g - 1, P, K), \tag{1}$$

where P is the number of trainable parameters represented in each score vector. Equation 1 makes group size an explicit identifiability constraint. In particular, an interface of width ten cannot realize ten centered directions with fewer than eleven candidates.

The audit walks the complete contract:

$$\begin{aligned} &\texttt{source} \to \texttt{DAG} \to \texttt{emission} \\ &\qquad \to \texttt{advantage} \to \texttt{gradient} \to \texttt{consumer}. \end{aligned} \tag{2}$$

Each stage records its schema, dimensions, provenance, and hash. The public implementation operates on a synthetic 50-rule proxy and open-weight models. Public claims require no private artifacts; proprietary reward recipes, weights, thresholds, prompts, responses, checkpoints, and raw telemetry are neither required nor disclosed.

LINEAR COLLAPSE

For completion i, let $s_i = \nabla_\theta \log \pi_\theta(y_i \mid x_i) \in \mathbb{R}^P$. We use the loss-gradient convention, so the corresponding ascent direction is $-g_k$. The theorem assumes common sampled completions, one common token support and credit map for all objectives, and a policy-gradient operator linear in the supplied advantage. Objective-specific masks, blocks, truncation, or length normalizers that change support violate this domain. Under these assumptions, the loss gradient for objective k is

$$g_k = -\frac{1}{g} \sum_{i=1}^{g} A_{ik} s_i. \tag{3}$$

Lemma 1 (fixed-linear collapse). Assume all objectives use the same sampled completions, token support, credit map, score matrix S, and policy-gradient operator linear in the supplied advantage. For any weight vector $w \in \mathbb{R}^K$ held constant while forming the current update, with no differentiation through A or w,

$$\sum_{k=1}^{K} w_k g_k = -\frac{1}{g} \sum_{i=1}^{g} (Aw)_i s_i. \tag{4}$$

Thus separately constructing K objective gradients and combining them with w is exactly equivalent to backpropagating one scalar advantage $a = Aw$. The identity is a direct consequence of linearity; its role is diagnostic rather than a claim of a new gradient theorem.

Proof. Substitute the definition of g_k, interchange two finite sums, and collect $\sum_k w_k A_{ik} = (Aw)_i$. No simplex constraint on w is required. □

Corollary 1 (mass-preserving merge quotient). Fix $K = 10$. Let $B_1, \ldots, B_d$ partition the standardized base objectives, define $P_d \in \mathbb{R}^{10 \times d}$ by $(P_d)_{kj} = |B_j|^{-1}$ when $k \in B_j$ and zero otherwise, and let $m_d \in \mathbb{R}^d$ have $(m_d)_j = |B_j|/10$. Holding the base advantage matrix A fixed and defining reduced channels as AP_d, we have $P_d m_d = \mathbf{1}_{10}/10$ exactly, so $(AP_d)m_d = A\mathbf{1}_{10}/10$. Thus mass-preserving merging commutes with equal averaging and licenses a zero-GPU logical control. If width changes the base advantages or credit map, this identity does not apply. We define its finite-precision implementation canonically by the registered $d = 10$ trajectory; separately executing alternative float32 reduction orders is not claimed bitwise identical.

Writing $S \in \mathbb{R}^{g \times P}$ with row $s_i^\top$ gives $G = -S^\top A/g$. The general product bound gives $D_G = \operatorname{rank}(S^\top A) \le \operatorname{rank}(A) = D_A$, without requiring centering. Centering adds $\mathbf{1}^\top A = 0$, so the columns of A lie in a $(g-1)$-dimensional subspace and $D_A \le \min(g-1, K)$; also $\operatorname{rank}(S^\top) \le \min(g, P)$. This yields Equation 1. We define gradient conflict in the trainable LoRA parameter space by $g_j^\top g_k < 0$; advantage dependence, gradient dependence, and negative inner products are different properties. The validator enforces $D_G \le D_A$ as a deterministic implementation invariant, while we never assume equality or a full-column-rank advantage-to-gradient map. Operationally, K_{consume} is defined only relative to a frozen batch family and a registered reconstruction hierarchy: fixed scalar weights, batch-dependent scalar weights, gradient-geometry-dependent consumers, and objective-specific credit maps. Equal averaging belongs to the first class and uses one reward-side scalar per completion. A nonlinear consumer is not assigned a universal dimension from one batch; instead we test whether one global fixed weighting reconstructs its updates on held-out batch clusters. None of these statements claims that a parameter gradient has one nonzero coordinate or that the sample-normalization Jacobian has rank one.

Equation 4 continues to hold for every fixed linear combination inside this factorization class. What can fail for a nonlinear or batch-dependent consumer is reconstruction by one weight vector shared across held-out batches: its update may still lie in $\operatorname{span}(G_b)$ and admit a batch-specific w_b, while no single global w reconstructs all batches. Consumers that route objectives to different parameter subspaces or assign objective-specific token or trajectory credit fall outside the theorem's scope because they no longer share one A and S factorization. Blockwise Advantage Estimation is an example: objective-specific advantages are applied only to corresponding text blocks (Pavlenko et al., 2026). PCGrad remains inside the shared-gradient input setting but is a nonlinear consumer whose conflict projections depend on pairwise gradient geometry (Yu et al., 2020). Its activation alone does not establish downstream benefit.

Panel	Purpose
Aligned / zero conflict	Rank-one null; nonlinear consumer must reduce to linear control.
Reward tradeoff	Multiple verifier directions without forced parameter-gradient conflict.
Structural conflict	Controlled semantic bundles that force identifiable conflicting directions.
Natural	Frozen-model samples; measures existence and prevalence within frozen task/model/decoder strata, without adaptive search.

Table 2: Panels isolate different claims rather than pooling them into one score.

OBJCAPBENCH

ObjCapBench (Objective-Capacity Benchmark) is designed to prevent three substitutions: reward tradeoff for gradient conflict, nonzero rank for consumer activation, and update divergence for external benefit. Every task exposes 16 candidates and ten deterministic public verifiers. The group size exceeds the minimum $K + 1 = 11$ needed to realize ten centered directions, while leaving five extra samples for nondegenerate group structure under the frozen budget. Fifty public proxy rules are assigned to nine frozen contiguous tiers; the tiers generate 18 cross-DAG coordinates, which are then compressed to ten emitted channels. The proxy tests the declared $50 \to 9 \to 18 \to 10$ architecture without revealing a private reward recipe. Immutable early preregistration artifacts retain a legacy benchmark identifier; the presentation rename changed no task, threshold, seed, result, schema field, path, or sealed byte.

The three task families are quantitative multi-field answering, evidence-grounded question answering, and structured extraction. The verifier contract covers correctness, coverage, evidence, format validity, concision, consistency, field completeness, citation grounding, refusal calibration, and instruction adherence. Directional mutation tests require every verifier's targeted mutation to change its own score while leaving other axes below a fixed cross-sensitivity tolerance.

For each model, seed, panel, and task, scalar, equal-linear, and PCGrad consumers share an identical rollout bank and objective-gradient materialization. In v2.1, fixed-weight reducibility is tested out of sample: one global w^* is fitted on 20 blocked templates and evaluated on ten held-out templates. This avoids the vacuous per-batch observation that any update in $\mathrm{span}(G_b)$ has some post-hoc coefficient vector w_b.

CONFIRMATORY PROTOCOL AND CLAIM SEPARATION

The v2.1 protocol includes two model families, five seeds per model, three panels, 30 probe templates, and 16 responses per template: 900 groups and 14,400 shared-bank sequences. These are observations, not 900 or 14,400 independent statistical replicates; the top-level uncertainty units are fixed model strata crossed with seed and task blocks. Model revisions, task bytes, code bytes, decoding settings, numerical rank thresholds, and stopping gates were hash-bound before any v2.1 cell. Mechanism-fit and mechanism-evaluation files are physically distinct. Natural samples use temperature 0.8, top-p 0.95, and 64 new tokens, with no adaptive panel expansion. The sealed Qwen generation file resolves top-k to 20, whereas SmolLM2 resolves it to none; the natural gate is therefore a per-model, fixed-model–decoder existence test rather than an equal-decoder comparison.

The zero-control gate requires no conflict or projection and update-relative error at most 10^{-6}. Controlled activation requires positive fixed-weight reconstruction error in at least four of five seeds for both models and a blocked-bootstrap lower confidence bound above 0.02. Natural existence requires both the pooled and each-model prevalence lower bounds to exceed zero, with at least four of five positive seeds per model; "material" additionally requires every lower bound to exceed 0.10. A separately labeled post-hoc sensitivity analysis independently reweights crossed task and nested seed factors and multiplies the resulting observation weights, following multiway product reweighting (Owen & Eckles, 2012); it cannot alter a preregistered gate.

Downstream benefit is a separate estimand and is not inferred from activation or non-reducibility. We therefore preregistered a fresh outcome program with matched 60-update budgets, split-isolated tasks, a distinct post-update evaluator, and second-process full replay. Its primary G3 estimand

Result gate	Qwen	SmolLM2	Pooled estimate [95% CI]	Threshold	Status
Zero-control equivalence	5/5	5/5	10/10 equivalent	error $\leq 10^{-6}$	Pass
Controlled activation	5/5	5/5	0.0902 $[0.0760, 0.1026]$	LCB > 0.02	Pass
Natural activation	1/5 0.0067	3/5 0.0533	0.0300 $[0.0067, 0.0567]$	each model $\geq 4/5$, LCB > 0	Fail
Formal downstream static envelope	0/5	0/5	-0.00083 $[-0.00333, 0]$	effect ≥ 0.03, LCB > 0	Fail

Table 3: Sealed confirmatory results. Ratios are positive seeds. The downstream row compares PCGrad with the paired maximum over three frozen static challengers. Failed gates remain scoped negative results rather than being retuned.

is the PCGrad endpoint minus the paired per-cell maximum over three frozen static challengers. Forty physical conditions—two fixed model families, five seed clusters, and four arms per block—completed training, held-out evaluation, replay validation, and inclusion with no failed or terminal condition. The result and its gate decision were independently rebuilt and timestamp-sealed. G4 was hierarchically gated on a valid positive G3 decision; because G3 failed closed, G4 was not authorized or executed. G4 is therefore unobserved by design, not a positive or negative result.

Empirical Evidence

We then completed the sealed v2.1 confirmatory matrix: two model families, five seeds, and three panels yielded 30/30 hash-valid, second-process validator-checked cells. The zero-conflict control passed in every seed of both models. The structural panel activated the nonlinear consumer in every cell, and the held-out fixed-weight reconstruction-error increment was 0.0902 with a 10,000-replicate blocked-bootstrap 95% interval $[0.0760, 0.1026]$. Natural activation was observed but rare and model-dependent: pooled prevalence was 0.0300 $[0.0067, 0.0567]$, with only 1/5 positive Qwen seeds and 3/5 positive SmolLM2 seeds. Consequently the original preregistered natural-existence gate did not pass; its threshold and result are retained unchanged.

A separately labeled crossed-factor sensitivity synchronized task draws across all model–seed cells and paired seed multiplicities across the structural and natural panels. Across 10,000 raw draws, all fixed-weight replicates remained jointly full rank. The controlled error increment was 0.09023 $[0.06013, 0.11895]$, with positive intervals in both model strata. Natural prevalence was 0.0300 $[0.0033, 0.0633]$ pooled, but the Qwen interval remained $[0, 0.0400]$ while the SmolLM2 interval was $[0.0067, 0.1200]$. This post-hoc dependence sensitivity strengthens the controlled-mechanism readout but neither replaces nor reverses the original G2 failure.

CAB fixed-rank representation intervention. The separately preregistered CAB v1.1 holds exact algebraic rank and the mass-preserving equal-linear estimand invariant within the registered positive-α family, while varying contrast geometry. Let $M = \mathbf{1}\mathbf{1}^\top/10$ and $T_\alpha = M + \alpha(I - M)$. CAB transforms the base objective Gram by $Q_\alpha = T_\alpha^\top Q_0 T_\alpha$ before width projection and uses the same α for full and reduced consumers. T_α has eigenvalue 1 once and α nine times, so $\det(T_\alpha) = \alpha^9$ and it is invertible for every $\alpha > 0$; hence $\mathrm{rank}(Q_\alpha) = \mathrm{rank}(Q_0)$. The $\alpha = 0$ condition is a separate rank-one null and is not the predictive boundary. The registered holdout boundary is $\alpha_- = 0.50$ versus $\alpha_+ = 0.75$, both strictly positive. Its K-scaling comparisons use the same operational $\alpha = 1$ and congruent ten-objective basis. The sealed analysis receipt, reproduced byte-exactly by an author-side second-process schema replay, reports a held-out activation-prevalence jump of 0.5265 $[0.4667, 0.5833]$ across that boundary and a pooled held-out fixed-weight error increment of 0.0828 $[0.0670, 0.0976]$. The pooled point estimate equally weights the two fixed model strata; Qwen and SmolLM2 increments are 0.0383 and 0.1272, respectively, and uncertainty resamples the registered seed and archetype clusters.

CAB readout	Estimate [95% CI]	Registered requirement	Receipt
Activation jump	0.5265 $[0.4667, 0.5833]$	jump ≥ 0.25, LCB > 0	Pass
Fixed-weight error increment	0.0828 $[0.0670, 0.0976]$	LCB > 0.02, positive by family	Pass
Boundary vs. smooth prediction	0.0426 $[0.0358, 0.0492]$	improvement ≥ 0.10, LCB > 0	Fail: magnitude
Consumer–capacity interaction	0.0317 $[0.0251, 0.0377]$	positive by family, LCB > 0	Pass
Effective vs. nominal LOLO	0.0223 $[-0.0047, 0.0535]$	improvement ≥ 0.10, LCB > 0	Fail
Effective vs. matched-DF nominal	0.5395 $[0.3457, 0.6795]$	improvement > 0	Pass
Ridge effective vs. nominal	0.0654 $[0.0423, 0.0915]$	improvement ≥ 0.10	Fail: magnitude

Table 4: Sealed CAB v1.1 analysis output. The original composite adjudication is Fail: 31 of 35 gates pass and four remain false. Fixed-weight increments are 0.0383 for Qwen and 0.1272 for SmolLM2. The ridge sensitivity reverses direction for Qwen. An author-side second-process schema replay reproduces all tabled estimates and the unchanged 31/35 composite Fail adjudication byte-exactly; no downstream outcome claim is authorized by this table.

The CAB failure is localized rather than global. The intervention produces a sharp activation crossover, positive held-out non-reducibility, and a positive consumer–capacity interaction (0.0317 $[0.0251, 0.0377]$); the corresponding family slopes are positive but substantially different (0.00682 for Qwen and 0.05667 for SmolLM2). What does not survive the preregistered standard is model-agnostic predictive dominance: the frozen boundary predictor improves over its smooth comparator by only 4.26%, below the 10% gate, while the primary effective-versus-nominal LOLO interval includes zero. A matched-degree nominal comparator is substantially weaker, but the ridge comparison remains below threshold and reverses for Qwen. Thus CAB documents PCGrad-specific representation sensitivity and a positive registered interaction within the two fixed model strata, but not a universal capacity predictor or external scaling law. The original thresholds and composite failure remain unchanged.

Formal downstream outcome. The completed G3 program tests a harder claim than activation: whether PCGrad improves the final worst-verifier endpoint over the strongest paired member of a registered three-challenger static set. The primary estimate is -0.0008333, with a 10,000-replicate crossed blocked-bootstrap 95% percentile interval $[-0.0033333, 0]$. It misses the registered $+0.03$ minimum-effect gate, its 0.025 bootstrap quantile is not strictly positive, and both Qwen and SmolLM2 have 0/5 strictly positive seed clusters. Individual contrasts are exactly zero for the convex-midpoint and screen-rank-1 challengers; screen-rank-2 has estimate -0.0008333 and interval $[-0.0033333, 0]$. All 40 conditions completed independent 60-update replay and all four-stage condition chains were hash-verified; the null result is not an engineering failure.

The endpoint is nevertheless weakly informative away from the registered gate. In nine of ten model–seed blocks, both dynamic and static-envelope worst-verifier scores are zero. In one Qwen block, screen-rank-2 is 0.0083333 while PCGrad is zero. This zero-floor concentration can hide small improvements and limits effect-size resolution, but it cannot justify changing the endpoint, threshold, challenger set, or gate after observation. The correct conclusion is scoped: under these two checkpoints, five fixed seed clusters, current tasks, 60 updates, and this worst-verifier endpoint, activation and non-reducibility did not yield positive downstream benefit over the static envelope. It is not evidence that PCGrad is universally ineffective.

A spectrum-preserving counterexample. CAB preserves algebraic rank through an invertible, generally non-orthogonal congruence, so it can change the singular spectrum. The following construction asks the separate question of whether rank and the full spectrum suffice. Let objective gradients be rows of $G \in \mathbb{R}^{50\times 64}$ and $M = \mathbf{1}\mathbf{1}^\top/50$. We construct banks with a fixed mean direction and norm, fixed centered contrast energy, and $\mathrm{rank}((I-M)G) \in \{1,3,5,9,18,32,49\}$. Orthogonal $R \in \mathbb{R}^{50\times 50}$ acts on objective-index space by left multiplication and satisfies $R\mathbf{1} = \mathbf{1}$. Such R commutes with M, so the rotated centered bank is $R(I-M)G$; its rank is preserved because R is invertible, and the full singular spectrum because R is orthogonal. Writing $g_i = \mu + c_i$ with $\sum_i c_i = 0$ and $E_{\mathrm{contrast}} = \sum_i \|c_i\|^2$, the cross terms vanish on averaging and the ordered off-diagonal average is fixed exactly:

$$\overline{g_i^\top g_j} = \|\mu\|^2 - \frac{E_{\mathrm{contrast}}}{K(K-1)}. \tag{5}$$

Orthogonality plus $R\mathbf{1} = \mathbf{1}$ implies $\mathbf{1}^\top R = \mathbf{1}^\top$, so uniform fixed-linear aggregation $\mathbf{1}^\top G/50$ is exactly invariant; the observed maximum difference is 8.6×10^{-16}. PCGrad is not invariant: at rank one its update changes by 0.0519–0.0538 in ℓ_2, and at rank three by 0.0318–0.0417, while the same construction is numerically invariant at ranks five and above under the registered mean/energy setting. The rotated rank-one negative-pair fraction moves from 8.16×10^{-4} to 0.0310–0.0359, and projection count from 2 to 38–42. A separate $7 \times 9 \times 64 = 4{,}032$ row phase map verifies the registered invariants across 63 rank–contrast/mean cells and exhibits within-cell rotation sensitivity. This is a constructive synthetic counterexample to any consumer law based only on nominal width, rank, mean pairwise conflict, or singular spectrum. It is not language-model outcome evidence, and the apparent invariance of MGDA in this symmetric construction is not claimed generally.

Related Work and Ownership Boundary

Public research lineage. Three earlier public technical reports form the direct research lineage but do not contain the present claims. Zhang (2026a) reported a single financial-MoE case study of multi-reward interference, tiered normalization, and apparent phase-transition behavior. Zhang (2026c) documented the nine-tier, 50-sub-reward, ten-channel architecture, while Zhang (2026b) proposed a broader architecture-sensitive wisdom-scaling hypothesis. The present work neither transfers their empirical counts nor treats their hypotheses as evidence. It replaces that exploratory framing with source-to-consumer accounting, an exact fixed-linear boundary, controlled mechanism tests, a spectrum-preserving counterexample, and a completed two-model downstream test. We cite these reports in the third person and retain their dates only as provenance for the problem lineage.

Reward-interface neighbors and boundary. Recent work makes multi-reward interfaces more expressive through hierarchical holistic/aspect rewards, multi-objective gated heads, hybrid rule/model rewards, and dynamic dimension selection (Lai et al., 2024; Wang et al., 2024b; Gulhane & Indurthi, 2026; Chen et al., 2026c). MORLAIF scalarizes principle-specific rewards before PPO (Williams, 2024); these works define or consume reward interfaces, whereas our contribution is the source-to-consumer capacity audit and falsifiable reducibility test. Composite-RM constraints, preference mixtures, weight averaging, and conservative reward-model ensembles address proxy overoptimization or preference diversity (Moskovitz et al., 2024; Chakraborty et al., 2024; Rame et al., 2024; Coste et al., 2024). They are optimization or reward-model interventions, not evidence that nominal interface width equals consumed objective capacity; the distinction is central to our ownership boundary.

Direct-alignment reward overoptimization has a separate proxy-versus-gold scaling law (Rafailov et al., 2024); it is not a law for objective capacity consumed by a nonlinear policy-update consumer. Our Equation 5 counterexample also separates spectral summaries from coordinate-sensitive conflict topology: fixed spectral information can coexist with different PCGrad projections. Multi-objective optimization has long separated Pareto structure from simple scalarization. MGDA finds a common descent direction, PCGrad projects conflicting gradients, and CAGrad balances average improvement against conflict (Sener & Koltun, 2018; Yu et al., 2020; Liu et al., 2021). These methods are baselines and probes, not our inventions. Earlier work already learns steerable or Pareto multi-objective language policies and model interpolations (Rame et al., 2023; Yang et al., 2024; Wang et al., 2024a; Zhou et al., 2024; Wang et al., 2024c); reward-dimension reduction, graph-aware

rubric aggregation, and saturation-aware reweighting are likewise prior algorithmic contributions (Park & Sung, 2025; Lv et al., 2026; Huang et al., 2026). Recent language-model and generative-model post-training work also studies multi-reward GRPO normalization, reward grouping, geometric aggregation, decorrelation, block-specific credit, and multi-objective or multi-group policy optimization (Liu et al., 2026b;a; Montero et al., 2026; Gupta et al., 2026; Bai et al., 2026; Pavlenko et al., 2026; Lyu et al., 2025; Xiong & Singh, 2025). Reward-PRPO explicitly computes dimension-specific relative advantages and optimizes them in parallel (Tang et al., 2026); it is an optimizer design, whereas we audit how much of a declared interface survives aggregation and consumption. Gradient-geometry consumers include GAPO, RACO, MGDA-Decoupled, Pareto-Lenient Consensus, and bucket-local MOO (Li et al., 2025; Chen et al., 2026b; Vári-Kakas et al., 2026; Tan et al., 2026; Hoang et al., 2026); these are also external designs. RDPO normalizes heterogeneous rewards and whitens correlated active reward subspaces before scalar aggregation; our audit instead traces what dimensionality survives through the emitted advantage, parameter-gradient, and consumer contracts. Blockwise credit changes the token–objective credit map and therefore lies outside the shared-score-vector assumption of Lemma 1.

Strong scalarization results show why comparison only against equal weighting is insufficient (Kurin et al., 2022; Hu et al., 2023; Royer et al., 2023). Dynamic reward weighting and covariance-targeted adaptation directly optimize multi-objective alignment and cross-objective interference; APEX further identifies variance hijacking and gradient conflict under heterogeneous objectives (Lu et al., 2025; Lu & Jiang, 2026; Chen et al., 2026a). These algorithms and their covariance laws are prior work. Reward-consistency sampling instead mitigates data-side conflict, while sign-safe advantage composition constrains auxiliary rewards to preserve verifier direction (Xu et al., 2026; Seth, 2026). We ask a different upstream question: how many registered objective directions survive aggregation and are actually consumed, before separately testing benefit. Consistent with that separation, gradient conflict can track optimization harm without predicting generalization (Mueller et al., 2024).

Reward discovery, scalar signal strength, and objective capacity are also distinct. Search-driven reward design reports that downstream quality depends strongly on which rewards are selected rather than on ensemble count alone (Ahmadi et al., 2026); it optimizes reward specifications, whereas our audit holds the specification fixed and studies its induced interface. The Alignment Auditor infers latent reward functions from model behavior using Bayesian inverse RL, while Reward Auditor tests reward-model suitability under perturbations (Bou et al., 2026; Zang et al., 2025). We instead audit declared reward interfaces and realized optimizer updates without inferring hidden objectives or treating reward-model robustness as consumed capacity. Reward-model teaching effectiveness can depend on policy-induced reward variance even when reward accuracy is high (Razin et al., 2025); scalar variance is therefore an important activation precondition but does not identify advantage rank, objective-gradient span, or nonlinear consumption. Finally, concave expectation-level scalarization introduces plug-in gradient bias and associated sample-complexity effects (Ganesh & Aggarwal, 2026); that setting lies outside our pointwise fixed-linear-collapse lemma and reinforces why the consumer class must be declared explicitly.

GDPO shows that joint GRPO normalization can map distinct reward combinations to identical advantages, while AVSPO studies homogeneous within-group rewards that produce near-zero advantages (Liu et al., 2026b; He et al., 2026). These collapse objects differ from fixed-linear equivalence across objective-gradient channels. Mishra et al. instead derive a token-credit bottleneck for output-only GRPO and report low effective rank in a rollout-gradient matrix (Mishra et al., 2026); their objects of analysis are token-level credit and the rollout-gradient matrix under output-only GRPO. Ours separates reward-interface advantage rank from objective-gradient span and consumer capacity, tests fixed-weight reducibility, and reserves external outcomes as a separate estimand. RewardBench and RewardBench 2 benchmark reward models rather than this end-to-end capacity actually consumed by a policy optimizer (Lambert et al., 2024; Malik et al., 2026). Reward-model overoptimization scaling concerns proxy optimization versus gold reward (Gao et al., 2023), while diversified-preference scaling concerns model and data size and reward-model calibration (Zeng et al., 2024). We preregistered a conditional held-width comparison of nominal interface width and mechanism-measured consumed capacity at fixed public checkpoints. The required G3 gate failed closed, so the hierarchical G4 comparison was not authorized; no completed result supports a universal or first scaling law.

Our contribution is the source-to-consumer accounting contract, collapse and held-out reducibility diagnostics, fixed-rank intervention, spectrum-preserving counterexample, and preregistered nega-

tive downstream test. We claim neither ownership of the cited algorithms nor a universal scaling law.

Limitations, Disclosure, and Reproducibility

The two small checkpoints are fixed strata, not samples from a model population; uncertainty covers only seeds and task blocks within these families. Natural prevalence is specific to the frozen tasks, decoders, and checkpoints. CAB is model-heterogeneous, and PCGrad's scoped null need not transfer to other consumers. The worst-verifier endpoint is zero in 19 of 20 dynamic/static values, limiting resolution near the floor despite clearly missing the $+0.03$ gate. H50 is synthetic mechanism evidence, not language-model quality evidence. Rank also depends on tolerance; we report 10^{-6}, 10^{-4}, and 10^{-3} sensitivities and enforce $D_G \leq D_A$.

The anonymous review artifact contains public tasks/verifiers, sanitized rows, hashes/receipts, CAB/H50/G3 evidence, and recomputation code. It passes member-hash, double-blind, review-license, secret, and focused-replay checks. It excludes private reward logic, production prompts/responses, checkpoints, raw matrices, and telemetry; local readiness grants neither public redistribution nor external-upload status.

Reproducibility Statement

The appendix gives Lemma 1's assumptions and proof, audit construction, statistical units, all 35 CAB gates, G3 endpoint/challengers, zero-floor audit, and hash/RFC3161 bindings. The anonymous artifact contains authorized synthetic assets and deterministic checks. All preregistered gates are reported unchanged.

Ethics Statement

The experiments use public synthetic tasks and open-weight checkpoints, with no human subjects or private user data. The main risks are overclaiming optimizer efficacy and exposing proprietary reward logic; we mitigate them through scoped fixed-strata claims, complete negative-result reporting, and exclusion of commercial reward formulas and production artifacts.

AI Use Statement

Generative AI systems, including GLM and Codex, assisted literature discovery and triage; conceptual framing and hypothesis refinement; critique of mathematical claims and proofs; experimental-protocol design and red-teaming; code, test, and analysis-script generation and debugging; data formatting and result interpretation; and manuscript drafting and editing. They were not treated as independent validators. The authors manually checked source papers, derivations, executable code, data transformations, statistical outputs, and every reported claim, and assume full responsibility.

Conclusion

Objective capacity must be traced from aggregation and emission through advantage rank, gradient span, and the update consumer. Fixed linear consumers collapse exactly; beyond them, controlled activation and non-reducibility coexist with failed natural, predictive, and downstream gates. H50 also shows that rank and spectrum need not determine a coordinate-sensitive consumer. Under the registered models, tasks, budget, and endpoint, mechanism activation did not imply benefit over a strong static envelope. The contribution is therefore an auditable hierarchy of positive, negative, and unexecuted claims, not an optimizer-superiority claim or universal law.

REFERENCES

Arash Ahmadi, Sarah Sharif, and Yaser (Mike) Banad. Enhanced LLM reasoning by optimizing reward functions with search-driven reinforcement learning. *arXiv preprint arXiv:2605.02073*, 2026. doi: 10.48550/arXiv.2605.02073. URL https://arxiv.org/abs/2605.02073.

Yang Bai, Kaiyuan Liu, Ziyuan Zhuang, Jiahong Zhou, Rongxiang Weng, Xin Chen, Jingang Wang, and Xunliang Cai. Multi-objective and mixed-reward reinforcement learning via reward-decorrelated policy optimization. *arXiv preprint arXiv:2605.13641*, 2026.

Matthieu Bou, Nyal Patel, Arjun Jagota, Satyapriya Krishna, and Sonali Parbhoo. The alignment auditor: A bayesian framework for verifying and refining LLM objectives. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=CH7TfRLqSF.

Souradip Chakraborty, Jiahao Qiu, Hui Yuan, Alec Koppel, Dinesh Manocha, Furong Huang, Amrit Bedi, and Mengdi Wang. MaxMin-RLHF: Alignment with diverse human preferences. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 6116–6135. PMLR, 2024. URL https://proceedings.mlr.press/v235/chakraborty24b.html.

Dongliang Chen, Xinlin Zhuang, Junjie Xu, Luojian Xie, Zehui Wang, Jiaxi Zhuang, Haolin Yang, Liang Dou, Xiao He, Xingjiao Wu, and Ying Qian. APEX: Learning adaptive priorities for multi-objective alignment in vision-language generation. *arXiv preprint arXiv:2601.06574*, 2026a. doi: 10.48550/arXiv.2601.06574. URL https://arxiv.org/abs/2601.06574.

Peter Chen, Xiaopeng Li, Xi Chen, and Tianyi Lin. Reward-free alignment for conflicting objectives. *arXiv preprint arXiv:2602.02495*, 2026b. doi: 10.48550/arXiv.2602.02495. URL https://arxiv.org/abs/2602.02495. Accepted to ICML 2026.

Qiyuan Chen, Hongsen Huang, Jiahe Chen, Qian Shao, Jintai Chen, Hongxia Xu, Renjie Hua, Ren Chuan, and Jian Wu. Learning what matters: Dynamic dimension selection and aggregation for interpretable vision-language reward modeling. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 19779–19792. Association for Computational Linguistics, 2026c. doi: 10.18653/v1/2026.acl-long.906. URL https://aclanthology.org/2026.acl-long.906/.

Victor Chernozhukov, Sokbae Lee, and Adam M. Rosen. Intersection bounds: Estimation and inference. *Econometrica*, 81(2):667–737, 2013.

Thomas Coste, Usman Anwar, Robert Kirk, and David Krueger. Reward model ensembles help mitigate overoptimization. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/dda7f9378a210c25e470e19304cce85d-Abstract-Conference.html.

Swetha Ganesh and Vaneet Aggarwal. Breaking the bias barrier in concave multi-objective reinforcement learning. *arXiv preprint arXiv:2603.08518*, 2026. doi: 10.48550/arXiv.2603.08518. URL https://arxiv.org/abs/2603.08518.

Leo Gao, John Schulman, and Jacob Hilton. Scaling laws for reward model overoptimization. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 10835–10866. PMLR, 2023. URL https://proceedings.mlr.press/v202/gao23h.html.

Radha Gulhane and Sathish Reddy Indurthi. Beyond monolithic rewards: Hybrid multi-aspect reward optimization. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 26519–26533. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.findings-acl.1320. URL https://aclanthology.org/2026.findings-acl.1320/.

Taneesh Gupta, Pragya Srivastava, Rahul Madhavan, Karthikeyan Shanmugam, and Aravindan Raghuveer. GEOMA: Geometric and econometric objectives for multi-reward alignment. In *ICLR 2026 Workshop on Scalable Post-Training (SPOT)*, 2026. URL https://openreview.net/forum?id=0Adevc4M0L.

Xixiang He, Qiyao Sun, Ao Cheng, Xingming Li, Xuanyu Ji, Hailun Lu, Runke Huang, and Qingyong Hu. Advantage collapse in group relative policy optimization: Diagnosis and mitigation. *arXiv preprint arXiv:2605.21125*, 2026. doi: 10.48550/arXiv.2605.21125. URL https://arxiv.org/abs/2605.21125.

M. O. Hill. Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54(2): 427–432, 1973.

Long P. Hoang, Yiran Zhao, Wei Lu, and Wenxuan Zhang. Multilingual fine-tuning via localized gradient conflict resolution. *arXiv preprint arXiv:2606.05613*, 2026.

Yuzheng Hu, Ruicheng Xian, Qilong Wu, Qiuling Fan, Lang Yin, and Han Zhao. Revisiting scalarization in multi-task learning: A theoretical perspective. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Yu Huang, Zihua Zhao, Zhaoxin Huan, Wanli Gu, Feng Hong, Xinmu Ge, Lin Yuan, Weichang Wu, Qiang Hu, Xiaolu Zhang, Jun Zhou, and Jiangchao Yao. Focal reward: Balanced reinforcement learning under rubric-based rewards. *arXiv preprint arXiv:2605.26579*, 2026. doi: 10.48550/arXiv.2605.26579. URL https://arxiv.org/abs/2605.26579.

Vitaly Kurin, Alessandro De Palma, Ilya Kostrikov, Shimon Whiteson, and M. Pawan Kumar. In defense of the unitary scalarization for deep multi-task learning. In *Advances in Neural Information Processing Systems*, 2022.

Yuhang Lai, Siyuan Wang, Shujun Liu, Xuanjing Huang, and Zhongyu Wei. ALaRM: Align language models via hierarchical rewards modeling. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 7817–7831. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-acl.465. URL https://aclanthology.org/2024.findings-acl.465/.

Nathan Lambert, Valentina Pyatkin, Jacob Morrison, LJ Miranda, Bill Yuchen Lin, Khyathi Chandu, Nouha Dziri, Sachin Kumar, Tom Zick, Yejin Choi, Noah A. Smith, and Hannaneh Hajishirzi. Rewardbench: Evaluating reward models for language modeling. *arXiv preprint arXiv:2403.13787*, 2024.

Chengao Li, Hanyu Zhang, Yunkun Xu, Hongyan Xue, Xiang Ao, and Qing He. Gradient-adaptive policy optimization: Towards multi-objective alignment of large language models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 11214–11232. Association for Computational Linguistics, 2025. doi: 10.18653/v1/2025.acl-long.549. URL https://aclanthology.org/2025.acl-long.549/.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. In *Advances in Neural Information Processing Systems*, 2021.

Haotian Liu, Yihao Liu, Jingwei Ni, Siyuan Huang, Xinpeng Liu, Pengyu Cheng, Jiajun Song, Ruijin Ding, Junfeng Li, Zhechao Yu, Mengyu Zhou, Hongteng Xu, Xiaoxi Jiang, and Guanjun Jiang. GD^{2}PO: Mitigating multi-reward conflicts via group-dynamic reward-decoupled policy optimization. *arXiv preprint arXiv:2606.16771*, 2026a.

Shih-Yang Liu, Xin Dong, Ximing Lu, Shizhe Diao, Peter Belcak, Mingjie Liu, Min-Hung Chen, Hongxu Yin, Yu-Chiang Frank Wang, Kwang-Ting Cheng, Yejin Choi, Jan Kautz, and Pavlo Molchanov. GDPO: Group reward-decoupled normalization policy optimization for multi-reward RL optimization. *arXiv preprint arXiv:2601.05242*, 2026b.

Yining Lu and Meng Jiang. Uncovering cross-objective interference in multi-objective alignment. *arXiv preprint arXiv:2602.06869*, 2026.

Yining Lu, Zilong Wang, Shiyang Li, Xin Liu, Changlong Yu, Qingyu Yin, Zhan Shi, Zixuan Zhang, and Meng Jiang. Learning to optimize multi-objective alignment through dynamic reward weighting. *arXiv preprint arXiv:2509.11452*, 2025.

Can Lv, Mingju Chen, Heng Chang, and Shiji Zhou. Mitigating false credit propagation: Probabilistic graphical reward aggregation for rubric-based reinforcement learning. *arXiv preprint arXiv:2606.03361*, 2026. doi: 10.48550/arXiv.2606.03361. URL https://arxiv.org/abs/2606.03361.

Qiang Lyu, Zicong Chen, Chongxiao Wang, Haolin Shi, Shibo Gao, Ran Piao, Youwei Zeng, Jianlou Si, Fei Ding, Jing Li, Chun Pong Lau, and Weiqiang Wang. Multi-grpo: Multi-group advantage estimation for text-to-image generation with tree-based trajectories and multiple rewards. *arXiv preprint arXiv:2512.00743*, 2025.

Saumya Malik, Valentina Pyatkin, Sander Land, Jacob Morrison, Noah A. Smith, Hannaneh Hajishirzi, and Nathan Lambert. Rewardbench 2: Advancing reward model evaluation. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=fb0G86Dewb.

Amritansh Mishra, Supriyo Chakraborty, and Berkcan Kapusuzoglu. On the policy gradient foundations of group relative policy optimization: Credit assignment, gradient sparsity, and rank collapse. *arXiv preprint arXiv:2606.29238*, 2026.

Ivan Montero, Tomasz Jurczyk, and Bhuwan Dhingra. RVPO: Risk-sensitive alignment via variance regularization. *arXiv preprint arXiv:2605.05750*, 2026.

Ted Moskovitz, Aaditya Singh, DJ Strouse, Tuomas Sandholm, Ruslan Salakhutdinov, Anca Dragan, and Stephen McAleer. Confronting reward model overoptimization with constrained RLHF. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/5eee634cb9729b8bcc2ec9f2a46a74ae-Abstract-Conference.html.

David Mueller, Mark Dredze, and Nicholas Andrews. Can optimization trajectories explain multi-task transfer? *Transactions on Machine Learning Research*, 2024.

Art B. Owen and Dean Eckles. Bootstrapping data arrays of arbitrary order. *The Annals of Applied Statistics*, 6(3):895–927, 2012. doi: 10.1214/12-AOAS547. URL https://projecteuclid.org/euclid.aoas/1346418577.

Giseung Park and Youngchul Sung. Reward dimension reduction for scalable multi-objective reinforcement learning. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=ssRdQimeUI.

Kirill Pavlenko, Alexander Golubev, Simon Karasik, and Boris Yangel. Blockwise advantage estimation for multi-objective RL with verifiable rewards. *arXiv preprint arXiv:2602.10231*, 2026.

Hong Qian and Yang Yu. Solving high-dimensional multi-objective optimization problems with low effective dimensions. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

Rafael Rafailov, Yaswanth Chittepu, Ryan Park, Harshit Sikchi, Joey Hejna, W. Bradley Knox, Chelsea Finn, and Scott Niekum. Scaling laws for reward model overoptimization in direct alignment algorithms. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-4009. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/e45caa3d5273d105b8d045e748636957-Abstract-Conference.html.

Alexandre Rame, Guillaume Couairon, Corentin Dancette, Jean-Baptiste Gaya, Mustafa Shukor, Laure Soulier, and Matthieu Cord. Rewarded soups: Towards pareto-optimal alignment by interpolating weights fine-tuned on diverse rewards. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/e12a3b98b67e8395f639fde4c2b03168-Abstract-Conference.html.

Alexandre Rame, Nino Vieillard, Leonard Hussenot, Robert Dadashi-Tazehozi, Geoffrey Cideron, Olivier Bachem, and Johan Ferret. WARM: On the benefits of weight averaged reward models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 42048–42073. PMLR, 2024. URL https://proceedings.mlr.press/v235/rame24a.html.

Noam Razin, Zixuan Wang, Hubert Strauss, Stanley Wei, Jason D. Lee, and Sanjeev Arora. What makes a reward model a good teacher? an optimization perspective. In *Advances in Neural Information Processing Systems*, volume 38, pp. 59162–59222, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/554e056fe2b6d9fd27ffcd3367ae1267-Paper-Conference.pdf.

Alfréd Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pp. 547–561, 1961.

Olivier Roy and Martin Vetterli. The effective rank: A measure of effective dimensionality. In *15th European Signal Processing Conference*, pp. 606–610, 2007.

Amelie Royer, Tijmen Blankevoort, and Babak Ehteshami Bejnordi. Scalarization for multi-task and multi-domain learning at scale. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. In *Advances in Neural Information Processing Systems*, 2018.

Aryan Seth. Sign-safe advantage composition for multi-objective RL. Master's thesis, University of California, Los Angeles, 2026. URL https://escholarship.org/uc/item/666912nq.

Renxuan Tan, Rongpeng Li, Zhifeng Zhao, and Honggang Zhang. Beyond compromise: Pareto-lenient consensus for efficient multi-preference LLM alignment. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 37684–37705. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.findings-acl.1879. URL https://aclanthology.org/2026.findings-acl.1879/.

Jiajin Tang, Gaoyang, Wenjie Wang, Sibei Yang, and Xing Chen. Chart deep research in LVLMs via parallel relative policy optimization. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2603.06677.

Andor Vári-Kakas, Ji Won Park, and Natasa Tagasovska. MGDA-Decoupled: Geometry-aware multi-objective optimisation for DPO-based LLM alignment. *arXiv preprint arXiv:2604.20685*, 2026. doi: 10.48550/arXiv.2604.20685. URL https://arxiv.org/abs/2604.20685.

Haoxiang Wang, Yong Lin, Wei Xiong, Rui Yang, Shizhe Diao, Shuang Qiu, Han Zhao, and Tong Zhang. Arithmetic control of LLMs for diverse user preferences: Directional preference alignment with multi-objective rewards. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 8642–8655. Association for Computational Linguistics, 2024a. doi: 10.18653/v1/2024.acl-long.468. URL https://aclanthology.org/2024.acl-long.468/.

Haoxiang Wang, Wei Xiong, Tengyang Xie, Han Zhao, and Tong Zhang. Interpretable preferences via multi-objective reward modeling and mixture-of-experts. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 10582–10592. Association for Computational Linguistics, 2024b. doi: 10.18653/v1/2024.findings-emnlp.620. URL https://aclanthology.org/2024.findings-emnlp.620/.

Kaiwen Wang, Rahul Kidambi, Ryan Sullivan, Alekh Agarwal, Christoph Dann, Andrea Michi, Marco Gelmi, Yunxuan Li, Raghav Gupta, Kumar Avinava Dubey, Alexandre Rame, Johan Ferret, Geoffrey Cideron, Le Hou, Hongkun Yu, Amr Ahmed, Aranyak Mehta, Leonard Hussenot, Olivier Bachem, and Edouard Leurent. Conditional language policy: A general framework for steerable multi-objective finetuning. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 2153–2186. Association for Computational Linguistics, 2024c. doi: 10.18653/v1/2024.findings-emnlp.118. URL https://aclanthology.org/2024.findings-emnlp.118/.

Marcus Williams. Multi-objective reinforcement learning from ai feedback. *arXiv preprint arXiv:2406.07295*, 2024. doi: 10.48550/arXiv.2406.07295. URL https://arxiv.org/abs/2406.07295.

Nuoya Xiong and Aarti Singh. Projection optimization: A general framework for multi-objective and multi-group RLHF. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 69017–69058. PMLR, 2025. URL https://proceedings.mlr.press/v267/xiong25c.html.

Zhihao Xu, Yongqi Tong, Xin Zhang, Jun Zhou, and Xiting Wang. Understanding conflicts in multi-objective alignment through reward consistency. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 5450–5472. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.findings-acl.269. URL https://aclanthology.org/2026.findings-acl.269/.

Rui Yang, Xiaoman Pan, Feng Luo, Shuang Qiu, Han Zhong, Dong Yu, and Jianshu Chen. Rewards-in-context: Multi-objective alignment of foundation models with dynamic preference adjustment. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 56276–56297, 2024. URL https://proceedings.mlr.press/v235/yang24q.html.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. In *Advances in Neural Information Processing Systems*, 2020.

Jianxiang Zang, Yongda Wei, Ruxue Bai, Shiyu Jiang, Nijia Mo, Binhong Li, Qiang Sun, and Hui Liu. Reward auditor: Inference on reward modeling suitability in real-world perturbed scenarios. *arXiv preprint arXiv:2512.00920*, 2025. doi: 10.48550/arXiv.2512.00920. URL https://arxiv.org/abs/2512.00920.

Dun Zeng, Yong Dai, Pengyu Cheng, Longyue Wang, Tianhao Hu, Wanshun Chen, Nan Du, and Zenglin Xu. On diversified preferences of large language model alignment. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 9194–9210. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-emnlp.538. URL https://aclanthology.org/2024.findings-emnlp.538/.

Mian Zhang. When rewards collide: Structural interference and phase transitions in multi-objective grpo, April 2026a. URL https://doi.org/10.5281/zenodo.19665969.

Mian Zhang. The second scaling law: Intelligence scales with parameters, wisdom scales with architecture, April 2026b. URL https://doi.org/10.5281/zenodo.19895990.

Mian Zhang. Cognitive reward topology: A nine-tier architecture for measuring and shaping intelligence via multi-reward grpo, April 2026c. URL https://doi.org/10.5281/zenodo.19666829.

Zhanhui Zhou, Jie Liu, Jing Shao, Xiangyu Yue, Chao Yang, Wanli Ouyang, and Yu Qiao. Beyond one-preference-fits-all alignment: Multi-objective direct preference optimization. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 10586–10613. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-acl.630. URL https://aclanthology.org/2024.findings-acl.630/.

SUPPLEMENTARY MATERIAL

EVIDENCE STATUS AND SCOPE

This supplement reports the completed, integrity-valid v2.1 confirmatory probe matrix, the separately sealed CAB v1.1 fixed-rank intervention, and the completed formal G3 downstream test. It does not convert a failed sequential or composite gate into a success. The zero-conflict control (G0) and controlled-activation test (G1) passed; the natural-existence gate (G2) failed. CAB passes 31 of 35 gates but retains its composite Fail. The later formal G3 program completed all 40 physical conditions and fails its preregistered static-envelope benefit gate. G4 was gated on a valid positive G3 decision and was therefore not authorized or executed. Each positive, negative, and unexecuted claim remains separate.

The v2.1 result receipt records `paper_upload_authorized=false`; that field belongs to the original v2.1 scientific lineage and cannot retroactively authorize a broader claim. The anonymous review candidate has since passed local completeness, hash, extraction, double-blind, review-only license-scope, secret-scanning, and focused replay audits. This statement records local readiness only, not a public redistribution license, external upload, venue receipt, or server-side persistence.

OUTCOME PROGRAM: FROZEN DESIGN AND EXCLUSION BOUNDARY

Earlier successor attempts were closed before admissible outcome completion and contribute no scientific values to this manuscript. Before fresh formal outcome access, V1.4.2 froze a set-valued three-challenger static envelope, exact tasks, models, seeds, endpoints, thresholds, execution sources, and replay validator. The Linux execution amendment changed only the resilient execution topology and infrastructure-retry contract before formal outcome access; it retained the original estimand, challenger identities, 60-update budget, bootstrap, and direction gates and obtained an independent RFC3161 timestamp.

The fresh G3 stage then completed 40/40 training commits, 40/40 final-evaluation commits, 40/40 full-trajectory replay validations, and 40/40 inclusions, with zero infrastructure or terminal failures. Every condition is an independent state machine and every committed artifact is hash-bound. The 40-condition closure, independent cell rebuild, deterministic aggregate, independent aggregate validation, and result-gate decision are all closed; the formal manifest, Linux execution contract, aggregate contract, and result decision each have verified RFC3161 source-byte and query bindings. The completed negative result is therefore a scientific outcome, not an interrupted-run or engineering-failure label.

The sealed outcome design uses two model families, five outcome-seed clusters per family, split-isolated prompt-archetype clusters, and 30/15/30 train/development/final tasks over three balanced domains. Each final endpoint uses eight samples per task after 60 physical updates. Generation fixes group size 16, prompt length 224, at most 64 new tokens, microbatch size 4, temperature 0.8, top-p 0.95, and top-k $= 0$ in both families. Optimization fixes AdamW, learning rate 5×10^{-4}, zero weight decay, KL coefficient 0.01, gradient clipping at 1.0, and LoRA rank/alpha/dropout $8/16/0$. Every update has exactly ten policy VJPs and eleven total VJPs in every physical arm. All nuisance randomness uses the canonical $K = 10$ namespace and excludes the actual arm and intervention width; only the deterministic PCGrad projection order is keyed by actual K. Each trace binds `intervention_K`, `nuisance_rng_K=10`, and a recomputable `projection_order_sha256`. The training scorer and audit verifier are separate code paths. This separation tests direct scorer reuse; it is not an independent research implementation or independent-machine replication.

V1.4.2 retains the same outcome-blind 16-candidate construction but does not select a unique winner. Both models evaluate all candidates for 15 updates under each of two fresh screen seeds. Within each model, the registered cross-seed maximin–mean–ID ranking retains the first two candidates and adds their $0.5/0.5$ convex midpoint, which remains feasible because the capped simplex is convex. The formal comparator is the paired per-cell maximum over these three static challengers. The maximum is recomputed after every shared task-archetype and outcome-cluster resample; fixing the

point-estimate maximizer or choosing a favorable challenger is forbidden. This is a conservative dev-screened static ambiguity set, not a globally optimal static oracle.

The six-level parent design uses $K \in \{1, 2, 3, 5, 8, 10\}$; $K = 4, 6$ are neither executed nor interpolated. It contains 80 scientific physical conditions and 50 theorem-backed equal-linear aliases. Every alias receives a zero-GPU algebraic reconstruction. Eight physical shadows test $K = 1$ versus $K = 10$ equal-linear trajectories and unitary versus equal-linear implementations; they check task order, rollout/evaluation seeds, response hashes, scalar advantages, gradients, optimizer states, parameter updates, adapter states, and final evaluation step by step. Response hashes and discrete identities must agree exactly; scalar loss uses absolute tolerance 10^{-7}, while gradients, optimizer states, and parameter updates use full-vector relative ℓ_2 tolerance 10^{-6}, and the final score permits at most 0.005 absolute difference. Any failed check invalidates the implementation gate. Shadows enter no effect estimate. V1.4.2 leaves all parent K-level PCGrad, equal-linear, alias, and shadow identities unchanged; its extra static-challenger cells must be enumerated and sealed in a separate formal extension before execution. No development artifact can silently create or authorize a final-test condition.

G3-C, completed and fail-closed. The primary endpoint is the final-test worst-verifier score. With generator seed 34031001, the 10,000-replicate crossed bootstrap resamples CAB seed clusters and final prompt archetypes, retains both tasks within every sampled archetype, pairs every draw across all four $K = 10$ physical arms and both fixed model strata, and recomputes both the worst-verifier endpoint and the strongest static challenger. The primary V1.4.2 contrast is PCGrad minus that three-challenger envelope. The registered gates require a point improvement of at least 0.03, a positive two-sided 95% percentile lower bound, and positive direction in at least 4/5 seeds for each model family. The observed estimate is -0.0008333333333333333, with percentile interval $[-0.003333333333333333, 0]$; the lower bound is not strictly positive, and both model families have 0/5 strictly positive seeds. Convex-midpoint and screen-rank-1 contrasts are exactly zero. Screen-rank-2 has estimate -0.0008333333333333333 and interval $[-0.003333333333333333, 0]$. The conjunctive primary gate is therefore **fail**.

This comparator is more precisely a *blockwise switching-oracle upper envelope*, not one deployable static policy. If D_b is the dynamic score and S_{bj} is challenger j in paired block b, then

$$\Delta_{\text{env}} = \frac{1}{B} \sum_b \left(D_b - \max_j S_{bj} \right) \tag{6}$$

$$= \frac{1}{B} \sum_b \min_j (D_b - S_{bj}) \tag{7}$$

$$\leq \min_j \frac{1}{B} \sum_b (D_b - S_{bj}). \tag{8}$$

Thus the unchanged envelope gate is stronger than beating every single fixed challenger on average; it does not establish optimality against all static weights. Because five paired clusters induce only 126 distinct multinomial cluster-count vectors, 10,000 bootstrap draws reduce Monte Carlo error but do not create additional seed information. Nine of ten paired model–seed blocks have dynamic and static-envelope worst-verifier scores equal to zero. In the remaining Qwen block, screen-rank-2 scores 0.008333333333333333 while PCGrad scores zero. This endpoint floor limits sensitivity to small improvements but cannot change the registered gate. The non-gating leave-one-cluster, positive-Dirichlet, challenger-switch, and exhaustive multinomial audits are reported without rescue authority. Neighboring intersection-bound methods are cited only as statistical context (Chernozhukov et al., 2013).

G4-O, registered but not authorized. The outcome panel contains exactly $2 \times 5 \times 6 \times 30 = 1{,}800$ PCGrad rows at model-family by CAB-seed-cluster by K by final-task grain. Its geometry is supported by 180 distinct model–seed–K–domain blocks, not 1,800 independent geometry observations; each fold has 60 eligible training blocks per model after excluding the held seed and width. Thirty joint leave-K-and-seed folds predict every row once: the test cell has the held K and seed, while training excludes every row sharing either; model families are fit separately and continuous standardization is fit inside each fold. The scientific geometry vector is realized gradient rank, update-retention ratio, stable rank, entropy rank, and conflict density. Let $Q_{10} \in \mathbb{R}^{10 \times 10}$ denote

the same-α, dimensionless CAB objective Gram in its ten-objective coefficient basis; this is distinct from the parameter-by-objective gradient matrix G in Definition 1. For update retention within one CAB row,

$$\rho_K = 1 - \frac{(c_K - c_{10})^\top Q_{10}(c_K - c_{10})}{c_{10}^\top Q_{10} c_{10}}, \tag{9}$$

where c_{10} is the counterfactual full-width coefficient vector under the same projection order and candidate bank. Cross-physical-cell subtraction is forbidden. Each of the ten registered structural tasks in an exact model–seed–K–domain block must satisfy $(c_{10}^\top Q_{10} c_{10})/\operatorname{tr}(Q_{10}) \geq (10^{-4})^2$; because the selected Gram has positive trace, this makes the denominator strictly positive and bounded away from numerical zero. Only then is the arithmetic mean taken over all ten tasks. Any missing or ineligible task invalidates the entire block: there is no imputation, task deletion, or mean over a surviving subset.

For the other geometry features, let Q_K be the projected dimensionless $K \times K$ Gram, let $\lambda_1 \geq \cdots \geq \lambda_K \geq 0$ be its numerically clipped eigenvalues, set $s_i = \sqrt{\lambda_i}$, and write $\tau_K = \sum_i \lambda_i$. Realized rank is $\#\{i : s_i > \max(10^{-4} s_1, 10^{-12})\}$; stable rank is $\tau_K / \max(\lambda_1, 10^{-30})$; and entropy rank is $\exp(-\sum_{i:p_i>0} p_i \log p_i)$ with $p_i = \lambda_i / \max(\tau_K, 10^{-30})$ (Roy & Vetterli, 2007). More generally, the Rényi–Hill profile is $D_q = (\sum_i p_i^q)^{1/(1-q)}$, with D_0 equal to realized support rank, D_1 to entropy rank, $D_2 = (\sum_i p_i^2)^{-1}$ to participation ratio, and $D_\infty = 1/\max_i p_i$ to stable rank (Rényi, 1961; Hill, 1973). It obeys $D_0 \geq D_1 \geq D_2 \geq D_\infty$. The sealed G4-O predictor remains unchanged: D_2 is descriptive and non-gating and may enter only a separately labeled journal-extension audit. Conflict density is the fraction of unordered off-diagonal pairs (i, j) with $(Q_K)_{ij} < 0$ among $\binom{K}{2}$ pairs, defined as zero at $K = 1$. These values are averaged over all ten physically disjoint CAB structural tasks only after every source row in the block passes eligibility; otherwise the entire block fails closed. Baseline task score, prompt-token count, and completion-token budget are the only non-identity nuisance features. K, seed, outcome seed, task, domain, and model identities are forbidden scientific predictors. Seed/task/domain identity controls and a leave-K-only leakage diagnostic are executed but non-gating.

The registered G4-O outcome was $y_t = 8^{-1} \sum_{s=1}^{8} \min_v a_{tsv}$, whereas G3-C first averages each verifier across the fixed task/sample bank and then takes the minimum. These are distinct endpoint definitions and are not described as identical. The archived plan also bound a non-gating G3-aligned endpoint, $\min_v 30^{-1} \sum_t 8^{-1} \sum_s a_{tsv}$, and a paired nested sample-index bootstrap sensitivity when common-random-number hashes match across arms. G4 analysis and execution were not authorized after the G3 primary gate failed closed.

All primary fits use ridge with fixed $\lambda = 1$. Geometry is compared with the strongest pooled, per-model, per-K, or model-by-K nominal envelope of first-, second-, and fifth-degree polynomials in $\log K$ plus identical nuisance features. Envelope choice, fold standardization, and all models are recomputed for each resample. With seed 34041401, the primary procedure accepts exactly 10,000 nondegenerate crossed seed-cluster–task-archetype bootstrap draws; single-unique-seed draws, whose probability is 0.0016, are rejected, redrawn, and counted in the receipt. G4-O uses one-sided 95% lower bounds, defined as the 0.05 quantile; this is distinct from G3-C's two-sided 95% percentile interval and 0.025 lower quantile. The G4-O lower bound is explicitly a small-cluster approximation under exchangeability of the five pseudorandom clusters and 15 fixed archetypes, not exact coverage. A 2,000-draw positive-Dirichlet Bayesian analysis with seed 34041402 is non-gating. All five delete-one-seed refits must be positive. A leave-K-only identity warning at 0.05 relative RMSE and fixed ridge sensitivities at $\lambda \in \{0.1, 0.3, 1, 3, 10\}$ are descriptive only. Legacy fixed-ridge thresholds of 0.05 relative and 0.0025 absolute improvement are retained as subordinate diagnostics and cannot replace the primary 0.10/0.005 gate. No exact randomization, sign-flip, jackknife-confidence, or population-wide seed/model claim is made.

The conjunctive G4 gate would require at least 10% relative and 0.005 absolute RMSE improvement, nominal RMSE and target standard deviation each at least 0.01, positive lower bounds for pooled relative improvement, pooled absolute reduction, the minimum model-family improvement, and geometry-only improvement, plus positive point directions in both families. All five delete-one-seed refits would have to be positive; pooled improvements would have to be positive at $K = 1, 10$ and at least five of six levels, each model at least four of six levels with nonnegative endpoints. Because hierarchical authorization was false, these G4 gates have neither pass nor fail status.

FROZEN CONFIRMATORY DESIGN

The v2.1 matrix fixes two open-weight model families, five seeds per family, and three panels: zero conflict, structural conflict, and natural samples. Each physical cell contains 30 task templates with 16 responses per template. Across the full matrix this yields 30 physical cells, 900 task groups, and 14,400 shared-bank sequences. Scalar-direct, equal per-objective linear, and PCGrad consumers are evaluated on the same rollout bank within each group. Rollouts are not treated as independent top-level replicates.

- Models: Qwen2.5-0.5B-Instruct and SmolLM2-360M-Instruct.
- Qwen revision: `7ae557604adf67be50417f59c2c2f167def9a775`.
- SmolLM2 revision: `a10cc1512eabd3dde888204e902eca88bddb4951`.
- Seeds: 104729, 130363, 155921, 196613, and 262147.
- Panels: zero conflict, structural conflict, and natural.
- Matrix: $2 \times 5 \times 3 = 30$ physical cells; 900 groups; 14,400 sequences.
- Interface: 10 declared objective coordinates.

The frozen natural sampler uses temperature 0.8, top-p 0.95, and 64 new tokens. The sealed Qwen configuration resolves top-k to 20, whereas SmolLM2 resolves it to none. Thus G2 is a fixed model–decoder existence test, not an equal-decoder comparison.

The frozen optimization settings are LoRA rank 8, LoRA alpha 16, dropout 0, learning rate 5×10^{-4}, weight decay 0, KL coefficient 0.01, and gradient clipping at norm 1.0. Numerical-rank sensitivity is recorded at relative tolerances 10^{-6}, 10^{-4}, and 10^{-3}.

REVIEWER-VISIBLE PUBLIC PROXY CONTRACT

The public benchmark is a synthetic, auditable proxy; it is not a disclosure of a production reward recipe. Let $s_1, \dots, s_{10} \in [0,1]$ be public verifier scores for exactness, coverage, unit/type consistency, derivation, non-contradiction, JSON validity, field-contract validity, evidence grounding, abstention, and token efficiency. The frozen public expansion produces five rules per verifier,

$$\begin{aligned} q_\ell &= 0.96 + 0.01(\ell \bmod 3), \\ r_{5(j-1)+\ell+1} &= \mathrm{clip}(s_j q_\ell, 0, 1), \quad j = 1{:}10,\ \ell = 0{:}4. \end{aligned} \tag{10}$$

This deterministic expansion is a nominal-rule control: its 50 coordinates are not asserted to be 50 statistically or causally independent objectives.

The nine registered contiguous tier blocks are

$$\begin{aligned} &B_1 = 1{:}5,\ B_2 = 6{:}10,\ B_3 = 11{:}16,\ B_4 = 17{:}24, \\ &B_5 = 25{:}29,\ B_6 = 30{:}33,\ B_7 = 34{:}39, \\ &B_8 = 40{:}45,\ B_9 = 46{:}50, \end{aligned} \tag{11}$$

with $t_a = |B_a|^{-1} \sum_{i \in B_a} r_i$. The public `nine_tier_cross` DAG emits the 18-vector

$$z = (t_1, \dots, t_9, t_1t_2, t_2t_3, \dots, t_8t_9, t_9t_1). \tag{12}$$

The frozen compression uses NumPy's ordered `array_split` of these 18 coordinates into 10 contiguous buckets and averages within each bucket. Explicitly,

$$\begin{aligned} c = \big(&(t_1+t_2)/2, (t_3+t_4)/2, (t_5+t_6)/2, (t_7+t_8)/2, \\ &(t_9+t_1t_2)/2, (t_2t_3+t_3t_4)/2, \\ &(t_4t_5+t_5t_6)/2, (t_6t_7+t_7t_8)/2, t_8t_9, t_9t_1\big). \end{aligned} \tag{13}$$

Thus the reviewer-visible path is fully specified as $s \mapsto r \in \mathbb{R}^{50} \mapsto t \in \mathbb{R}^9 \mapsto z \in \mathbb{R}^{18} \mapsto c \in \mathbb{R}^{10}$. The separately distributed standalone synthetic generator has a simpler ten-channel diagnostic path; it is not substituted for this confirmatory aggregation path.

This is a deterministic representation chain, not evidence for 50 independent latent causes. Deterministic aggregation cannot create new source information, while the nonlinear $9 \to 18$ feature lift can change what is linearly accessible to a downstream audit. Accordingly, 18 derived coordinates are not counted as 18 new rewards; source count, representation width, empirical rank, and consumer activation remain separate ledger fields.

Production systems may use private formulas, thresholds, weights, routing rules, prompts, responses, checkpoints, and telemetry. None is required to reproduce the public proxy claims, and no claim in this paper identifies the public neutral tiers $B_1{:}B_9$ with proprietary semantic tiers. The disclosed object is the public proxy DAG and its evidence receipts, not a production reward implementation.

Formal Audit Objects, Claims, and Proofs

Definition 1 (source-to-consumer audit). For a shared rollout group of n samples and K emitted objectives, let $J \in \mathbb{R}^{n \times p}$ be the shared per-sample log-policy score Jacobian, with row $J_{i:} = \nabla_\theta \log \pi_\theta(y_i \mid x_i)^\top$ on the common token support and credit map. Let $A = [a_1, \ldots, a_K] \in \mathbb{R}^{n \times K}$ contain objective advantages. We use the loss-gradient convention throughout; the corresponding ascent direction is $-g_k$. Define

$$G = -\frac{1}{n} J^\top A = [g_1, \ldots, g_K] \in \mathbb{R}^{p \times K}. \tag{14}$$

The audit records nominal rule count, emitted width K, $D_A = \operatorname{rank}(A)$, $D_G = \operatorname{rank}(G)$, pairwise conflict and consumer-activation events, and held-out fixed-weight reconstruction error. Numerical rank is always accompanied by the relative tolerance and singular spectrum. "Consumed capacity" is consumer-specific in this paper: it refers to the rank/geometry coordinates that change the consumer output, not to nominal rule count alone. It is neither Shannon channel capacity nor VC dimension, and it is not a universal integer shared across consumers, batches, representations, or numerical tolerances.

Proposition 1 (fixed-linear collapse). Assume all objectives use the same rollout samples, token support, credit map, and score matrix J; their policy-gradient estimator is linear in the advantage; and $w \in \mathbb{R}^K$ is fixed with no differentiation through w or the advantage construction. Then

$$\sum_{k=1}^{K} w_k g_k = -\frac{1}{n} J^\top (Aw), \tag{15}$$

which is exactly the update obtained from the scalar advantage $a_{\mathrm{scalar}} = Aw$.

Proof. By the common-support assumption, $g_k = -n^{-1} J^\top a_k$. Linearity gives $\sum_k w_k g_k = -n^{-1} J^\top \sum_k w_k a_k = -n^{-1} J^\top Aw$. No approximation or distributional assumption is used. □

The proposition does not cover objective-specific rollout distributions, masks, token credit, nonlinear normalization before or after aggregation, adaptive or parameter-dependent weights, constrained/multi-gradient consumers, or differentiation through the reward/weight computation. These are explicit out-of-domain cases rather than hidden exceptions.

Proposition 2 (advantage-to-gradient rank bound). Under Definition 1, the first inequality below holds generally; the second additionally uses group centering:

$$D_G \le D_A \le \min(n-1, K) \tag{16}$$

when each advantage column is centered within the group.

Proof. Because multiplication by the nonzero scalar $-n^{-1}$ does not change rank and $\operatorname{rank}(XY) \le \min(\operatorname{rank}(X), \operatorname{rank}(Y))$, $G = -n^{-1} J^\top A$ gives $\operatorname{rank}(G) = \operatorname{rank}(J^\top A) \le \operatorname{rank}(A) = D_A$, regardless of centering. Equality $D_G = D_A$ holds exactly when $\ker(J^\top) \cap \operatorname{col}(A) = \{0\}$, i.e., $J^\top$ is injective on the advantage column space. If the advantages are centered, $A^\top \mathbf{1}_n = 0$, so every column of A lies in the $(n-1)$-dimensional subspace $\mathbf{1}_n^\perp$ and $D_A \le n - 1$; the column count also gives $D_A \le K$. □

Proposition 3 (mass-preserving merge invariance). Partition K source objectives into d nonempty blocks $C_1, \ldots, C_d$. Define $P \in \mathbb{R}^{K \times d}$ by $P_{kj} = |C_j|^{-1}$ for $k \in C_j$ and zero otherwise, and define reduced equal-linear weights $m_j = |C_j|/K$. Then the reduced equal-linear update is identical to the full equal-linear update:

$$(GP)m = G\frac{\mathbf{1}_K}{K}. \tag{17}$$

Proof. For every $k \in C_j$, $(Pm)_k = |C_j|^{-1}|C_j|/K = 1/K$. Therefore $Pm = \mathbf{1}_K/K$, and left-multiplication by G proves the identity. □

This proposition licenses lower-width equal-linear rows as theorem-backed aliases only when all ten base objectives use the same frozen standardization constants, no K-specific renormalization follows merging, and nuisance randomness uses the canonical $K = 10$ namespace. Under those implementation premises the scalar advantage, gradient, AdamW state, and parameters agree step by step. The aliases contain zero new rollouts, updates, VJPs, or independent samples. Nonlinear consumers must still be physically evaluated at each tested interface.

Proposition 4 (CAB positive-contrast rank invariance at $K = 10$). For the registered ten-objective CAB basis, let $M = \mathbf{1}_{10}\mathbf{1}_{10}^\top/10$ and $T_\alpha = M + \alpha(I - M)$. For a base objective Gram matrix Q_0, CAB uses

$$Q_\alpha = T_\alpha^\top Q_0 T_\alpha \tag{18}$$

before every width projection and applies the same α to the full and reduced consumers. For every $\alpha > 0$, $\operatorname{rank}(Q_\alpha) = \operatorname{rank}(Q_0)$.

Proof. M is the projector onto the mean direction in the registered $K = 10$ space. T_α has eigenvalue 1 on that direction and eigenvalue α on its nine-dimensional orthogonal complement. Hence T_α is invertible exactly when $\alpha > 0$. Rank is invariant under left and right multiplication by invertible matrices, proving the claim. At $\alpha = 0$, $T_0 = M$ has rank one, so the positive-α premise is necessary. □

The CAB receipt additionally verifies the forward congruence and inverse recovery numerically at a registered relative tolerance. That numerical certificate checks implementation of the theorem; it does not turn the author-side replay into independent replication.

Proposition 5 (CAB exact contrast-energy scaling). Let $C = I - M$. For any $m \in \operatorname{Im}(M)$ and $c \in \operatorname{Im}(C)$,

$$m^\top Q_\alpha m = m^\top Q_0 m, \tag{19}$$

$$c^\top Q_\alpha c = \alpha^2 c^\top Q_0 c, \tag{20}$$

$$m^\top Q_\alpha c = \alpha m^\top Q_0 c, \tag{21}$$

$$\operatorname{tr}(Q_\alpha) = \operatorname{tr}(Q_0 M) + \alpha^2 \operatorname{tr}(Q_0 C). \tag{22}$$

Proof. The orthogonal projectors satisfy $M^2 = M$, $C^2 = C$, $MC = CM = 0$, $Mm = m$, $Cm = 0$, $Mc = 0$, and $Cc = c$, so $T_\alpha m = m$ and $T_\alpha c = \alpha c$. Substitution into $Q_\alpha = T_\alpha^\top Q_0 T_\alpha$ gives the first three identities. Moreover, T_α is symmetric and $T_\alpha T_\alpha^\top = T_\alpha^2 = (M + \alpha C)^2 = M + \alpha^2 C$. Hence cyclicity gives $\operatorname{tr}(Q_\alpha) = \operatorname{tr}(Q_0 T_\alpha T_\alpha^\top) = \operatorname{tr}(Q_0 M) + \alpha^2 \operatorname{tr}(Q_0 C)$. □

This is an exact mechanism-level scaling identity: CAB holds rank fixed for every positive α while contrast energy changes quadratically. It is not, by itself, a law for language-model quality. The uncompleted G4-O design would have tested whether a mechanism-derived profile predicts held-width outcomes beyond nominal K; no such result exists.

Proposition 6 (objective-basis invariants and consumer sensitivity). For any invertible $T \in \mathbb{R}^{K \times K}$, define $A' = AT$ and $G' = GT$. Then $\operatorname{rank}(A') = \operatorname{rank}(A)$, $\operatorname{rank}(G') = \operatorname{rank}(G)$, and $\operatorname{col}(G') = \operatorname{col}(G)$. Moreover, every fixed-linear update Gw is reproduced as $G'(T^{-1}w)$. Nonlinear consumers that use coordinatewise objective geometry, including PCGrad, need not be invariant to this change of basis.

Proof. Right multiplication by an invertible matrix preserves rank. Since T is onto, $\mathrm{col}(GT) = \{GTw : w \in \mathbb{R}^K\} = \{Gv : v \in \mathbb{R}^K\} = \mathrm{col}(G)$, where $v = Tw$. Also $(GT)(T^{-1}w) = Gw$. The final statement is a boundary rather than an invariance claim: coordinatewise projections depend on the transformed objective Gram and generally do not commute with arbitrary T. □

Thus rank is a coordinate-free upper bound on linear reachability, while operational consumption is consumer- and representation-specific. CAB exploits this distinction deliberately; it does not treat PCGrad activation as another name for rank.

Corollary 2 (all-ones-preserving rotation counterexample). Let $G \in \mathbb{R}^{K\times P}$ contain objective gradients as rows and let R be orthogonal with $R\mathbf{1} = \mathbf{1}$. Then RG preserves the mean gradient, Frobenius contrast energy, rank, and full singular spectrum. Orthogonality and $R\mathbf{1} = \mathbf{1}$ imply $R^\top\mathbf{1} = \mathbf{1}$, hence $\mathbf{1}^\top R = \mathbf{1}^\top$ and uniform fixed-linear aggregation is invariant: $K^{-1}\mathbf{1}^\top RG = K^{-1}\mathbf{1}^\top G$. PCGrad need not be invariant because the signs and locations of pairwise dot products and its ordered coordinate projections can change.

An explicit $K = 3$, $P = 2$ instance is

$$G = \begin{bmatrix} -2 & -2 \\ -2 & -2 \\ -1 & 2 \end{bmatrix}, \qquad R = \frac{2}{3}\mathbf{1}\mathbf{1}^\top - I. \tag{23}$$

R is orthogonal and fixes $\mathbf{1}$. Both uniform updates equal $(-5/3, -2/3)$. Under the registered PCGrad seed/order rule, $\mathrm{PCGrad}(G) = (-2.1, -0.3)$ with three projections, whereas $\mathrm{PCGrad}(RG) = (-5/3, -2/3)$ with zero projections; their ℓ_2 difference is 0.5676. This finite construction proves non-invariance independently of the larger H50 grid.

The independently revalidated H50 construction fixes $K = 50$, parameter dimension 64, mean norm 1, contrast energy 25, ranks $1, 3, 5, 9, 18, 32, 49$, and three registered rotations. All 21 rows preserve the mean, singular spectrum, and average-pairwise-dot identity. Uniform fixed-linear differences are at most 8.6×10^{-16}, while PCGrad differs in all six rank-one/rank-three rows: 0.0519–0.0538 at rank one and 0.0318–0.0417 at rank three. A separate 4,032-row map covers 63 rank–contrast/mean cells and 64 rotations per cell. These are constructive synthetic mechanism results only; they authorize no language-model, GRPO-quality, or universal scaling-law claim.

For completeness, the unordered-pair dot sum satisfies

$$\sum_{i<j} g_i^\top g_j = \frac{1}{2}\left(\left\|\sum_i g_i\right\|^2 - \|G\|_F^2\right). \tag{24}$$

Left multiplication by orthogonal R preserves $\|G\|_F$, and $R\mathbf{1} = \mathbf{1}$ preserves $\sum_i g_i = \mathbf{1}^\top G$ because $\mathbf{1}^\top R = \mathbf{1}^\top$. Hence the average off-diagonal dot product is invariant as claimed.

Definition 2 (held-out fixed-weight reducibility). On cluster-isolated fit groups $\mathcal{F}$, fit one global vector

$$w^\star = \arg\min_w \sum_{b\in\mathcal{F}} \|G_b w - u_b\|_2^2, \tag{25}$$

then report relative reconstruction error only on held-out groups $\mathcal{H}$,

$$E_{\mathcal{H}} = \left(\frac{\sum_{b\in\mathcal{H}} \|G_b w^\star - u_b\|_2^2}{\sum_{b\in\mathcal{H}} \|u_b\|_2^2}\right)^{1/2}. \tag{26}$$

This relative ℓ_2 endpoint is eligible only when its held-out denominator is strictly positive. Batch-specific or evaluation-fitted weights are diagnostics and cannot satisfy this definition.

Equivalently, vertically stack $\mathcal{G} = [G_1; \ldots; G_B] \in \mathbb{R}^{BP\times K}$ and $\sqcap = [u_1; \ldots; u_B] \in \mathbb{R}^{BP}$. A shared exact fixed weight exists if and only if $\sqcap \in \mathrm{col}(\mathcal{G})$; when it exists, it is unique if and only if $\mathrm{rank}(\mathcal{G}) = K$. Otherwise, for any particular solution $w^\star$, the full solution set is the affine space $w^\star + \ker(\mathcal{G})$. The audit therefore reports the stacked singular spectrum, numerical rank/nullity, full-column-rank condition number when defined, fit residual, held-out residual, and every per-batch out-of-span residual. This prevents a low aggregate residual from hiding an ill-conditioned or batch-local reconstruction.

OBJECTIVE-CAPACITY AUDIT PROCEDURE

Inputs: hash-bound public reward definitions and DAG; shared rollout groups; consumer implementations; rank, activation, split, and decision thresholds fixed before confirmatory analysis.

1. Verify source, task, model-revision, configuration, and preregistration hashes; fail closed on any mismatch.
2. Evaluate all public rule coordinates on one shared response bank. Reject missing/duplicate identities and unexpected dimensions.
3. Apply the declared DAG exactly, retaining the rule, tier, emitted-channel, and consumer-interface widths as different ledger fields.
4. Construct centered per-objective advantages A under the declared normalization order; store the spectrum and D_A at every registered tolerance.
5. On identical token support, compute the shared score matrix J and $G = -n^{-1}J^\top A$; reject any cell with $D_G > D_A$ beyond numerical tolerance.
6. Execute the scalar-direct, fixed equal-linear, and nonlinear consumer arms on the same bank and budget. Record conflicts, projections, update coefficients, VJPs, and nondegeneracy checks.
7. Fit at most one global reconstruction vector on cluster-isolated fit groups. Score it on held-out groups only; never treat a per-group fit as evidence of reducibility.
8. For width interventions, bind full and reduced consumers to the same α, use mass-preserving theorem aliases only for exact equal-linear controls, and physically execute nonlinear rows.
9. Bootstrap registered estimands over physical model/seed/archetype units. Never resample derived α, width, or alias rows as independent observations.
10. Emit a machine-readable receipt containing every gate, including false gates, plus validator hashes and claim/upload authorization fields. No composite pass may be inferred from a subset of positive estimates.

DEVELOPMENT-ONLY HARNESS EVIDENCE

Before sealing v2.1, implementation and positive-control checks exercised the mechanism and validator but were excluded from confirmatory estimation. Qwen's aligned panel had $D_A = D_G = 1$, zero projections, and held-out fixed-weight error 5.0×10^{-8}. Its structural panel had $D_A = D_G = 3$, 30 projections, and error 0.0800. A SmolLM2 structural compatibility smoke likewise had rank three, 27 projections, and error 0.0839. The Qwen structural panel activated conflict in all ten groups and rejected one global fixed weighting, whereas its aligned control yielded identical linear and PCGrad updates. Six deliberate artifact mutations were rejected by the validator. These checks are provenance for the harness, not confirmatory effect estimates.

SEQUENTIAL GATE RESULTS

The intervals below are 95% blocked-bootstrap intervals with 10,000 replicates.

G0, zero control (pass). All 10 expected cells are complete. Every paired adapter is equal after the zero-conflict update, and the aggregate records no integrity violation.

G1, controlled activation (pass). Both models show the positive direction in 5/5 seeds. The activation fraction is 1.0 in every structural cell. At the preregistered primary relative tolerance 10^{-4}, all 300/300 unique structural groups have $D_A = D_G = 10$, while all 300/300 unique zero-conflict groups have $D_A = D_G = 1$. This statement is a deterministic recomputation over 1,200 hash-bound diagnostic rows and is authorized only as confirmatory mechanism evidence; it does not establish natural prevalence, algorithmic superiority, or outcome scaling. The fixed-weight reconstruction-error increment is 0.09023, with 95% CI $[0.07605, 0.10258]$.

G2, natural existence (fail). Pooled activation prevalence is 0.0300, with 95% CI $[0.00667, 0.05667]$. Qwen has 1/5 positive seeds and prevalence 0.00667, with CI $[0, 0.02667]$. SmolLM2 has 3/5 positive seeds and prevalence 0.05333, with CI $[0.00667, 0.10667]$. The requirement of at least 4/5 positive seeds in each model was not met, and Qwen's model-specific lower bound was not positive.

Formal G3-C (completed; primary gate FAIL_CLOSED). The fresh Linux formal stage completed 40/40 training, evaluation, full second-process replay validation, and inclusion records with zero failure or terminal markers. The primary estimate is PCGrad minus the paired per-cell maximum over the three registered static challengers: -0.0008333333333333333, with 95% percentile interval $[-0.00333333333333333, 0]$. The minimum-effect gate is $+0.03$, the bootstrap lower bound is not strictly positive, and strictly positive seed directions are Qwen 0/5 and SmolLM2 0/5. The individual convex-midpoint and screen-rank-1 contrasts are zero; screen-rank-2 has the same negative estimate and interval. The result is a formal downstream null/negative result, not missing data and not an engineering failure.

Formal G4-O (registered but not authorized or executed). The hierarchy requires a valid positive G3 decision before G4 opens. Because G3 failed closed, G4 was not authorized and has no result. This is an intentional gate outcome, not a G4 failure or pass, and no G4 computation is used to rescue G3.

CAB v1.1 fixed-rank intervention (composite fail; author-side byte-exact schema replay complete). CAB holds exact algebraic rank and the mass-preserving equal-linear estimand invariant while varying representation geometry. Its sealed receipt, reproduced byte-exactly by the author-side second-process schema replay, records an activation-prevalence jump of 0.52646 with 95% CI $[0.46667, 0.58333]$, and a pooled held-out fixed-weight error increment of 0.08277 $[0.06696, 0.09757]$. The Qwen and SmolLM2 increments are respectively 0.03830 and 0.12724. The consumer–capacity interaction is positive, 0.03171 $[0.02510, 0.03774]$, with positive but heterogeneous family slopes.

The failure is localized to stronger predictive-dominance requirements. Boundary-versus-smooth improvement is 0.04257 $[0.03577, 0.04916]$, below the registered 0.10 magnitude gate. Effective-versus-nominal leave-one-level-out improvement is 0.02226 $[-0.00467, 0.05352]$, so its interval includes zero and its magnitude is below 0.10. A matched-degree nominal comparator is substantially weaker, 0.53945 $[0.34571, 0.67955]$, but fixed-ridge improvement is 0.06542 $[0.04227, 0.09147]$, below threshold, and reverses sign for Qwen. Thus 31/35 gates pass and four remain false. The receipt documents PCGrad-specific representation sensitivity and a positive registered interaction within the two fixed model strata, not a universal capacity predictor, external outcome scaling, or a CAB pass. The author-side second-process schema replay reproduces all reported estimates and the unchanged composite Fail adjudication byte-exactly.

CAB Sample Ledger and All 35 Gates

CAB contains 10 physical gradient cells and 300 physical task groups: two fixed model families, five seeds per family, and 30 tasks per cell. Prompt archetypes are role-isolated into 18 fit and 12 held-out tasks over 15 archetype clusters. The 3,000 α rows and 4,800 width/consumer rows are deterministic derivatives of those physical groups and are not independent samples. Primary uncertainty uses 10,000 crossed, blocked bootstrap replicates with fixed model strata and resampling over physical seed clusters and prompt-archetype clusters. Derived width, α, and theorem-alias rows never become bootstrap replicates.

The discovery-only boundary selection fixed $\alpha_- = 0.50$, $\alpha_+ = 0.75$, and operational $\alpha = 1.00$ before holdout confirmation. No holdout boundary reselection occurred. The table below reproduces every registered Boolean; **T** and **F** mean true and false in the sealed receipt, not statistical significance in isolation.

B0: completeness and isolation.

Gate	Registered condition	Value
B0.1	Complete valid ten-cell holdout	**T**
B0.2	Prompt-archetype clusters role-isolated	**T**
B0.3	Fit/held-out task counts equal 18/12	**T**

B1: algebraic contrast contract.

Gate	Registered condition	Value
B1.1	$\alpha = 0$ exact no-activation control	**T**
B1.2	Positive-α exact rank-invariance bound	**T**
B1.3	Centered geometry has finite conflict onset	**T**

B2: operational activation and reducibility.

Gate	Registered condition	Value
B2.1	Operational reference updates nondegenerate	**T**
B2.2	PCGrad epsilon activation stable	**T**
B2.3	PCGrad epsilon projection count stable	**T**
B2.4	Positive direction in 4/5 seeds per model	**T**
B2.5	Activated fraction at least one half	**T**
B2.6	All fixed-weight point cells eligible	**T**
B2.7	Fixed-weight delta lower bound above 0.02	**T**

B3: boundary prediction.

Gate	Registered condition	Value
B3.1	Activation jump at least 0.25	**T**
B3.2	Activation-jump bootstrap lower bound positive	**T**
B3.3	Positive direction in 4/5 seeds per model	**T**
B3.4	Boundary prediction improvement at least 10%	**F**
B3.5	Boundary prediction bootstrap lower bound positive	**T**
B3.6	Boundary prediction positive in each model	**T**
B3.7	Rank/error crossing stable within one grid step	**T**

S0: scaling invariants.

Gate	Registered condition	Value
S0.1	Equal-linear mass invariance	**T**
S0.2	Nominal rule-replication invariance	**T**
S0.3	All scaling rows full-update eligible	**T**
S0.4	Same α explicitly bound	**T**
S0.5	Same-α congruence certified	**T**

S1–S2: realized levels and interaction.

Gate	Registered condition	Value
S1.1	At least six eligible width levels per model	**T**
S1.2	PCGrad active in 4/5 seeds per model	**T**
S2.1	Positive consumer interaction in each model	**T**
S2.2	Consumer-interaction bootstrap lower bound positive	**T**
S2.3	Positive order sensitivity in at least 80%	**T**

S3: predictive dominance.

Gate	Registered condition	Value
S3.1	Effective LOLO point improvement at least 10%	**F**
S3.2	Effective LOLO bootstrap lower bound positive	**F**
S3.3	No model-family sign reversal	**T**
S3.4	Effective beats matched-DF nominal baseline	**T**
S3.5	Fixed-ridge effective improvement at least 10%	**F**

Exactly B3.4, S3.1, S3.2, and S3.5 are false. The unchanged composite adjudication is therefore **fail** (31/35), even though the mechanism and interaction gates are positive.

G2's failure does not mean that natural activation was exactly zero: the frozen panel observed rare, model-dependent activation. It means that the stronger preregistered cross-model existence criterion was not satisfied. The authorized empirical statements are therefore limited to zero-control equivalence, controlled consumer activation, and rare model-dependent natural activation in this frozen panel.

Statistical Unit and Integrity Controls

The completed v2.1 matrix uses a 10,000-replicate blocked paired bootstrap over fixed model strata, seeds nested within model, benchmark domain, and template. Its evidence ledger forbids optional stopping, deletion of observed seeds or models, and adaptive panel expansion. This does not prohibit V1.4's preregistered five delete-one-seed analysis refits, which never remove source rows from the ledger. Missing v2.1 runs remain in its failure ledger; V1.4 formal records instead become `STARTED_NO_RETRY` before outcome-bound reads and cannot use an infrastructure rerun to replace the same formal identity. Reducing registered repetition counts, changing a registered seed or resampling kind, omitting a started identity, or executing a reduced rerun is labeled `DIAGNOSTIC_NON_GATING`; it cannot overwrite, rescue, or contribute to the formal adjudication.

All 30 expected cells are present and valid, with zero missing or unexpected cells and no aggregate integrity errors. The final receipt records 12 protocol and aggregation tests passed and zero failed. Model revisions, task bytes, source bytes, seed identities, and the aggregate are hash-bound.

Preregistration payload SHA-256:

```
f57a095637c570fa5d661f843c81a0d5094a3ac6691d62ea8c1c47f6b34d94b6
```

Completed aggregate SHA-256:

```
b8bbaf2bf7ea854108da0d66f5c715d7730e64787488766f4a6835de45b0a4ed
```

Contribution, Reproducibility, and Disclosure Boundary

External methods and concepts (not ours). PCGrad, MGDA, CAGrad, reward normalization/whitening, blockwise credit assignment, multi-objective preference optimization, dimension reduction, standard rank and congruence facts, blocked bootstrap inference, LoRA, and the Qwen/SmolLM checkpoints are external foundations or comparators, not our contributions.

The closest title-level terminology we found is Qian and Yu's low-effective-dimension multi-objective optimization (Qian & Yu, 2017). Their object is a low-dimensional decision/solution subspace enabling random embedding for derivative-free search and Pareto-front preservation. This work instead audits the realized capacity of a declared reward interface along reward definitions $\rightarrow$ advantages $\rightarrow$ policy gradients $\rightarrow$ a specified consumer. We do not claim their random-embedding method, objective-space theorem, or terminology as ours.

New in this paper. The claimed contribution is the objective-capacity accounting contract; fixed-linear collapse and mass-preserving aliases as falsifiable controls; the held-out one-global-weight reducibility audit; ObjCapBench with hash-bound shared-bank validation; CAB as a fixed-rank, same-α intervention with its 31/35 composite Fail retained; the H50 spectrum-preserving synthetic counterexample; and the formally completed G3 negative downstream result against a three-challenger

static envelope. G4-O remains registered but unauthorized. Internal artifact identifiers are omitted during review.

Public reproducibility versus private production assets. The core claims require no private reward recipe. The local anonymous reviewer artifact contains public task bytes and verifiers, merge matrices, model revisions, sanitized derived rows, source hashes, validation/resampling receipts, the complete 600-row rank-control replay surface, ten public CAB cell artifacts, H50 generators, and the formal G3 aggregate, validation, closure, gate, and timestamp-seal receipts. Credentials, local paths, commercial reward formulas, production thresholds, weights, production prompts and responses, raw public or production responses, adapters, checkpoints, raw production Jacobians/Gram matrices, and raw telemetry remain excluded. The extracted package passes its member-set/hash verifier and focused mechanism replays; the G3 component verifies the released aggregate and its hash-bound decision chain rather than reconstructing it from undisclosed response-level artifacts. This appendix establishes no OpenReview upload or server-side persistence; final uploaded bytes must still be server-downloaded and hash-verified before external submission is recorded as complete.