BATCH DIVERSITY UNDER DISTORTION CONSTRAINTS FOR LANGUAGE MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

A batch of responses to one prompt is more useful when the responses differ and each response stays close in distribution to the model. We introduce batch diversity, which quantifies the spread of responses within a batch, and distortion, which quantifies the deviation of the batch's average feature distribution from the model's. We characterize the maximum diversity at every distortion bound. Release selects a batch from a pool of independent responses under the bound, and Release diversity converges to the supremum on the support of the model's features as the pool grows and the optimization error vanishes. Diversity-Adaptive Release (DAR) chooses the number of additional responses from a pilot and keeps the bound. On two language models, Release has higher diversity than independent sampling at zero distortion, and DAR has lower generation cost than Release. Larger distortion bounds tend to raise utility and answer coverage and lower the constant-reward fraction, majority accuracy, and response accuracy.

1 INTRODUCTION

Language models often generate a batch of responses to the same prompt. We ask how batch diversity can be quantified and how far it can be increased. Diversity alone, however, does not quantify how much the output distribution differs from the model distribution. We therefore define measures of batch diversity and distortion and develop a theory for diversity maximization under distortion constraints. We propose Release to select diverse batches from a response pool under a prescribed distortion bound. To reduce generation cost, we propose Diversity-Adaptive Release (DAR), which adapts the pool size using a pilot sample.

In open-ended tasks, distinct responses provide users with useful alternatives (Zhang et al., 2025). In mathematical reasoning and code generation, multiple sampled solutions can improve success rates (Wang et al., 2023; Chen et al., 2021; Brown et al., 2024). Group relative policy optimization (GRPO) computes advantages from the rewards of responses to the same prompt (Shao et al., 2024). Batch diversity matters in all three settings, yet repeated sampling often produces similar responses (Zhang et al., 2025; Jiang et al., 2025; Banayeeanzade et al., 2026).

Several studies propose decoding or prompting methods to control response diversity. Nucleus sampling truncates the next-token distribution using a cumulative probability threshold (Holtzman et al., 2020), whereas min-p sampling sets its cutoff relative to the highest next-token probability (Nguyen et al., 2025). Diverse beam search penalizes similarity across beam groups (Vijayakumar et al., 2018). In-context regeneration requests a different answer after each response while retaining previous answers in the conversation (Zhang et al., 2025). Verbalized sampling prompts the model to generate multiple responses with probability estimates (Zhang et al., 2026).

Other methods generate dependent samples jointly. Arithmetic sampling generates dependent responses while preserving the model distribution at each output position (Vilnis et al., 2023). Stochastic beam search samples sequences without replacement by selecting the largest Gumbel-perturbed sequence log-probabilities (Kool et al., 2019).

Selection methods choose outputs from a pool using diversity, reward, or expected utility as criteria. MMR promotes diversity by balancing query relevance against similarity to previously selected items (Carbonell & Goldstein, 1998). k-DPPs consider diversity when assigning sampling probabilities to subsets of size k (Kulesza & Taskar, 2011). Minimum Bayes risk decoding selects transla-

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.

tions by maximizing estimated expected utility under the model (Eikema & Aziz, 2020). Speculative rejection (Sun et al., 2024) and TreeBoN (Qiu et al., 2025) reduce generation cost through early rejection and tree pruning, respectively.

Diversity and distributional alignment are also studied during training and beyond text. DQO (Chen et al., 2026b) optimizes a determinantal point process objective for quality and semantic diversity. Diversity-aware policy optimization (Yao et al., 2025) adds a token-level diversity objective to reinforcement learning for mathematical reasoning, and GAPO (Anschel et al., 2025) computes GRPO rewards over the whole group to encourage diverse completions. S2L-PO (Ren et al., 2026) diversifies GRPO groups with rollouts from smaller models of the same family. For diffusion models, particle guidance modifies joint diffusion dynamics using a repulsive potential in image synthesis and molecular conformer generation (Corso et al., 2024), and EDDY uses interacting diffusion dynamics that preserve each particle's marginal distribution (Vinograd et al., 2026). CADS (Sadat et al., 2024) anneals the conditioning signal during sampling to diversify images, and AIG (Jena et al., 2025) is an inference-time guidance that retains the diversity of the base model. Diverse samples are also used for uncertainty quantification in inverse problems (Chung et al., 2023; Wen et al., 2024). OTPrune (Chen et al., 2026a) prunes visual tokens by minimizing a 2-Wasserstein distance between the full and pruned token distributions. Our methods Release and DAR select responses from a fixed model and bound the distortion of the returned batch.

Our contributions are as follows. (i) We define distortion and diversity, determine the maximum diversity under each distortion bound, and derive the minimum distortion for a target diversity. For pools of independent model responses, we give selection constraints that control distortion and prove that the maximum pool diversity converges to the supremum on the model's feature support as the pool grows. (ii) We propose Release to approximately maximize expected batch diversity over distributions on subsets of a fixed-size pool under a distortion constraint. We also propose Diversity-Adaptive Release (DAR), which lowers generation cost by setting the number of further responses from a pilot sample and selects from all generated responses. We show that both methods control distortion and that the diversity of each method converges to the supremum on the model's feature support under the convergence conditions. (iii) On two language models, we examine how Release diversity changes with pool size, measure the generation cost of DAR against Release, and compare both methods with baselines on NoveltyBench, mathematical pass@B, and GRPO group statistics.

2 A Theory of Batch Diversity

For a fixed prompt, we first derive an upper bound on diversity in terms of distortion. We then determine the maximum diversity under each prescribed upper bound on distortion and the minimum distortion required for a given diversity. For pools of independent model responses, we give sufficient conditions for batch selection to satisfy a prescribed distortion bound. We also prove that the maximum pool diversity converges to the supremum under the feature support constraint at the same distortion bound. Finally, we relate diversity and distortion to existing metrics. Appendix A contains the proofs of Proposition 2.2, Theorem 2.3, Corollary 2.4, and Theorem 2.5.

Setting. Fix integers $d \geq 1$ and $B \geq 2$, and put $c_B = B/(B-1)$. For a Borel probability measure P on $\mathbb{R}^d$ with finite second moment, define

$$m_P = \int y \, dP(y), \qquad s_P = \left(\int \|y - m_P\|^2 \, dP(y) \right)^{1/2}.$$

Let Y be the random response of a language model to the fixed prompt, with values in a measurable space $(\mathcal{Y}, \mathcal{A})$ and let $\phi : \mathcal{Y} \to \mathbb{R}^d$ be measurable. Assume $P_0 = \mathcal{L}(\phi(Y))$ has finite second moment and $s_0 = s_{P_0} > 0$, and write $m_0 = m_{P_0}$. Here $\mathcal{L}$ denotes distribution, norms are Euclidean, and W_2 is the 2-Wasserstein distance (Villani, 2009).

All joint distributions Π considered below are Borel probability measures on $(\mathbb{R}^d)^B$ with finite second moment. For $X = (X_1, \ldots, X_B) \sim \Pi$, write $\Pi_i = \mathcal{L}(X_i)$, $\bar{X} = B^{-1} \sum_i X_i$, and $\bar{\Pi} = B^{-1} \sum_i \Pi_i$. Thus $\bar{\Pi} = \mathcal{L}(X_U)$ when U is independent of X and uniform on $\{1, \ldots, B\}$.

Definition 2.1. The *distortion* D and the *diversity* Div are

$$D(\Pi) = \frac{W_2(\bar{\Pi}, P_0)}{s_0}, \qquad \mathrm{Div}(\Pi) = \frac{\sum_{1 \leq i < j \leq B} \mathbb{E}_\Pi \|X_i - X_j\|^2}{s_0^2 B(B-1)}.$$

If $X_1, \dots, X_B$ are sampled independently from P_0, then $(D(\Pi), \mathrm{Div}(\Pi)) = (0, 1)$.

Define

$$M_B(P_0) = \inf_{\Pi_i = P_0,\ 1 \le i \le B} \mathbb{E}_\Pi \|\bar{X} - m_0\|^2.$$

Intuitively, $M_B(P_0)$ measures the irreducible fluctuation of the batch mean around m_0. The infimum is attained, as shown in the proof of Theorem 2.3 in Appendix A. Taking $\Pi = P_0^{\otimes B}$ gives $0 \le M_B(P_0) \le s_0^2/B$.

The following proposition gives an upper bound on batch diversity in terms of distortion.

Proposition 2.2. *Every joint distribution* Π *satisfies*

$$\mathrm{Div}(\Pi) \le c_B\big(1 + D(\Pi)\big)^2. \tag{1}$$

If $M_B(P_0) = 0$*, then for each* $\rho \ge 0$ *there is a joint distribution* Π *with* $D(\Pi) = \rho$ *and* $\mathrm{Div}(\Pi) = c_B(1+\rho)^2$.

By (1), $\mathrm{Div}(\Pi) > c_B$ implies $D(\Pi) > 0$.

We next determine $\sup_{D(\Pi) \le \rho} \mathrm{Div}(\Pi)$ without assuming $M_B(P_0) = 0$.

Theorem 2.3. *For every* $\rho \ge 0$*,*

$$\sup_{D(\Pi) \le \rho} \mathrm{Div}(\Pi) = c_B \left(\sqrt{1 - \frac{M_B(P_0)}{s_0^2}} + \rho \right)^2, \tag{2}$$

and the supremum is attained by a joint distribution Π *with* $D(\Pi) = \rho$.

If $M_B(P_0) > 0$, then

$$c_B \left(\sqrt{1 - \frac{M_B(P_0)}{s_0^2}} + \rho \right)^2 < c_B(1+\rho)^2 \qquad \text{for every } \rho \ge 0.$$

Thus the upper bound in Proposition 2.2 is not attainable when $M_B(P_0) > 0$.

Theorem 2.3 gives the maximum diversity under the constraint $D(\Pi) \le \rho$. Corollary 2.4 gives the minimum distortion required for a given diversity.

Corollary 2.4. *For every* $v \ge 0$*,*

$$\min_{\Pi:\, \mathrm{Div}(\Pi) \ge v} D(\Pi) = \max\left\{0,\ \sqrt{\frac{v}{c_B}} - \sqrt{1 - \frac{M_B(P_0)}{s_0^2}}\right\}, \tag{3}$$

and the minimum is attained.

Finite pools. We now consider selecting a batch from a finite pool of independent samples from P_0. Let $E = \operatorname{supp} P_0$. Since $P_0(E) = 1$, selection from a finite sample from P_0 gives a joint distribution Π satisfying $\Pi(E^B) = 1$. For $\rho \ge 0$, define

$$\mathrm{Div}^\star_E(\rho) = \sup\{\mathrm{Div}(\Pi) : D(\Pi) \le \rho,\ \Pi(E^B) = 1\}.$$

At $\rho = 0$, $\mathrm{Div}^\star_E(0)$ equals the maximum in (2). For $\rho > 0$, $\mathrm{Div}^\star_E(\rho)$ can be strictly smaller than the maximum in (2), because the supremum defining $\mathrm{Div}^\star_E(\rho)$ additionally requires $\Pi(E^B) = 1$.

For each integer $n \ge B$, put $\mathcal{S}_n = \{J \subseteq \{1, \dots, n\} : |J| = B\}$. For $N \ge B$, we call a list $x = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$ a pool. Set

$$P_x = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}, \qquad s_x = s_{P_x} = \left(\frac{1}{N}\sum_{i=1}^N \|x_i - m_{P_x}\|^2\right)^{1/2}.$$

Let $q : \mathcal{S}_N \to [0,1]$ satisfy $\sum_{J \in \mathcal{S}_N} q(J) = 1$. Define an $\mathcal{S}_N$-valued random variable S by $\Pr(S = J) = q(J)$. Conditional on S, let $(I_1, \dots, I_B)$ be a uniformly random ordering of the indices in S. The joint distribution of $(x_{I_1}, \dots, x_{I_B})$ is

$$\Pi_{x,q} = \sum_{J \in \mathcal{S}_N} \frac{q(J)}{B!} \sum_{\substack{(i_1,\dots,i_B) \\ \{i_1,\dots,i_B\} = J}} \delta_{(x_{i_1},\dots,x_{i_B})}.$$

Define

$$\mathrm{Div}_x(\Pi_{x,q}) = \begin{cases} \dfrac{\sum_{1\le i<j\le B} \mathbb{E}_{\Pi_{x,q}}\|X_i - X_j\|^2}{s_x^2 B(B-1)}, & s_x > 0, \\ 0, & s_x = 0. \end{cases}$$

For $\eta \ge 0$, set

$$\mathrm{Div}_x^\star(\eta) = \max_{q:\, W_2(\bar{\Pi}_{x,q}, P_x) \le \eta s_x} \mathrm{Div}_x(\Pi_{x,q}).$$

Let $Y_1, Y_2, \ldots$ be independent copies of Y, put $\xi_i = \phi(Y_i)$, and set $x_N = (\xi_1, \ldots, \xi_N)$ for each $N \ge B$. For each $x \in (\mathbb{R}^d)^N$, let q_x be a probability mass function on $\mathcal{S}_N$, with $x \mapsto q_x(J)$ Borel measurable for every $J \in \mathcal{S}_N$. Draw S with $\Pr(S = J \mid x_N) = q_{x_N}(J)$ for $J \in \mathcal{S}_N$, and order S uniformly conditional on (S, x_N) to obtain $(I_1, \ldots, I_B)$. Define

$$\Pi_N = \mathcal{L}(\xi_{I_1}, \ldots, \xi_{I_B}) = \mathbb{E}[\Pi_{x_N, q_{x_N}}]. \tag{4}$$

For $\rho \ge 0$, put

$$\eta_N(\rho) = \frac{\rho}{\sqrt{1 - 1/N}}. \tag{5}$$

Theorem 2.5. *For every $\rho \ge 0$,*

(a)

$$W_2(\bar{\Pi}_{x_N, q_{x_N}}, P_{x_N}) \le \eta_N(\rho) s_{x_N} \quad \textit{almost surely} \quad \textit{implies} \quad D(\Pi_N) \le \rho. \tag{6}$$

(b) As $N \to \infty$,

$$\mathrm{Div}_{x_N}^\star(\eta_N(\rho)) \longrightarrow \mathrm{Div}_E^\star(\rho) \qquad \textit{almost surely}. \tag{7}$$

Theorem 2.5(a) shows how to constrain batch selection within each observed pool to guarantee a prescribed distortion bound. Theorem 2.5(b) shows that the largest diversity attainable under the selection constraint converges to the supremum under the feature support constraint as the pool grows.

Diversity and existing metrics. $\mathrm{Div}(\Pi)$ is the expected average pairwise squared Euclidean distance under Π, normalized by the independent-sampling baseline $2s_0^2$.

For unit vectors, squared Euclidean distance is twice cosine distance. If P_0 and $\bar{\Pi}$ are supported on the unit sphere, $\mathrm{Div}(\Pi)$ is also the expected average pairwise cosine distance under Π, normalized by the independent-sampling baseline s_0^2. Appendix B.1 derives the relationships between $\mathrm{Div}(\Pi)$, squared Euclidean distance, and cosine distance.

Distortion and existing metrics. For class features with C classes, an answer in class j is represented by the standard basis vector $e_j \in \mathbb{R}^C$. If $P_0 = \sum_j p_j \delta_{e_j}$ and $\bar{\Pi} = \sum_j r_j \delta_{e_j}$, then $D(\Pi)^2 = 2\,\mathrm{TV}(p, r)/(1 - \sum_j p_j^2)$, where $\mathrm{TV}(p, r) = \frac{1}{2}\sum_j |p_j - r_j|$.

For correctness features, a correct answer is represented by 1 and an incorrect answer by 0. If $P_0 = (1-p)\delta_0 + p\delta_1$ with $0 < p < 1$ and $\bar{\Pi} = (1-r)\delta_0 + r\delta_1$, then $D(\Pi)^2 = |r - p|/[p(1-p)]$. Appendix B.2 relates $D(\Pi)$ to total variation distance between class distributions and to changes in the probability of correctness.

Appendix B.3 relates diversity to existing objectives at zero distortion.

3 METHOD

We keep the notation and assumptions of Section 2. We define Release to approximately maximize diversity under the finite-pool constraint, then extend the construction to Diversity-Adaptive Release (DAR), which uses a pilot sample to determine how many additional responses to generate. We show that both methods satisfy the prescribed distortion bound and approach $\mathrm{Div}_E^\star(\rho)$ as the sample size grows and the optimization error vanishes.

3.1 RELEASE

Fix $\rho \geq 0$. Release generates a pool of $N \geq B$ independent responses and returns a random batch of B responses from the pool. For a pool x of size N, Release computes a probability mass function $\hat{q}$ on $\mathcal{S}_N$ that satisfies the constraint and approximates the maximum of

$$\max_q \ \mathrm{Div}_x(\Pi_{x,q}) \quad \text{subject to} \quad W_2(\bar{\Pi}_{x,q}, P_x) \leq \eta_N(\rho) s_x, \tag{8}$$

where η_N is defined in (5). The maximum in (8) is $\mathrm{Div}_x^\star(\eta_N(\rho))$. Algorithm 1 gives the procedure, and Appendix C.1 gives the theoretical background and implementation details.

Algorithm 1 Release: parameters ρ, N

1: Generate N independent responses $Y_1, \ldots, Y_N$ and form $x_N = (\phi(Y_1), \ldots, \phi(Y_N))$.
2: Compute $\hat{q}$ that satisfies the constraint and approximates the maximum of (8) at $x = x_N$.
3: Draw S with $\Pr(S = J) = \hat{q}(J)$, order S uniformly as $(I_1, \ldots, I_B)$, and return $(Y_{I_1}, \ldots, Y_{I_B})$.

The features of the batch returned by Algorithm 1 have the distribution Π_N in (4). For the $\hat{q}$ computed in Algorithm 1, let $\varepsilon_N \geq 0$ bound the approximation gap $\mathrm{Div}_{x_N}^\star(\eta_N(\rho)) - \mathrm{Div}_{x_N}(\Pi_{x_N,\hat{q}})$ almost surely and uniformly over pools.

Proposition 3.1. *Release satisfies $D(\Pi_N) \leq \rho$. If $\varepsilon_N \to 0$, then $\mathrm{Div}(\Pi_N) \to \mathrm{Div}_E^\star(\rho)$ as $N \to \infty$.*

The inequality $D(\Pi_N) \leq \rho$ follows from Theorem 2.5(a) and holds whenever Release uses a feasible solution of (8), so optimization error lowers only diversity. Appendix C.2 gives the proof of Proposition 3.1.

Relation to survey sampling. Under zero distortion and equal inclusion probabilities for all pool indices, the batch feature mean is the Horvitz–Thompson estimator of the pool feature mean (Horvitz & Thompson, 1952). For a pool with positive feature variance, maximizing diversity is equivalent to minimizing the total variance of the batch feature mean. Balanced sampling attains the zero-distortion diversity bound when every selected batch has the same feature mean as the pool (Deville & Tillé, 2004). Appendix C.3 gives the derivation.

3.2 DIVERSITY-ADAPTIVE RELEASE

Fix integers $m \geq 1$ and $T \geq m + B$ for the pilot size and total response limit. Using calibration data independent of the evaluation responses, fix a measurable map $g_{T,\rho} : \mathcal{Y}^m \to \{B, \ldots, T - m\}$. Generate the pilot $Y_1, \ldots, Y_m$, set $N_T = g_{T,\rho}(Y_1, \ldots, Y_m)$, and generate N_T further independent responses. Put $x_{N_T}^+ = (\xi_{m+1}, \ldots, \xi_{m+N_T})$ and $x = x_{m+N_T}$, and define

$$R_T = \frac{m}{T} P_{x_m} + \left(1 - \frac{m}{T}\right) P_{x_{N_T}^+}. \tag{9}$$

The pilot mass m/T is fixed before observing the responses, so $\mathbb{E}R_T = P_0$ even though N_T depends on the pilot. DAR computes a probability mass function $\hat{q}$ on $\mathcal{S}_{m+N_T}$ that satisfies the constraint and approximates the maximum of

$$\max_q \ \mathrm{Div}_x(\Pi_{x,q}) \quad \text{subject to} \quad W_2(\bar{\Pi}_{x,q}, R_T) \leq \eta_{N_T}(\rho) s_{x_{N_T}^+}. \tag{10}$$

In (10), $s_{x_{N_T}^+}$ uses only the additional responses, which are independent given the pilot. All generated responses remain candidates. Algorithm 2 gives the procedure, and Appendix C.4 gives details.

Algorithm 2 Diversity-Adaptive Release (DAR): parameters $\rho, m, T, g_{T,\rho}$

1: Generate m independent responses $Y_1, \ldots, Y_m$ and set $N_T = g_{T,\rho}(Y_1, \ldots, Y_m)$.
2: Generate N_T further independent responses $Y_{m+1}, \ldots, Y_{m+N_T}$, and form x and R_T.
3: Compute $\hat{q}$ that satisfies the constraint and approximates the maximum of (10).
4: Draw S with $\Pr(S = J) = \hat{q}(J)$, order S uniformly as $(I_1, \ldots, I_B)$, and return $(Y_{I_1}, \ldots, Y_{I_B})$.

Let Π_T be the distribution of the features of the batch returned by Algorithm 2. For the $\hat{q}$ computed in Algorithm 2, let $\varepsilon_T \geq 0$ bound the difference between the maximum in (10) and $\mathrm{Div}_x(\Pi_{x,\hat{q}})$ almost surely and uniformly over pools.

Proposition 3.2. *DAR satisfies* $D(\Pi_T) \le \rho$. *For fixed* m, *if* $N_T \to \infty$ *almost surely and* $\varepsilon_T \to 0$, *then* $\mathrm{Div}(\Pi_T) \to \mathrm{Div}^\star_E(\rho)$ *as* $T \to \infty$.

DAR therefore retains the distortion bound while adapting the number of responses and approaches the same supremum as Release under the stated conditions. Appendix C.5 proves Proposition 3.2. Appendix C.6 shows that each response returned by Release or DAR at zero distortion has the distribution $\mathcal{L}(Y)$.

4 Experiments

We examine how the diversity of Release changes with pool size, measure the generation cost of DAR against Release, and compare Release and DAR with baselines on three downstream tasks. We use Llama-3.1-8B-Instruct (Grattafiori et al., 2024) and Qwen3-8B in thinking mode (Yang et al., 2025) and call them Llama and Qwen. Each model uses 100 calibration prompts and 100 held-out prompts per benchmark. All reported results use the held-out prompts. Appendix D gives the experimental details.

4.1 Diversity and distortion of Release

For open-ended responses with bge-base features (Xiao et al., 2024), we solve (8) by column generation at $B = 4$, $16 \le N \le 256$, and $\rho \in \{0, 1, 2\}$. Figure 1 shows the mean diversity of Release against N for each ρ over 100 Llama prompts. Table 1 compares Release with $N = 64$ and 128 and independent sampling at $\rho = 0$ over the same 100 prompts for Llama and Qwen.

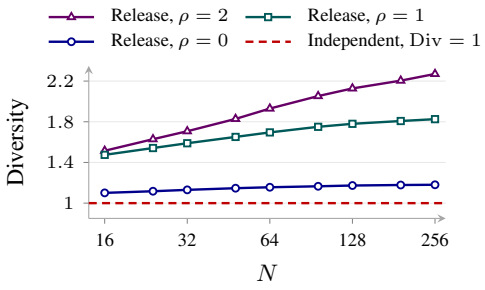

Figure 1: Computed diversity of Release for each ρ, with one point per pool size N, and analytic diversity of independent sampling, which is 1.

Model	Method	$\widehat{\mathrm{Div}}$	$\widehat{D}$
Llama	Independent	1.007±0.017	1.141±0.010
	Release, $N = 64$	1.156±0.007	1.123±0.009
	Release, $N = 128$	**1.173**±0.005	**1.116**±0.011
Qwen	Independent	0.994±0.014	1.164±0.010
	Release, $N = 64$	1.140±0.008	1.143±0.011
	Release, $N = 128$	**1.158**±0.004	**1.140**±0.011

Table 1: Computed diversity and distortion of independent sampling and Release at $\rho = 0$. Entries are estimates ± SE across prompts. Bold marks the highest diversity and the lowest distortion.

At $\rho = 0$, mean diversity rises from 1.1009 at $N = 16$ to 1.1805 at $N = 256$, which exceeds independent sampling by 18%. From $N = 192$ to 256, diversity increases by 0.0022 at $\rho = 0$, 0.0183 at $\rho = 1$, and 0.0658 at $\rho = 2$. Saturation in N occurs later under larger distortion bounds. Appendix E.1 shows a similar pattern for Release diversity on Qwen. In Table 1, Release with $N = 128$ has higher diversity and lower estimated distortion than independent sampling. Estimated distortion is positive even for independent sampling because 16 outputs per prompt are compared with a separate set of 128 responses. Distortion estimates are therefore read relative to the independent row.

4.2 Generation cost of DAR

We compare Release with $N = 64$ and 128 and DAR with $T = 64$ and 128 on open-ended tasks. All settings use bge-base features, $B = 4$, and $\rho = 0$. Cost includes all thinking and answer tokens. Table 2 reports token counts, savings, diversity, and distortion.

At $T = 64$, DAR reduces mean generation cost by 61.9% for both models relative to Release at $N = 64$. Mean diversity decreases from 1.156 to 1.117 on Llama and from 1.140 to 1.103 on Qwen. At $T = 128$, savings relative to $N = 128$ are 61.2% on Llama and 80.4% on Qwen. Appendix E.2 reports additional results on the generation cost of DAR.

Model	Method	Tokens, 10^3	Saving, % vs. $N = 64$	Saving, % vs. $N = 128$	$\widehat{\text{Div}}$	$\widehat{D}$
Llama	Release, $N = 64$	50.9±5.3	0.0±0.0	50.4±0.4	1.156±0.007	1.123±0.009
	Release, $N = 128$	102.6±10.5	−101.7±1.6	0.0±0.0	1.173±0.005	1.116±0.011
	DAR, $T = 64$	**19.4**±2.0	**61.9**±0.9	**81.1**±0.5	1.117±0.012	1.138±0.010
	DAR, $T = 128$	39.8±4.9	21.8±5.3	61.2±2.6	1.148±0.009	1.129±0.010
Qwen	Release, $N = 64$	85.5±4.0	0.0±0.0	50.4±0.4	1.140±0.008	1.143±0.011
	Release, $N = 128$	172.6±8.3	−101.8±1.8	0.0±0.0	1.158±0.004	1.140±0.011
	DAR, $T = 64$	**32.6**±1.5	**61.9**±0.5	**81.1**±0.3	1.103±0.011	1.174±0.009
	DAR, $T = 128$	33.8±1.5	60.5±0.9	80.4±0.5	1.097±0.011	1.185±0.009

Table 2: Results on 100 open-ended prompts per model. Entries are estimates ± SE across prompts. Savings compare mean token counts. Bold marks the lowest token count and the largest savings.

4.3 Downstream evaluation

Baselines. We compare Release with $N = 64$ and DAR with $T = 64$ against independent sampling and the following six baselines. In-context regeneration (Zhang et al., 2025) and verbalized sampling (Zhang et al., 2026) diversify responses through prompting. Arithmetic sampling (Vilnis et al., 2023) preserves output marginals, and stochastic beam search (Kool et al., 2019) avoids repeated sequences. MMR (Carbonell & Goldstein, 1998) and k-DPP (Kulesza & Taskar, 2011) compare alternative selection objectives on the same $N = 64$ pools as Release. Release, MMR, and k-DPP therefore have the same generation cost. The prompting and decoding baselines generate no candidate pool. For Qwen, Arithmetic sampling and stochastic beam search change decoding only in the answer that follows the thinking block. Appendix D.6 describes each baseline in detail.

Open-ended generation. On open-ended prompts, Release and DAR use one-hot indicators of classes formed by the NoveltyBench DeBERTa classifier (Zhang et al., 2025). We score each returned batch with the NoveltyBench distinct and utility metrics. Figure 2 shows distinct and utility for every method. Appendix D.7 describes the NoveltyBench classifier and metrics in detail.

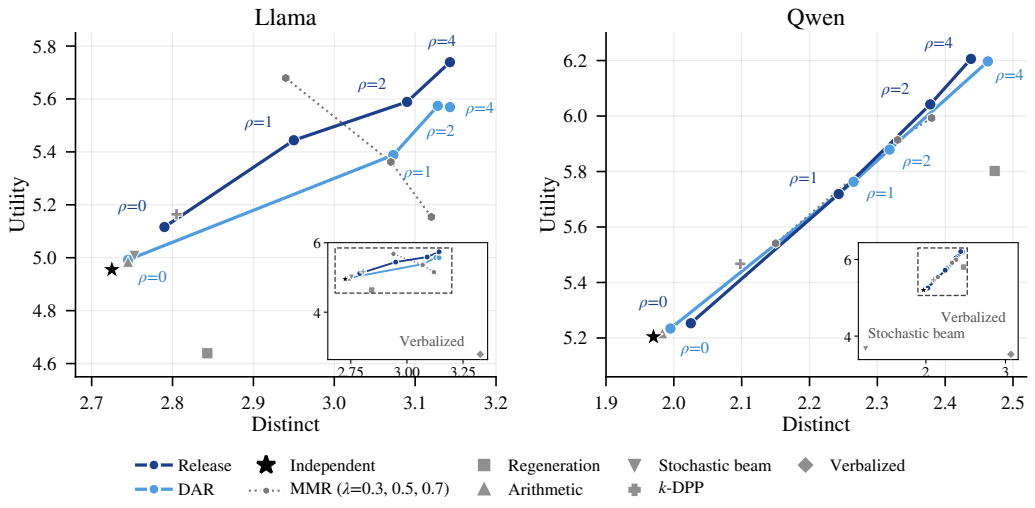

Figure 2: NoveltyBench distinct and utility. Higher values are better on both axes. Release and DAR are connected in order of $\rho \in \{0, 1, 2, 4\}$, and MMR is connected in order of $\lambda \in \{0.3, 0.5, 0.7\}$. Each inset shows every method, and the dashed box marks the range of the main panel. Table 22 in Appendix E.3 lists the means and standard errors of distinct and utility across prompts.

For Release, increasing ρ from 0 to 2 raises mean distinct from 2.790 to 3.090 on Llama and from 2.025 to 2.378 on Qwen. Utility rises from 5.116 to 5.589 and from 5.253 to 6.042, respectively.

At $\rho = 2$, DAR attains utilities 5.574 and 5.879. Release utility continues to rise beyond $\rho = 2$ and reaches 5.739 on Llama and 6.206 on Qwen at $\rho = 4$. Because Release selects only among pooled responses, the pool bounds the gain from a larger ρ, and a larger pool raises the bound. Verbalized sampling has the highest distinct but lower utility than independent sampling. Appendix E.4 reports bge-base diversity and distortion for all methods, compares class and bge-base features for Release, and reports Release and DAR through $\rho = 5$. Appendix E.5 reports MMR with $\lambda \in \{0.1, 0.9\}$. Appendix E.6 reports temperature, top-p, and min-p settings on Llama.

Mathematical reasoning. On GSM8K and MATH-500 problems (Cobbe et al., 2021; Hendrycks et al., 2021; Lightman et al., 2024), we report pass@8 and majority@8 in Table 3. Pass@8 is the fraction of problems with a correct response in the batch, and majority@8 is the fraction whose most frequent final answer is correct. Release and DAR use the final answer of each response as the feature and do not use which answer is correct. Independent sampling draws uniformly without replacement from the same pools.

	Llama				Qwen			
	GSM8K		MATH-500		GSM8K		MATH-500	
Method	Pass@8	Maj@8	Pass@8	Maj@8	Pass@8	Maj@8	Pass@8	Maj@8
Independent	95.0±1.8	88.1±2.6	71.6±4.0	59.2±4.3	97.2±1.4	**96.0**±2.0	**96.0**±2.0	95.9±2.0
Regeneration	74.5±3.6	32.0±3.5	59.5±4.2	27.0±3.4	97.5±1.5	58.0±4.3	95.5±2.0	91.5±2.5
Arithmetic	97.0±1.2	88.0±2.9	71.0±4.3	55.5±4.6	97.0±1.7	**96.0**±2.0	**96.0**±2.0	95.0±2.2
Stochastic beam	96.0±1.7	**91.0**±2.8	68.0±4.4	56.5±4.6	95.0±2.2	95.0±2.2	95.0±2.2	95.0±2.2
Verbalized	89.0±2.8	74.0±3.7	55.0±4.5	42.0±4.3	97.5±1.3	95.5±1.9	80.5±3.6	72.5±4.3
MMR, $\lambda = 0.3$	93.0±2.6	86.0±3.5	67.0±4.7	50.0±5.0	98.0±1.4	95.0±2.2	**96.0**±2.0	95.0±2.2
MMR, $\lambda = 0.5$	96.0±2.0	90.0±3.0	74.0±4.4	55.0±5.0	98.0±1.4	95.0±2.2	**96.0**±2.0	**96.0**±2.0
MMR, $\lambda = 0.7$	97.0±1.7	88.0±3.3	75.0±4.4	**64.0**±4.8	98.0±1.4	95.0±2.2	**96.0**±2.0	**96.0**±2.0
k-DPP	95.0±2.2	86.5±3.1	72.5±4.2	61.5±4.6	97.5±1.3	**96.0**±2.0	**96.0**±2.0	95.5±2.0
Release, $\rho = 0$	95.9±1.7	90.5±2.8	74.9±4.1	63.4±4.7	97.4±1.4	**96.0**±2.0	**96.0**±2.0	**96.0**±2.0
Release, $\rho = 1.0$	**97.3**±1.5	78.4±3.6	77.1±4.0	47.2±4.8	98.7±1.0	**96.0**±2.0	**96.0**±2.0	95.3±2.1
Release, $\rho = 2.0$	97.2±1.5	61.1±4.7	**77.2**±4.0	22.8±4.0	**99.0**±1.0	95.0±2.2	**96.0**±2.0	94.6±2.2
DAR, $\rho = 0$	96.1±1.7	89.7±2.9	74.2±4.1	63.0±4.7	97.2±1.5	**96.0**±2.0	**96.0**±2.0	95.5±2.0
DAR, $\rho = 1.0$	96.4±1.7	75.4±4.1	75.5±4.2	46.8±4.9	98.0±1.4	**96.0**±2.0	**96.0**±2.0	95.0±2.2
DAR, $\rho = 2.0$	97.0±1.7	65.4±4.7	75.5±4.2	26.6±4.4	**99.0**±1.0	95.0±2.2	**96.0**±2.0	94.7±2.2

Table 3: Pass@8 and majority@8, in percent ± SE in percentage points, on 100 prompts per dataset and model. Bold marks the highest displayed means.

Averaged over GSM8K and MATH-500, mean pass@8 with Release on Llama increases from 85.4% at $\rho = 0$ to 87.2% at $\rho = 2$, while majority@8 decreases from 77.0% to 42.0%. A larger distortion bound lets Release spread the batch over more distinct final answers, which raises the chance of including the correct answer and lowers the chance that the correct answer is the most frequent. At $\rho = 0$ on Llama, Release also raises pass@8 and majority@8 over independent sampling, and Appendix E.7 reports paired intervals. At $\rho = 2$, DAR attains 86.3% pass@8. Qwen pass@8 reaches 97.5% with both methods. Increasing the Release pool to $N = 128$ gives 88.8% on Llama at $\rho = 2$. Appendix E.8 reports Release with $N = 128$ and diversity and distortion.

GRPO group statistics. Using the GSM8K and MATH-500 pools and final-answer features of the mathematical reasoning experiment, we evaluate GRPO group statistics (Shao et al., 2024) at $B = 16$. Groups with identical rewards have zero centered advantages. We therefore report the constant-reward fraction and the response accuracy in Table 4. Appendix E.9 reports additional results.

Averaged over GSM8K and MATH-500 on Llama at $\rho = 0$, the constant-reward fraction is 21.3% for Release and 26.2% for independent sampling, while response accuracy is 56.9% and 57.0%. At $\rho = 2$, the fraction falls to 13.4% and response accuracy falls to 34.3%. On Qwen, raising ρ from 0 to 2 lowers the constant-reward fraction of Release from 94.2% to 91.5%, while response accuracy changes only from 95.6% to 94.7%. Positive distortion may therefore make GRPO groups more informative on models with high accuracy, and training experiments are needed to test the effect.

	Llama				Qwen			
	GSM8K		MATH-500		GSM8K		MATH-500	
Method	Constant	Accuracy	Constant	Accuracy	Constant	Accuracy	Constant	Accuracy
Independent	23.6±3.0	71.9±2.8	28.7±4.0	42.2±3.5	95.3±1.7	**95.9**±1.9	95.4±1.7	95.3±2.0
Regeneration	28.0±4.5	8.0±0.8	45.0±5.0	7.6±1.0	**41.0**±4.9	54.7±4.0	**82.0**±3.9	85.6±3.1
Arithmetic	21.0±4.1	72.2±2.8	28.0±4.5	41.9±3.6	96.0±2.0	95.8±1.9	95.0±2.2	95.0±2.1
Stochastic beam	20.0±4.0	72.1±2.8	24.0±4.3	42.6±3.6	99.0±1.0	95.1±2.1	100.0±0.0	95.0±2.2
Verbalized	15.0±3.6	60.6±2.9	35.0±4.8	25.9±2.9	94.0±2.4	**95.9**±1.7	**82.0**±3.9	75.7±3.9
MMR, $\lambda = 0.3$	21.0±4.1	66.6±2.9	25.0±4.4	36.8±3.2	93.0±2.6	95.2±2.0	91.0±2.9	94.8±2.0
MMR, $\lambda = 0.5$	23.0±4.2	71.6±2.8	26.0±4.4	41.1±3.4	93.0±2.6	95.4±1.9	94.0±2.4	95.0±2.0
MMR, $\lambda = 0.7$	30.0±4.6	**75.6**±2.7	28.0±4.5	**46.4**±3.6	94.0±2.4	95.5±1.9	95.0±2.2	**95.4**±2.0
k-DPP	23.0±4.2	69.8±3.0	25.0±4.4	39.3±3.5	93.0±2.6	95.8±1.8	94.0±2.4	95.2±2.0
Release, $\rho = 0$	17.7±3.2	71.7±2.8	24.9±4.1	42.1±3.4	94.1±2.2	**95.9**±1.9	94.4±2.0	95.2±2.0
Release, $\rho = 1.0$	9.8±2.5	59.3±3.2	22.6±4.0	30.1±3.0	92.5±2.6	95.8±1.8	91.9±2.6	94.7±2.1
Release, $\rho = 2.0$	**5.5**±2.2	48.1±3.1	**21.3**±4.0	20.6±2.4	92.0±2.7	95.6±1.7	91.0±2.9	93.9±2.2
DAR, $\rho = 0$	17.3±3.1	71.5±2.8	25.5±4.2	42.6±3.5	94.2±2.2	95.8±1.9	95.7±1.9	95.2±2.0
DAR, $\rho = 1.0$	11.0±2.8	59.3±3.1	24.4±4.3	31.0±3.0	93.2±2.5	95.6±1.9	95.5±2.0	94.9±2.1
DAR, $\rho = 2.0$	8.7±2.8	47.9±3.2	28.7±4.5	20.3±2.4	92.0±2.7	95.6±1.7	91.0±2.9	94.1±2.1

Table 4: GRPO statistics at $B = 16$, in percent ± SE in percentage points, on 100 prompts per dataset and model. Constant denotes the constant-reward fraction, and Accuracy denotes the response accuracy. Bold marks the lowest constant fractions and the highest accuracies.

5 Conclusion

We define batch diversity and distortion and characterize the maximum diversity at every distortion bound. Release selects from a pool of independent responses under the distortion bound, and Release diversity approaches the supremum on the support of the model's features when the pool grows and the approximation gap vanishes. DAR chooses the number of additional responses from a pilot and keeps the same distortion bound. In our experiments, Release has higher diversity than independent sampling at zero distortion, and diversity saturates in pool size later under larger distortion bounds. DAR has lower generation cost and lower diversity than Release with the same response limit. On downstream tasks, larger distortion bounds generally increase NoveltyBench utility and answer coverage and decrease majority accuracy, the constant-reward fraction of GRPO groups, and response accuracy. At zero distortion, Release has a lower constant-reward fraction than independent sampling on both models at nearly the same response accuracy.

6 Limitations and Future Work

We do not derive distortion guarantees for changes to decoding or prompting.

Our convergence results provide no finite-sample error bounds and require vanishing optimization error. DAR also requires a fixed pilot size and diverging additional sample sizes. DAR attains lower diversity than Release with the same response limit, and reducing the diversity loss of DAR remains an open direction. Figure 1 stops at $N = 256$ and does not establish proximity to the population supremum. At positive distortion, selection is also restricted to the support of P_0. Experiments cover two models and a limited set of batch sizes. Reported generation-token costs exclude feature computation and optimization. Appendix E.2 reports solver times.

Diversity depends on the chosen features, and maximizing diversity need not maximize downstream utility. We give no rule for choosing ρ. Final-answer classes do not distinguish reasoning strategies that produce the same answer. Future work could examine how feature choices and errors in semantic classification affect diversity and downstream performance.

Our GRPO experiments measure group reward statistics without policy training and do not show gains in learning speed or final performance. Groups built at positive distortion need not be on-policy, so policy-gradient estimates can be biased. A natural next step is to train GRPO on groups built by Release and DAR at several distortion bounds.

AI USE STATEMENT

In this work, we used generative AI tools to help develop the theoretical framework, to formulate mathematical claims, to provide critical ingredients for the proofs and help write the proofs, to give feedback on and refine the research methodology and the design of experiments, to implement methods, to clean and reformat datasets, to assist with translation, to draft and polish text, to find and check related work and references, and to interpret results. We did not use generative AI tools to generate synthetic datasets or to propose or refine hypotheses. Qualitative or thematic data analysis is not applicable to this work. The authors reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Appendices A and C give proofs and optimization procedures. Appendix D specifies the prompt sets, features, calibration, and evaluation. We plan to release the code and numerical results with the camera-ready version.

REFERENCES

Oron Anschel, Alon Shoshan, Adam Botach, Shunit Haviv Hakimi, Asaf Gendler, Emanuel Ben Baruch, Nadav Bhonker, Igor Kviatkovsky, Manoj Aggarwal, and Gerard Medioni. Group-aware reinforcement learning for output diversity in large language models. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2025.

Amin Banayeeanzade, Qingchuan Yang, Dhruv Tarsadiya, Fatemeh Bahrani, Leonardo Blas, Alfy Samuel, Robin Jia, Meisam Razaviyayn, and Sai Praneeth Karimireddy. Sampling more, getting less: Calibration is the diversity bottleneck in LLMs. *arXiv preprint arXiv:2605.11128*, 2026.

Bradley Brown, Jordan Juravsky, Ryan Ehrlich, Ronald Clark, Quoc V. Le, Christopher Ré, and Azalia Mirhoseini. Large language monkeys: Scaling inference compute with repeated sampling. *arXiv preprint arXiv:2407.21787*, 2024.

Jaime Carbonell and Jade Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 1998.

Mark Chen et al. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*, 2021.

Xiwen Chen, Wenhui Zhu, Gen Li, Xuanzhao Dong, Yujian Xiong, Hao Wang, Peijie Qiu, Qingquan Song, Zhipeng Wang, Shao Tang, Yalin Wang, and Abolfazl Razi. OTPrune: Distribution-aligned visual token pruning via optimal transport. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Findings (CVPR Findings)*, 2026a.

Yilei Chen, Souradip Chakraborty, Lorenz Wolf, Yannis Paschalidis, and Aldo Pacchiano. Post-training large language models for diverse high-quality responses. In *International Conference on Learning Representations (ICLR)*, 2026b.

Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations (ICLR)*, 2023.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Gabriele Corso, Yilun Xu, Valentin de Bortoli, Regina Barzilay, and Tommi Jaakkola. Particle guidance: non-I.I.D. diverse sampling with diffusion models. In *International Conference on Learning Representations (ICLR)*, 2024.

Jacques Desrosiers and Marco E. Lübbecke. A primer in column generation. In Guy Desaulniers, Jacques Desrosiers, and Marius M. Solomon (eds.), *Column Generation*, pp. 1–32. Springer, 2005.

Jean-Claude Deville and Yves Tillé. Efficient balanced sampling: The cube method. *Biometrika*, 91 (4):893–912, 2004.

Bryan Eikema and Wilker Aziz. Is MAP decoding all you need? The inadequacy of the mode in neural machine translation. In *International Conference on Computational Linguistics (COLING)*, 2020.

Angela Fan, Mike Lewis, and Yann Dauphin. Hierarchical neural story generation. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2018.

Aaron Grattafiori et al. The Llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. DeBERTaV3: Improving DeBERTa using ELECTRA-style pre-training with gradient-disentangled embedding sharing. In *International Conference on Learning Representations (ICLR)*, 2023.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the MATH dataset. In *Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS Datasets and Benchmarks)*, 2021.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. The curious case of neural text degeneration. In *International Conference on Learning Representations (ICLR)*, 2020.

Daniel G. Horvitz and Donovan J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

Rohit Jena, Ali Taghibakhshi, Sahil Jain, Gerald Shen, Nima Tajbakhsh, and Arash Vahdat. Elucidating optimal reward-diversity tradeoffs in text-to-image diffusion models. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2025.

Liwei Jiang, Yuanjun Chai, Margaret Li, Mickel Liu, Raymond Fok, Nouha Dziri, Yulia Tsvetkov, Maarten Sap, and Yejin Choi. Artificial Hivemind: The open-ended homogeneity of language models (and beyond). In *Advances in Neural Information Processing Systems Datasets and Benchmarks Track (NeurIPS Datasets and Benchmarks)*, 2025.

Wouter Kool, Herke van Hoof, and Max Welling. Stochastic beams and where to find them: The Gumbel-top-k trick for sampling sequences without replacement. In *International Conference on Machine Learning (ICML)*, 2019.

Alex Kulesza and Ben Taskar. k-DPPs: Fixed-size determinantal point processes. In *International Conference on Machine Learning (ICML)*, 2011.

Hunter Lightman, Vineet Kosaraju, Yura Burda, Harri Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. In *International Conference on Learning Representations (ICLR)*, 2024.

Chris Yuhao Liu, Liang Zeng, Jiacai Liu, Rui Yan, Jujie He, Chaojie Wang, Shuicheng Yan, Yang Liu, and Yahui Zhou. Skywork-Reward: Bag of tricks for reward modeling in LLMs. *arXiv preprint arXiv:2410.18451*, 2024.

Minh Nhat Nguyen, Andrew Baker, Clement Neo, Allen Roush, Andreas Kirsch, and Ravid Shwartz-Ziv. Turning up the heat: Min-p sampling for creative and coherent LLM outputs. In *International Conference on Learning Representations (ICLR)*, 2025.

Jiahao Qiu, Yifu Lu, Yifan Zeng, Jiacheng Guo, Jiayi Geng, Chenhao Zhu, Xinzhe Juan, Ling Yang, Huazheng Wang, Kaixuan Huang, Yue Wu, and Mengdi Wang. TreeBoN: Enhancing inference-time alignment with speculative tree-search and best-of-N sampling. In *Findings of the Association for Computational Linguistics: EMNLP (EMNLP Findings)*, 2025.

Yiming Ren, Yiran Xu, Zicheng Lin, Chufan Shi, Yukang Chen, Dingdong Wang, Tianhe Wu, Junjie Wang, Yujiu Yang, Yu Qiao, and Ruihang Chu. Smaller models are natural explorers for policy-level diversity in GRPO. In *International Conference on Machine Learning (ICML)*, 2026.

Seyedmorteza Sadat, Jakob Buhmann, Derek Bradley, Otmar Hilliges, and Romann M. Weber. CADS: Unleashing the diversity of diffusion models through condition-annealed sampling. In *International Conference on Learning Representations (ICLR)*, 2024.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Hanshi Sun, Momin Haider, Ruiqi Zhang, Huitao Yang, Jiahao Qiu, Ming Yin, Mengdi Wang, Peter L. Bartlett, and Andrea Zanette. Fast best-of-N decoding via speculative rejection. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Ashwin K. Vijayakumar, Michael Cogswell, Ramprasaath R. Selvaraju, Qing Sun, Stefan Lee, David Crandall, and Dhruv Batra. Diverse beam search for improved description of complex scenes. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

Luke Vilnis, Yury Zemlyanskiy, Patrick Murray, Alexandre Passos, and Sumit Sanghai. Arithmetic sampling: Parallel diverse decoding for large language models. In *International Conference on Machine Learning (ICML)*, 2023.

Gal Vinograd, Idan Achituve, and Ethan Fetaya. Diverse sampling in diffusion models with marginal preserving particle guidance. *arXiv preprint arXiv:2605.06553*, 2026.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *International Conference on Learning Representations (ICLR)*, 2023.

Jeffrey Wen, Rizwan Ahmad, and Philip Schniter. Task-driven uncertainty quantification in inverse problems via conformal prediction. In *European Conference on Computer Vision (ECCV)*, 2024.

Shitao Xiao, Zheng Liu, Peitian Zhang, Niklas Muennighoff, Defu Lian, and Jian-Yun Nie. C-Pack: Packed resources for general Chinese embeddings. In *International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 2024.

An Yang et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Jian Yao, Ran Cheng, Xingyu Wu, Jibin Wu, and Kay Chen Tan. Diversity-aware policy optimization for large language model reasoning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Jiayi Zhang, Simon Yu, Derek Chong, Anthony Sicilia, Michael R. Tomz, Christopher D. Manning, and Weiyan Shi. Verbalized sampling: How to mitigate mode collapse and unlock LLM diversity. In *International Conference on Machine Learning (ICML)*, 2026.

Yiming Zhang, Harshita Diddee, Susan Holm, Hanchen Liu, Xinyue Liu, Vinay Samuel, Barry Wang, and Daphne Ippolito. NoveltyBench: Evaluating language models for humanlike diversity. In *Conference on Language Modeling (COLM)*, 2025.

Wenting Zhao, Xiang Ren, Jack Hessel, Claire Cardie, Yejin Choi, and Yuntian Deng. WildChat: 1M ChatGPT interaction logs in the wild. In *International Conference on Learning Representations (ICLR)*, 2024.

A PROOFS FOR SECTION 2

Keep the notation and assumptions of Section 2. Let $\mathcal{C}(P_0)$ be the set of joint distributions with $\Pi_i = P_0$ for every $i = 1, \dots, B$.

Lemma A.1. *Let $X \sim \Pi$. Independently of X, let U be uniform on $\{1, \dots, B\}$ and let σ be uniform on the $B!$ permutations of $\{1, \dots, B\}$. Define $X_i^\sigma = X_{\sigma(i)}$. Then $X_U \sim \bar{\Pi}$ and $X_i^\sigma \sim \bar{\Pi}$ for every i. Moreover, $\bar{X^\sigma} = \bar{X}$ almost surely and $\mathrm{Div}(\mathcal{L}(X^\sigma)) = \mathrm{Div}(\Pi)$.*

Proof of Lemma A.1. For every Borel set $A \subset \mathbb{R}^d$ and every i,

$$\Pr(X_U \in A) = \frac{1}{B}\sum_{j=1}^{B} \Pi_j(A) = \bar{\Pi}(A), \qquad \Pr(X_i^\sigma \in A) = \frac{1}{B}\sum_{j=1}^{B} \Pi_j(A).$$

Permuting coordinates preserves their mean and the sum of pairwise squared distances, so $\bar{X^\sigma} = \bar{X}$ and $\mathrm{Div}(\mathcal{L}(X^\sigma)) = \mathrm{Div}(\Pi)$. □

Since W_2 is a metric, $D(\Pi) = 0$ if and only if $\bar{\Pi} = P_0$. Under zero distortion, Lemma A.1 gives $X_i^\sigma \sim P_0$ for every i.

For fixed P_0, Div is affine in Π.

Lemma A.2. *For every joint distribution Π in the setting of Section 2 and every $a \in \mathbb{R}^d$,*

$$\mathrm{Div}(\Pi) = \frac{c_B}{s_0^2}\left(\int \|y - a\|^2 \, d\bar{\Pi}(y) - \mathbb{E}_\Pi \|\bar{X} - a\|^2\right). \tag{11}$$

Proof of Lemma A.2. For $z_1, \dots, z_B, a \in \mathbb{R}^d$, put $\bar{z} = B^{-1}\sum_i z_i$. Expansion gives

$$\sum_{i<j} \|z_i - z_j\|^2 = B\sum_i \|z_i - \bar{z}\|^2 = B\left(\sum_i \|z_i - a\|^2 - B\|\bar{z} - a\|^2\right). \tag{12}$$

Substituting $z_i = X_i$, taking expectations, and dividing by $s_0^2 B(B-1)$ gives (11). □

For square-integrable random batches on a common probability space, write

$$\|X\|_{\mathcal{H}}^2 = \frac{1}{B}\sum_i \mathbb{E}\|X_i\|^2, \qquad R(X)^2 = \frac{1}{B}\sum_i \mathbb{E}\|X_i - \bar{X}\|^2.$$

Centering $X - Z$ gives

$$\frac{1}{B}\sum_i \mathbb{E}\|(X_i - \bar{X}) - (Z_i - \bar{Z})\|^2 = \|X - Z\|_{\mathcal{H}}^2 - \mathbb{E}\|\bar{X} - \bar{Z}\|^2 \le \|X - Z\|_{\mathcal{H}}^2.$$

The reverse triangle inequality and (12) imply

$$|R(X) - R(Z)| \le \|X - Z\|_{\mathcal{H}}, \qquad \mathrm{Div}(\mathcal{L}(X)) = \frac{c_B}{s_0^2} R(X)^2. \tag{13}$$

Lemma A.3. *Let $\mu, \nu_1, \dots, \nu_B$ be Borel probability measures on $\mathbb{R}^d$ with finite second moments, and let X have all coordinate marginals equal to μ. On an extension of the probability space there is a batch Z, with $Z_i \sim \nu_i$, such that*

$$\mathbb{E}\|X_i - Z_i\|^2 = W_2(\mu, \nu_i)^2, \qquad \|X - Z\|_{\mathcal{H}}^2 = \frac{1}{B}\sum_i W_2(\mu, \nu_i)^2. \tag{14}$$

Proof of Lemma A.3. For each i, the set of couplings of μ and ν_i is tight and weakly closed, hence weakly compact. The nonnegative continuous cost $(x, y) \mapsto \|x - y\|^2$ induces a weakly lower semicontinuous integral, and the product coupling has finite cost. Therefore an optimal coupling γ_i

exists. Since $\mathbb{R}^d$ is a standard Borel space, write $\gamma_i(dx, dy) = \mu(dx)K_i(x, dy)$ for a probability kernel K_i. Define the joint distribution

$$\mathcal{L}(X, Z)(dx, dz) = \mathcal{L}(X)(dx) \prod_{i=1}^{B} K_i(x_i, dz_i).$$

The (X_i, Z_i) marginal equals $\mu(dx_i)K_i(x_i, dz_i) = \gamma_i(dx_i, dz_i)$. Hence $Z_i \sim \nu_i$ and $\mathbb{E}\|X_i - Z_i\|^2 = W_2(\mu, \nu_i)^2$. Summing over i proves (14). □

Proof of Proposition 2.2. For a Borel probability measure μ on $\mathbb{R}^d$ with finite second moment, the only coupling with a point mass at m_0 gives

$$W_2(\mu, \delta_{m_0})^2 = \int \|y - m_0\|^2 \, d\mu(y).$$

By (11), $\mathbb{E}\|\bar{X} - m_0\|^2 \geq 0$, and the triangle inequality,

$$\mathrm{Div}(\Pi) \leq \frac{c_B}{s_0^2} W_2(\bar{\Pi}, \delta_{m_0})^2 \leq \frac{c_B}{s_0^2} \big(W_2(\bar{\Pi}, P_0) + s_0\big)^2.$$

Substitute $W_2(\bar{\Pi}, P_0) = s_0 D(\Pi)$. If $M_B(P_0) = 0$, the equality assertion follows from Theorem 2.3, proved below. □

Proof of Theorem 2.3. For compact $K \subset \mathbb{R}^d$,

$$\sup_{\Pi \in \mathcal{C}(P_0)} \Pi((K^B)^c) \leq B P_0(K^c).$$

Thus $\mathcal{C}(P_0)$ is tight. Coordinate projections are continuous, so $\mathcal{C}(P_0)$ is weakly closed and hence weakly compact. The nonnegative continuous integrand $\|\bar{X} - m_0\|^2$ gives a weakly lower semicontinuous objective. Therefore some $\Pi^* \in \mathcal{C}(P_0)$ attains $M_B(P_0)$. By (12),

$$R(X)^2 = s_0^2 - \mathbb{E}\|\bar{X} - m_0\|^2 \quad \text{for } \mathcal{L}(X) \in \mathcal{C}(P_0).$$

Set $r_0 = \sqrt{s_0^2 - M_B(P_0)}$. Since $M_B(P_0) \leq s_0^2/B$,

$$r_0 \geq s_0\sqrt{1 - 1/B} > 0, \qquad \max_{\mathcal{L}(X) \in \mathcal{C}(P_0)} R(X) = r_0.$$

Lemma A.1 preserves R under symmetrization and gives

$$\max_{D(\Pi)=0} \mathrm{Div}(\Pi) = \frac{c_B r_0^2}{s_0^2}.$$

Let $Z \sim \Pi$ and set $Q = \bar{\Pi}$. Apply the independent uniform permutation from Lemma A.1, so $Z_i^\sigma \sim Q$ and $R(Z^\sigma) = R(Z)$. Lemma A.3, with $\mu = Q$ and $\nu_i = P_0$, gives $\mathcal{L}(X) \in \mathcal{C}(P_0)$ and $\|Z^\sigma - X\|_{\mathcal{H}} = W_2(Q, P_0)$. Therefore

$$R(Z) = R(Z^\sigma) \leq R(X) + \|Z^\sigma - X\|_{\mathcal{H}} \leq r_0 + W_2(Q, P_0). \tag{15}$$

Equation (13) now bounds the diversity of every batch with $D \leq \rho$ by the right side of (2).

For $\rho \geq 0$, let $X \sim \Pi^*$ and set

$$Z_i = X_i + \frac{\rho s_0}{r_0}(X_i - \bar{X}), \qquad i = 1, \ldots, B. \tag{16}$$

Then $\bar{Z} = \bar{X}$ and $R(Z) = r_0 + \rho s_0$. Let U be independent of (X, Z) and uniform on $\{1, \ldots, B\}$. The pair (X_U, Z_U) couples P_0 and $Q = \mathcal{L}(Z_U)$, and

$$\mathbb{E}\|Z_U - X_U\|^2 = \frac{\rho^2 s_0^2}{r_0^2} \frac{1}{B} \sum_i \mathbb{E}\|X_i - \bar{X}\|^2 = \rho^2 s_0^2.$$

Hence $D(\mathcal{L}(Z)) \leq \rho$, and (13) gives $\mathrm{Div}(\mathcal{L}(Z)) = c_B(r_0 + \rho s_0)^2/s_0^2$. Conversely, (15) gives $\rho s_0 = R(Z) - r_0 \leq W_2(Q, P_0)$. Therefore $D(\mathcal{L}(Z)) = \rho$, and the diversity equals (2). □

Proof of Corollary 2.4. Fix $v \geq 0$ and let ρ denote the right-hand side of (3). If $\mathrm{Div}(\Pi) \geq v$, Theorem 2.3 gives $\sqrt{v/c_B} \leq \sqrt{1 - M_B(P_0)/s_0^2} + D(\Pi)$, and $D(\Pi) \geq 0$, so $D(\Pi) \geq \rho$. Theorem 2.3 also provides a joint distribution Π_ρ with $D(\Pi_\rho) = \rho$ and $\mathrm{Div}(\Pi_\rho) = c_B\big(\sqrt{1 - M_B(P_0)/s_0^2} + \rho\big)^2 \geq v$. □

For a probability mass function q on $\mathcal{S}_N$, put $\pi_i(q) = \sum_{J:\, i \in J} q(J)$.

Lemma A.4. *For every pool x and probability mass function q on $\mathcal{S}_N$,*

$$\bar{\Pi}_{x,q} = \sum_{i=1}^{N} \frac{\pi_i(q)}{B} \delta_{x_i}.$$

Thus $\pi_i(q) = B/N$ for every i implies $\bar{\Pi}_{x,q} = P_x$. Conversely, if $\bar{\Pi}_{x,q} = P_x$, there is a distribution q' with $\pi_i(q') = B/N$ for every i and $\Pi_{x,q'} = \Pi_{x,q}$.

Proof of Lemma A.4. Let $\Pr(S = J) = q(J)$ and, conditional on S, let I be uniform on S. Then $\Pr(I = i) = \pi_i(q)/B$, so $\bar{\Pi}_{x,q} = \mathcal{L}(x_I) = \sum_{i=1}^{N} \pi_i(q)\delta_{x_i}/B$.

For the converse, let G_x be the finite set of permutations τ of $\{1, \ldots, N\}$ satisfying $x_{\tau(i)} = x_i$ for every i. Independently of S, draw τ uniformly from G_x and let q' be the distribution of $\tau(S)$. Since $x_{\tau(i)} = x_i$, uniformly ordering the selected feature vectors gives $\Pi_{x,q'} = \Pi_{x,q}$. For each feature value v in the pool, let $I_v = \{i : x_i = v\}$. The assumption $\bar{\Pi}_{x,q} = P_x$ gives

$$\sum_{i \in I_v} \pi_i(q) = \frac{B|I_v|}{N}.$$

Uniform permutation within I_v therefore gives $\pi_i(q') = |I_v|^{-1} \sum_{j \in I_v} \pi_j(q) = B/N$ for every $i \in I_v$. Thus q' selects every candidate with probability B/N and $\Pi_{x,q'} = \Pi_{x,q}$. □

Proof of Theorem 2.5(a). Every output has finite second moment since $\|\xi_{I_j}\|^2 \leq \sum_{i=1}^{N} \|\xi_i\|^2$. Moreover,

$$\bar{\Pi}_N = \mathbb{E}[\bar{\Pi}_{x_N, q_{x_N}}], \qquad \mathbb{E}P_{x_N} = P_0, \qquad \mathbb{E}s_{x_N}^2 = (1 - 1/N)s_0^2.$$

Joint convexity of W_2^2 under mixtures gives

$$\begin{aligned} W_2(\bar{\Pi}_N, P_0)^2 &\leq \mathbb{E}\Big[W_2(\bar{\Pi}_{x_N, q_{x_N}}, P_{x_N})^2\Big] \\ &\leq \eta_N(\rho)^2 \mathbb{E}s_{x_N}^2 = \rho^2 s_0^2. \end{aligned}$$

Dividing by s_0^2 proves $D(\Pi_N) \leq \rho$. □

Lemma A.5. *The function $F(\rho) := \mathrm{Div}_E^\star(\rho)$ is finite, nondecreasing, and continuous on $[0, \infty)$. Moreover, $F(0)$ equals the maximum in (2) at $\rho = 0$.*

Proof of Lemma A.5. If $\bar{\Pi} = P_0$, then $B^{-1} \sum_i \Pi_i(E^c) = P_0(E^c) = 0$. Hence every $\Pi_i(E^c) = 0$, and $\Pi(E^B) = 1$ by the union bound. Therefore $F(0)$ equals the maximum in (2) at $\rho = 0$. The feasible sets increase with ρ, so F is nondecreasing. Proposition 2.2 gives $F(\rho) \leq c_B(1+\rho)^2 < \infty$.

To prove concavity of $t \mapsto F(\sqrt{t})$, fix $t_1, t_2 \geq 0$, $\lambda \in [0, 1]$, and joint distributions Π^k feasible for $F(\sqrt{t_k})$, $k = 1, 2$. Let γ_k be an optimal coupling of $\bar{\Pi}^k$ and P_0. The probability measure $\Pi = \lambda\Pi^1 + (1 - \lambda)\Pi^2$ is supported on E^B, and $\lambda\gamma_1 + (1 - \lambda)\gamma_2$ couples $\bar{\Pi}$ and P_0. Thus

$$W_2(\bar{\Pi}, P_0)^2 \leq s_0^2\big(\lambda t_1 + (1 - \lambda)t_2\big), \qquad \mathrm{Div}(\Pi) = \lambda \mathrm{Div}(\Pi^1) + (1 - \lambda)\mathrm{Div}(\Pi^2).$$

Taking the supremum over Π^1 and Π^2 gives

$$F\Big(\sqrt{\lambda t_1 + (1 - \lambda)t_2}\Big) \geq \lambda F(\sqrt{t_1}) + (1 - \lambda)F(\sqrt{t_2}).$$

A finite concave function on $[0, \infty)$ is continuous on $(0, \infty)$, so F is continuous at every $\rho > 0$. At zero, Theorem 2.3 gives

$$F(0) \leq F(\rho) \leq \big(\sqrt{F(0)} + \sqrt{c_B}\rho\big)^2.$$

Letting $\rho \downarrow 0$ proves right continuity at zero. □

Proof of Theorem 2.5(b). First prove convergence for each fixed $\eta \geq 0$. Let $P_N = P_{x_N}$, $s_N = s_{x_N}$, and $V_N(\eta) = \mathrm{Div}^\star_{x_N}(\eta)$. For $j = 1, \ldots, B$ and $L \geq 1$, define

$$P_{j,L} = \frac{1}{L} \sum_{\ell=0}^{L-1} \delta_{\xi_{j+\ell B}}.$$

Almost-sure weak convergence of empirical measures and the strong law of large numbers applied to squared norms give

$$W_2(P_N, P_0) \longrightarrow 0, \qquad \max_{1 \leq j \leq B} W_2(P_{j,L}, P_0) \longrightarrow 0 \quad \text{almost surely as } N, L \to \infty. \tag{17}$$

Also $\Pr(\xi_i \in E \text{ for every } i \geq 1) = 1$. Fix a realization satisfying (17) and $\xi_i \in E$ for every i. For every open set $A \subset \mathbb{R}^d$ with $A \cap E \neq \varnothing$, the definition of support gives $P_0(A) > 0$. By the Portmanteau theorem, for every $j = 1, \ldots, B$,

$$\liminf_{L \to \infty} P_{j,L}(A) \geq P_0(A) > 0.$$

Hence $\xi_{j+\ell B} \in A$ for some $\ell \geq 0$. Since E is closed and contains every ξ_i,

$$\overline{\{\xi_{j+\ell B} : \ell \geq 0\}} = E \qquad (j = 1, \ldots, B). \tag{18}$$

Convergence in W_2 gives convergence of the first and second moments, so $s_N^2 = \int \|y\|^2 \, dP_N(y) - \| \int y \, dP_N(y) \|^2 \to s_0^2$. In particular, $s_N > 0$ for all sufficiently large N. All remaining limits concern the fixed realization.

Put $\delta_N = W_2(P_N, P_0)$. The joint distribution Π obtained from any feasible q on the pool satisfies $\Pi(E^B) = 1$ and

$$W_2(\bar{\Pi}, P_0) \leq W_2(\bar{\Pi}, P_N) + \delta_N \leq \eta s_N + \delta_N, \qquad \mathrm{Div}(\Pi) = \frac{s_N^2}{s_0^2} \mathrm{Div}_{x_N}(\Pi).$$

Taking the maximum over feasible distributions q on the pool gives

$$\frac{s_N^2}{s_0^2} V_N(\eta) \leq F\left(\frac{\eta s_N + \delta_N}{s_0} \right). \tag{19}$$

Since $s_N \to s_0$, $\delta_N \to 0$, and F is continuous by Lemma A.5, $\limsup_N V_N(\eta) \leq F(\eta)$.

For $N = BL$ and $\eta > 0$, fix a joint distribution Π feasible for $F(\eta)$ and $t \in (0, 1)$. Set $\Pi^t = (1-t)\Pi + t P_0^{\otimes B}$. Let γ be an optimal coupling of $\bar{\Pi}$ and P_0. The measure $(1-t)\gamma + t \int \delta_{(y,y)} \, dP_0(y)$ couples $\bar{\Pi}^t$ and P_0. Hence

$$\Pi^t(E^B) = 1, \qquad W_2(\bar{\Pi}^t, P_0)^2 \leq (1-t)\eta^2 s_0^2 < \eta^2 s_0^2.$$

For $j = 1, \ldots, B$ and $y \in \mathbb{R}^d$, define the candidate index

$$I_{j,L}(y) = \min \operatorname*{arg\,min}_{i \in \{j, j+B, \ldots, j+(L-1)B\}} \|\xi_i - y\|.$$

Each map $I_{j,L}$ is Borel measurable. Equation (18) gives $\|\xi_{I_{j,L}(y)} - y\| \to 0$ for every $y \in E$, and

$$\|\xi_{I_{j,L}(y)} - y\|^2 \leq \|\xi_j - y\|^2.$$

Let $X \sim \Pi^t$ and define $Z_j^{(L)} = \xi_{I_{j,L}(X_j)}$. Since $X_j \in E$ almost surely and $\mathbb{E}\|\xi_j - X_j\|^2 < \infty$, dominated convergence gives $\|Z^{(L)} - X\|_{\mathcal{H}} \to 0$.

Set $S_L = \{I_{j,L}(X_j) : 1 \leq j \leq B\}$ and put $q_L(J) = \Pr(S_L = J)$ for $J \in \mathcal{S}_N$. Distinct values of j give indices in disjoint residue classes modulo B, so $|S_L| = B$. An independent uniform permutation of $Z^{(L)}$ has distribution Π_{x_N, q_L} and preserves diversity. Let U be independent of $(X, Z^{(L)})$ and uniform on $\{1, \ldots, B\}$. Then

$$\bar{\Pi}_{x_N, q_L} = \mathcal{L}(Z_U^{(L)}), \qquad \mathrm{Div}(\Pi_{x_N, q_L}) = \mathrm{Div}(\mathcal{L}(Z^{(L)})).$$

The pair $(Z_U^{(L)}, X_U)$ couples $\bar{\Pi}_{x_N,q_L}$ and $\bar{\Pi}^t$. Hence

$$W_2(\bar{\Pi}_{x_N,q_L}, P_N) \le \|Z^{(L)} - X\|_{\mathcal{H}} + W_2(\bar{\Pi}^t, P_0) + \delta_N < \eta s_N$$

for all sufficiently large L, since $W_2(\bar{\Pi}^t, P_0) < \eta s_0$ and $s_N \to s_0$. Thus q_L is feasible for $V_N(\eta)$. Equation (13) gives

$$\mathrm{Div}_{x_N}(\Pi_{x_N,q_L}) = \frac{c_B}{s_N^2} R(Z^{(L)})^2 \longrightarrow \mathrm{Div}(\Pi^t).$$

Therefore $\liminf_{L\to\infty} V_{BL}(\eta) \ge \mathrm{Div}(\Pi^t)$. Since $\mathrm{Div}(\Pi^t) = (1-t)\mathrm{Div}(\Pi) + t$, letting $t \downarrow 0$ and taking the supremum over feasible Π gives $\liminf_{L\to\infty} V_{BL}(\eta) \ge F(\eta)$.

For $N = BL$ and $\eta = 0$, Theorem 2.3 gives X with $X_j \sim P_0$ for every j and $\mathrm{Div}(\mathcal{L}(X)) = F(0)$. For each L, Lemma A.3 gives a coupling of X and $Z^{(L)}$ such that $Z_j^{(L)} \sim P_{j,L}$ and

$$\|Z^{(L)} - X\|_{\mathcal{H}}^2 = \frac{1}{B}\sum_{j=1}^{B} W_2(P_0, P_{j,L})^2 \longrightarrow 0.$$

Conditional on $Z^{(L)}$, draw $I_j^{(L)}$ independently and uniformly from $\{j + \ell B : 0 \le \ell < L,\ \xi_{j+\ell B} = Z_j^{(L)}\}$. For $i \in \{j, j+B, \dots, j+(L-1)B\}$,

$$\Pr(I_j^{(L)} = i) = \frac{P_{j,L}(\{\xi_i\})}{\#\{0 \le \ell < L : \xi_{j+\ell B} = \xi_i\}} = \frac{1}{L}.$$

The indices $I_1^{(L)}, \dots, I_B^{(L)}$ belong to disjoint residue classes. Therefore $S_L = \{I_1^{(L)}, \dots, I_B^{(L)}\}$ has cardinality B, and $q_L(J) = \Pr(S_L = J)$ defines a distribution selecting each index with probability $1/L = B/N$. By Lemma A.4, $\bar{\Pi}_{x_N,q_L} = P_N$, so q_L is feasible at $\eta = 0$. An independent uniform permutation of $Z^{(L)}$ has distribution Π_{x_N,q_L}. Equations (12) and (13) give $\mathrm{Div}_{x_N}(\Pi_{x_N,q_L}) = c_B R(Z^{(L)})^2/s_N^2 \to F(0)$. Hence $\liminf_{L\to\infty} V_{BL}(0) \ge F(0)$.

For any $N \ge B$, set $M = B\lfloor N/B \rfloor$ and $r = N - M$. Then $0 \le r < B$, $M/N \to 1$, $s_M/s_N \to 1$, and $W_2(P_M, P_N) \to 0$. The upper bound (19) holds for every N.

Fix $0 < \eta' < \eta$ and a probability mass function q_M on $\mathcal{S}_M$ attaining $V_M(\eta')$. Regarding q_M as a distribution on subsets of $\{1, \dots, N\}$ gives

$$W_2(\bar{\Pi}_{x_M,q_M}, P_N) \le \eta' s_M + W_2(P_M, P_N) < \eta s_N$$

for all sufficiently large N. Consequently,

$$V_N(\eta) \ge \frac{s_M^2}{s_N^2} V_M(\eta'). \tag{20}$$

Since M is a multiple of B, taking the lower limit in (20) and then letting $\eta' \uparrow \eta$ gives $\liminf_N V_N(\eta) \ge F(\eta)$ by continuity of F.

At $\eta = 0$, choose an optimal q_M with inclusion probability B/M for every index. Let u_N select all r indices in $\{M+1, \dots, N\}$ and a uniform $(B-r)$-element subset of $\{1, \dots, M\}$. Define

$$\widetilde{q}_N = \left(1 - \frac{B}{N}\right) q_M + \frac{B}{N} u_N.$$

For $i \le M$, the inclusion probability under $\widetilde{q}_N$ is

$$\left(1 - \frac{B}{N}\right)\frac{B}{M} + \frac{B}{N}\frac{B-r}{M} = \frac{B}{N}.$$

For $M < i \le N$, the inclusion probability is also B/N. Thus $\widetilde{q}_N$ is feasible at zero distortion. Nonnegativity of pairwise squared distances gives

$$V_N(0) \ge \left(1 - \frac{B}{N}\right)\frac{s_M^2}{s_N^2} V_M(0) \longrightarrow F(0). \tag{21}$$

Together with (19), the lower bounds (20) and (21) establish $V_N(\eta) \to F(\eta)$ for every $\eta \geq 0$ along all pool sizes.

Since $\eta_N(0) = 0$, equation (7) holds at $\rho = 0$. Fix $\rho > 0$ and a realization on which $V_N(t) \to F(t)$ for $t = \rho$ and $t = \rho + 1/k$, $k \geq 1$. Such realizations have probability one. For each $k \geq 1$, eventually

$$\rho \leq \eta_N(\rho) \leq \rho + 1/k, \qquad V_N(\rho) \leq V_N(\eta_N(\rho)) \leq V_N(\rho + 1/k).$$

Taking lower and upper limits gives

$$F(\rho) \leq \liminf_N V_N(\eta_N(\rho)) \leq \limsup_N V_N(\eta_N(\rho)) \leq F(\rho + 1/k).$$

Letting $k \to \infty$ and using Lemma A.5 proves (7). □

B Relations to Existing Diversity and Distortion Measures

B.1 Diversity and existing metrics

Under $P_0^{\otimes B}$, $\mathbb{E}\|X_i - X_j\|^2 = 2s_0^2$ for $i \neq j$. Hence

$$\mathrm{Div}(\Pi) = \frac{\sum_{i<j} \mathbb{E}_{\Pi}\|X_i - X_j\|^2}{\sum_{i<j} \mathbb{E}_{P_0^{\otimes B}}\|X_i - X_j\|^2}.$$

For unit vectors u, v, $\|u - v\|^2 = 2(1 - \langle u, v\rangle)$. Thus, if P_0 and $\bar{\Pi}$ are supported on the unit sphere, the same ratio holds for cosine distances. Under independent sampling, the expected average cosine distance is $s_0^2 = 1 - \|m_0\|^2$.

B.2 Distortion and existing metrics

Class probabilities at arbitrary distortion. Represent an answer in class j by the one-hot vector e_j. Let $P_0 = \sum_{j=1}^C p_j \delta_{e_j}$ be a nondegenerate distribution on the standard basis of $\mathbb{R}^C$, and suppose $\bar{\Pi} = \sum_{j=1}^C r_j \delta_{e_j}$. Then

$$D(\Pi)^2 = \frac{2\,\mathrm{TV}(p, r)}{1 - \sum_j p_j^2}, \qquad \mathrm{TV}(p, r) = \frac{1}{2}\sum_j |p_j - r_j|. \tag{22}$$

Put $t = \mathrm{TV}(p, r) = 1 - \sum_j \min(p_j, r_j)$. Every coupling matrix γ of p and r satisfies $\gamma_{jj} \leq \min(p_j, r_j)$, so $\sum_{j \neq k} \gamma_{jk} \geq t$. Since $\|e_j - e_k\|^2 = 2$ for $j \neq k$, every coupling has cost at least $2t$. If $t = 0$, the diagonal coupling has cost zero. If $t > 0$, define

$$\gamma_{jk} = \mathbf{1}_{\{j=k\}} \min(p_j, r_j) + \frac{(p_j - r_j)_+ (r_k - p_k)_+}{t}.$$

The row sums are p_j, the column sums are r_k, and the diagonal sum is $1 - t$. Hence the coupling has cost $2t$. Finally, $s_0^2 = 1 - \sum_j p_j^2$, which proves (22).

Mean individual accuracy at arbitrary distortion. Represent a correct answer by 1 and an incorrect answer by 0. Let $P_0 = (1 - p)\delta_0 + p\delta_1$, $0 < p < 1$, and suppose $\Pi(\{0, 1\}^B) = 1$. Define $r = B^{-1} \sum_i \Pr_{\Pi}(X_i = 1)$, the mean accuracy over the B output positions. Then $\bar{\Pi} = (1 - r)\delta_0 + r\delta_1$ and

$$D(\Pi)^2 = \frac{|r - p|}{p(1 - p)}. \tag{23}$$

In particular, $D(\Pi) \leq \rho$ if and only if $|r - p| \leq p(1 - p)\rho^2$. For the probability vectors $(1 - p, p)$ and $(1 - r, r)$, total variation equals $|r - p|$. The coupling lower bound and construction in the proof of (22), with squared distance 1 between 0 and 1, give $W_2(P_0, \bar{\Pi})^2 = |r - p|$. Division by $s_0^2 = p(1 - p)$ proves (23).

B.3 Diversity and existing objectives at zero distortion

Example B.1. For one-hot class features, let $P_0 = \sum_{j=1}^{C} p_j \delta_{e_j}$ be a nondegenerate distribution on the standard basis of $\mathbb{R}^C$. For $X \sim \Pi$ with $D(\Pi) = 0$, define $K_j = |\{i : X_i = e_j\}|$. Then $\mathbb{E}K_j = Bp_j$. Every diversity-maximizing distribution at zero distortion satisfies $K_j \in \{\lfloor Bp_j \rfloor, \lceil Bp_j \rceil\}$ almost surely for every j, and also attains

$$\max_{\Pi:\, D(\Pi)=0} \mathbb{E}_\Pi |\{j : K_j > 0\}| = \sum_{j=1}^{C} \min\{1, Bp_j\}. \tag{24}$$

Independent sampling gives $\sum_j [1 - (1-p_j)^B]$.

Proof of Example B.1. Since $\bar{\Pi} = P_0$ and $s_0^2 = 1 - \sum_j p_j^2 > 0$,

$$\sum_j K_j = B, \qquad \mathbb{E}K_j = Bp_j, \qquad \mathbb{E}\|\bar{X} - m_0\|^2 = \frac{1}{B^2} \sum_j \mathrm{Var}(K_j).$$

Conversely, any nonnegative integer counts with $\sum_j K_j = B$ and $\mathbb{E}K_j = Bp_j$ define a coupling in $\mathcal{C}(P_0)$: uniformly order K_j copies of each e_j. Then

$$\Pr(X_i = e_j \mid K) = K_j/B, \qquad \Pr(X_i = e_j) = p_j.$$

Let K be a square-integrable integer-valued random variable with mean $a + \theta$, where $a \in \mathbb{Z}$ and $0 \le \theta < 1$. Expansion gives

$$\mathrm{Var}(K) - \theta(1-\theta) = \mathbb{E}[(K-a)(K-a-1)] \ge 0. \tag{25}$$

For integer k, $(k-a)(k-a-1) \ge 0$, with equality only at $a, a+1$. Hence equality in (25) is equivalent to $K \in \{a, a+1\}$ almost surely.

Set $a_j = \lfloor Bp_j \rfloor$ and $\theta_j = Bp_j - a_j$. Equation (25) gives $\mathrm{Var}(K_j) \ge \theta_j(1-\theta_j)$ for every j. To attain every lower bound simultaneously, put $b_0 = 0$ and $b_j = B\sum_{\ell=1}^{j} p_\ell$. Draw V uniformly on $[0,1)$ and set

$$K_j = \big|\{k \in \{0, \dots, B-1\} : b_{j-1} \le V + k < b_j\}\big|.$$

The intervals partition $[0, B)$, so $\sum_j K_j = B$. Each interval has length Bp_j and contains either $\lfloor Bp_j \rfloor$ or $\lceil Bp_j \rceil$ of the equally spaced points. Integrating over V gives $\mathbb{E}K_j = Bp_j$, hence $\mathrm{Var}(K_j) = \theta_j(1-\theta_j)$. Uniformly permuting the list with K_j copies of e_j gives a joint distribution in $\mathcal{C}(P_0)$, so

$$M_B(P_0) = \frac{1}{B^2} \sum_j \theta_j(1-\theta_j). \tag{26}$$

Hence $M_B(P_0) = 0$ if and only if every Bp_j is an integer.

For every Π with $\bar{\Pi} = P_0$, Lemma A.2 gives

$$\mathrm{Div}(\Pi) = c_B \left(1 - \frac{1}{B^2 s_0^2} \sum_{j=1}^{C} \mathrm{Var}(K_j)\right).$$

Thus Π maximizes diversity at zero distortion if and only if, for every j,

$$\mathrm{Var}(K_j) = \theta_j(1-\theta_j), \quad \text{equivalently} \quad K_j \in \{\lfloor Bp_j \rfloor, \lceil Bp_j \rceil\} \quad \text{a.s.}$$

For every Π with $\bar{\Pi} = P_0$ and every j, $\mathbf{1}_{\{K_j > 0\}} \le K_j$, so

$$\Pr(K_j > 0) \le \min\{1, \mathbb{E}K_j\} = \min\{1, Bp_j\}.$$

For the minimizing counts,

$$\Pr(K_j > 0) = \begin{cases} \mathbb{E}K_j = Bp_j, & Bp_j < 1, \\ 1, & Bp_j \ge 1. \end{cases}$$

Summing over j proves (24). Under independent sampling, $\Pr(K_j = 0) = (1-p_j)^B$. □

Corollary B.2. *In the setting of Example B.1, assign quality $h_j \geq 0$ to every response in class j. Put $w_i = 0.8^{i-1} / \sum_{k=1}^{B} 0.8^{k-1}$ for $1 \leq i \leq B$ and $w_{B+1} = 0$. Define*

$$\tau_j = \min\big(\{i : X_i = e_j\} \cup \{B+1\}\big), \qquad \text{utility}(X) = \sum_{j=1}^{C} h_j w_{\tau_j}.$$

Among uniformly ordered batches with $D(\Pi) = 0$, every maximizer of $\text{Div}(\Pi)$ also maximizes $\mathbb{E}_\Pi[\text{utility}(X)]$.

Intuitively, additional copies of a class give diminishing gains in its expected utility.

Proof of Corollary B.2. Uniform ordering gives $\mathbb{E}[w_{\tau_j} \mid K_j] = f_B(K_j)$, where, with $\binom{n}{t} = 0$ for $t > n$,

$$\begin{aligned} f_B(k) &= \sum_{t=1}^{B} (w_t - w_{t+1}) \left(1 - \frac{\binom{B-k}{t}}{\binom{B}{t}}\right), \quad 0 \leq k \leq B, \\ f_B(k+1) - f_B(k) &= \sum_{t=1}^{B} (w_t - w_{t+1}) \frac{\binom{B-k-1}{t-1}}{\binom{B}{t}}, \qquad 0 \leq k < B. \end{aligned} \tag{27}$$

The differences in (27) are nonnegative and nonincreasing in k. Thus f_B is discretely concave. With $a_j = \lfloor Bp_j \rfloor$ and $\theta_j = Bp_j - a_j$, concavity and $\mathbb{E}K_j = Bp_j$ give

$$\mathbb{E}_\Pi[\text{utility}(X)] = \sum_j h_j \mathbb{E} f_B(K_j) \leq \sum_j h_j \big[(1-\theta_j) f_B(a_j) + \theta_j f_B(a_j+1)\big].$$

By Example B.1, every diversity maximizer has $K_j \in \{a_j, a_j+1\}$ almost surely for every j and attains equality. □

In NoveltyBench, quality can vary within a class and class assignments can depend on the batch.

Binary correctness.

Example B.3. For binary correctness at $D = 0$, take $P_0 = (1-p)\delta_0 + p\delta_1$, $0 < p < 1$. For $X \sim \Pi$, put $K = \sum_i X_i$ and $\text{pass@}\, B(\Pi) = \Pr_\Pi(K \geq 1)$. Every diversity-maximizing distribution at zero distortion also attains

$$\max_{\Pi:\, D(\Pi)=0} \text{pass@}\, B(\Pi) = \min\{1, Bp\}, \tag{28}$$

whereas independent sampling gives $1 - (1-p)^B$.

Proof of Example B.3. Zero distortion gives $\mathbb{E}K = Bp$. By Lemma A.2,

$$\text{Div}(\Pi) = c_B \left(1 - \frac{\text{Var}(K)}{B^2 p(1-p)}\right). \tag{29}$$

The integer-variance bound (25) is attained by $K \in \{\lfloor Bp \rfloor, \lceil Bp \rceil\}$ with $\mathbb{E}K = Bp$. Uniformly ordering K ones and $B - K$ zeros gives each marginal P_0.

For every admissible Π, $\Pr(K > 0) \leq \min\{1, \mathbb{E}K\} = \min\{1, Bp\}$. A diversity maximizer attains equality: if $Bp < 1$, then $K \in \{0, 1\}$ and $\Pr(K > 0) = Bp$, while if $Bp \geq 1$, then $K \geq 1$ almost surely. Independent sampling gives $\Pr(K = 0) = (1-p)^B$. □

For one-hot final-answer features, every zero-distortion diversity maximizer attains the pass@B maximum $\min\{1, Bp\}$ if the correct responses form exactly one class of probability p.

C Implementation of Release and DAR

C.1 Background and implementation of Release

Computed probabilities are jointly measurable in the feature pool and any solver seed, which is independent of the responses. Subset draws and uniform orderings are conditional on the responses and seed.

At $\rho = 0$, Release also imposes

$$\sum_{J:\, i\in J} q(J) = B/N, \qquad i = 1, \dots, N, \tag{30}$$

so that every response in the pool is selected with probability B/N. By Proposition C.1, (30) does not lower the maximum in (8).

Proposition C.1. *For fixed $N \geq B$ and $\eta \geq 0$, $\mathrm{Div}_x^\star(\eta)$ admits a maximizer q_x such that $x \mapsto q_x(J)$ is Borel measurable for every $J \in \mathcal{S}_N$. At $\eta = 0$, a maximizer can satisfy (30).*

Proof of Proposition C.1. Put $q_J = q(J)$, $w_i(q) = B^{-1} \sum_{J:\, i\in J} q_J$, and $d_{ij} = \|x_i - x_j\|^2$. For $s_x > 0$, set

$$h_x(J) = \frac{\sum_{i<j,\ i,j\in J} d_{ij}}{s_x^2 B(B-1)}.$$

Set $h_x(J) = 0$ when $s_x = 0$. For any $\eta \geq 0$, the pool optimization is

$$\begin{aligned} \max_{q,\gamma} \quad & \sum_J q_J h_x(J) \\ \text{subject to} \quad & q_J \geq 0, \qquad \gamma_{ij} \geq 0, \qquad \sum_{J\in\mathcal{S}_N} q_J = 1, \\ & \sum_j \gamma_{ij} = w_i(q), \qquad \sum_i \gamma_{ij} = \frac{1}{N}, \\ & \sum_{i,j} \gamma_{ij} d_{ij} \leq \eta^2 s_x^2. \end{aligned} \tag{31}$$

Indeed, $\sum_{i,j} \gamma_{ij} \delta_{(x_i, x_j)}$ couples $Q = \bar{\Pi}_{x,q}$ and P_x. Conversely, any coupling κ of Q and P_x lifts to

$$\gamma_{ij} = \begin{cases} \kappa(\{x_i\} \times \{x_j\}) \dfrac{w_i(q)}{Q(\{x_i\})} \dfrac{1/N}{P_x(\{x_j\})}, & Q(\{x_i\}) > 0, \\ 0, & Q(\{x_i\}) = 0. \end{cases}$$

Summing over equal feature values gives

$$\sum_j \gamma_{ij} = w_i(q), \qquad \sum_i \gamma_{ij} = \frac{1}{N}, \qquad \sum_{i,j} \gamma_{ij} d_{ij} = \int \|u - v\|^2 \, d\kappa(u, v).$$

Thus (31) has value $\mathrm{Div}_x^\star(\eta)$. Its feasible set is a compact polytope, since all variables lie in $[0, 1]$ and the uniform subset distribution with $\gamma_{ij} = \mathbf{1}_{\{i=j\}}/N$ is feasible. At $\eta = 0$, Lemma A.4 permits the additional constraints $w_i(q) = 1/N$.

Enumerate the square systems of active constraints. Each nonsingular system defines a Borel candidate vertex by Cramer's rule. Feasibility and objective comparisons are Borel, so the first feasible maximizing vertex in the enumeration is a Borel optimizer. Such a vertex exists by compactness. When $s_x = 0$, use the uniform subset distribution and diagonal transport matrix.

For any Borel measurable feasible family $(q_x)_x$, fix $q = q_x$ in (31) and replace the objective by zero. The vertex enumeration gives a Borel measurable feasible transport matrix $\gamma(x)$. □

Column generation. Column generation (Desrosiers & Lübbecke, 2005) restricts (31) to $q(J) = 0$ for $J \notin \mathcal{R}$, where $\mathcal{R} \subseteq \mathcal{S}_N$. All transport variables are retained. Initialize with

$$\mathcal{R} = \{J_t : 0 \le t < N\}, \qquad J_t = \{1 + ((t+\ell) \bmod N) : 0 \le \ell < B\}.$$

Uniform t gives inclusion probabilities B/N, hence feasibility for every $\eta \ge 0$.

For $\eta > 0$, the restricted dual is

$$\begin{aligned} \min_{a,u,v,\lambda} \quad & a + \frac{1}{N}\sum_j v_j + \lambda\eta^2 s_x^2, \\ \text{subject to} \quad & a + \frac{1}{B}\sum_{i\in J} u_i \ge h_x(J) \quad (J \in \mathcal{R}), \\ & u_i - v_j \le \lambda\|x_i - x_j\|^2 \quad (1 \le i, j \le N), \qquad \lambda \ge 0, \end{aligned} \tag{32}$$

with $a, u_i, v_j \in \mathbb{R}$. At an optimal dual solution, pricing computes

$$r^* = \max_{J\in\mathcal{S}_N} \left\{ h_x(J) - a - \frac{1}{B}\sum_{i\in J} u_i \right\}. \tag{33}$$

If $r^* > 0$, add a maximizing J to $\mathcal{R}$ and re-solve. If $r^* \le 0$, the restricted solution is optimal. Pricing is a finite maximization, equivalently a quadratic program over $z \in \{0,1\}^N$ with $\sum_i z_i = B$ and $J = \{i : z_i = 1\}$.

At $\eta = 0$, impose (30) and omit γ. The restricted program and pricing problem are

$$\max_{q\ge 0} \left\{ \sum_{J\in\mathcal{R}} q(J)h_x(J) : \sum_{J\in\mathcal{R}:\, i\in J} q(J) = \frac{B}{N} \text{ for every } i \right\}, \qquad r^* = \max_{J\in\mathcal{S}_N} \left\{ h_x(J) - \sum_{i\in J} y_i \right\}, \tag{34}$$

where y_i are optimal dual variables. Summing the inclusion constraints gives $\sum_J q(J) = 1$.

For either program, let L be the restricted optimum and $\overline{r} \ge r^*$ a certified upper bound. Dual feasibility gives

$$L \le \mathrm{Div}_x^\star(\eta) \le L + \max\{0, \overline{r}\}. \tag{35}$$

For (32), increase a by $\max\{0,\overline{r}\}$. At $\eta = 0$, increase every y_i by $\max\{0,\overline{r}\}/B$. Both changes increase the dual objective by $\max\{0,\overline{r}\}$. Stop when the certified gap meets the tolerance or the computation limit is reached. Return a numerically verified feasible distribution and its available gap bound. A heuristic pricing search without an upper bound does not certify optimality. If feasibility cannot be verified, use the initial cyclic distribution.

Release with one-hot features. Let the pool have C nonempty classes of sizes n_j, with $x_i = e_j$ in class j. Set

$$p_j = \frac{n_j}{N}, \qquad K_j = |\{i \in S : x_i = e_j\}|, \qquad a_j = \mathbb{E}K_j, \qquad r_j = \frac{a_j}{B}.$$

Counting pairs in distinct classes and using the transport construction in Appendix B.2 gives

$$s_x^2 \mathrm{Div}_x(\Pi_{x,q}) = \frac{B^2 - \sum_j \mathbb{E}K_j^2}{B(B-1)}, \qquad W_2(\bar{\Pi}_{x,q}, P_x)^2 = \sum_j |r_j - p_j|. \tag{36}$$

For $a_j \in [\ell, \ell+1]$, (25) gives

$$\mathbb{E}K_j^2 \ge (2\ell+1)a_j - \ell(\ell+1),$$

with equality precisely when $K_j \in \{\lfloor a_j \rfloor, \lceil a_j \rceil\}$ almost surely. Thus the optimal expected counts solve

$$\begin{aligned} \min_{a,t,u} \quad & \sum_{j=1}^{C} t_j \\ \text{subject to} \quad & 0 \le a_j \le n_j, \qquad \sum_j a_j = B, \\ & t_j \ge (2\ell+1)a_j - \ell(\ell+1) \quad (1 \le j \le C,\ 0 \le \ell < n_j,\ \ell \in \mathbb{Z}), \\ & u_j \ge a_j/B - p_j, \qquad u_j \ge p_j - a_j/B, \qquad \sum_j u_j \le \eta^2 s_x^2. \end{aligned} \tag{37}$$

At $\eta = 0$, $a_j = Bp_j$.

Let a solve (37). Set $b_0 = 0$, $b_j = \sum_{h=1}^{j} a_h$, and, for $V \sim \mathrm{Unif}[0, 1)$, define

$$K_j = \big|\{k \in \{0, \dots, B-1\} : b_{j-1} \le V + k < b_j\}\big|.$$

As in Appendix B.3,

$$\sum_j K_j = B, \qquad \mathbb{E}K_j = a_j, \qquad K_j \in \{\lfloor a_j \rfloor, \lceil a_j \rceil\} \subseteq \{0, \dots, n_j\}.$$

Choose K_j indices uniformly without replacement from class j, then uniformly order all B selected indices. Each index in class j has inclusion probability a_j/n_j. The induced distribution q satisfies $\mathrm{Div}_x(\Pi_{x,q}) = \mathrm{Div}_x^\star(\eta)$ and, at $\eta = 0$, satisfies (30).

C.2 Proof of Proposition 3.1

Proof of Proposition 3.1. Theorem 2.5(a) gives $D(\Pi_N) \le \rho$. Independent solver seeds are handled by conditioning and convexity under mixtures. Put

$$V_N = \mathrm{Div}^\star_{x_N}(\eta_N(\rho)).$$

Theorem 2.5(b) gives $V_N \to \mathrm{Div}^\star_E(\rho)$ almost surely. Applying Proposition 2.2 with P_{x_N} in place of P_0 yields

$$0 \le V_N \le c_B(1 + \sqrt{2}\rho)^2.$$

The bound also holds when $s_{x_N} = 0$. Moreover, $s^2_{x_N} \to s_0^2$ almost surely and in L^1, since $\mathbb{E}s^2_{x_N} = (1 - 1/N)s_0^2$. Consequently,

$$\frac{\mathbb{E}[s^2_{x_N} V_N]}{s_0^2} \longrightarrow \mathrm{Div}^\star_E(\rho).$$

Linearity of the pairwise-distance objective gives

$$0 \le \frac{\mathbb{E}[s^2_{x_N} V_N]}{s_0^2} - \mathrm{Div}(\Pi_N) \le \frac{\varepsilon_N \mathbb{E}s^2_{x_N}}{s_0^2} = (1 - 1/N)\varepsilon_N.$$

Let $N \to \infty$. □

C.3 Relation to survey sampling

Proposition C.2. *Let $x \in (\mathbb{R}^d)^N$ and let q be a probability mass function on $\mathcal{S}_N$ satisfying (30). For $X \sim \Pi_{x,q}$,*

$$\mathbb{E}\bar{X} = m_{P_x}, \qquad s_x^2 \mathrm{Div}_x(\Pi_{x,q}) = c_B\left(s_x^2 - \mathrm{Tr}\,\mathrm{Cov}(\bar{X})\right). \tag{38}$$

Proof of Proposition C.2. Uniform ordering and (30) give

$$\mathcal{L}(X_j) = P_x, \qquad \mathbb{E}\bar{X} = m_{P_x}, \qquad \frac{1}{B}\sum_j \mathbb{E}\|X_j - m_{P_x}\|^2 = s_x^2.$$

Apply (12) and $\mathrm{Tr}\,\mathrm{Cov}(\bar{X}) = \mathbb{E}\|\bar{X} - m_{P_x}\|^2$. □

Thus maximizing $\mathrm{Div}_x(\Pi_{x,q})$ under (30) minimizes the objective defining $M_B(P_x)$ over distributions $\Pi_{x,q}$.

C.4 Implementation of DAR

Write $q_{T,x}$ for the probability mass function computed by DAR. Then

$$\Pi_T = \mathbb{E}[\Pi_{x_{m+N_T}, q_{T,x_{m+N_T}}}].$$

The measurability and sampling conventions of Appendix C.1 apply to $q_{T,x}$.

At $\rho = 0$, DAR also imposes

$$\sum_{J:\, i\in J} q(J) = \begin{cases} B/T, & 1 \le i \le m, \\ B(T-m)/(TN_T), & m < i \le m + N_T, \end{cases} \tag{39}$$

so that pilot responses have the same inclusion probabilities as in Release with T responses. At $\rho = 0$, ε_T bounds the objective gap in (10) under (39).

Fix T, and write $x^+ = x^+_{N_T}$. Set

$$\alpha_i = \begin{cases} 1/T, & i \le m, \\ (T-m)/(TN_T), & i > m. \end{cases}$$

The linear program is (31) on $\mathcal{S}_{m+N_T}$ with target column sums α_j and transport bound $\eta_{N_T}(\rho)^2 s^2_{x^+}$:

$$\begin{aligned} \max_{q,\gamma} \quad & \sum_J q_J h_x(J) \\ \text{subject to} \quad & q_J \ge 0, \quad \gamma_{ij} \ge 0, \quad \sum_J q_J = 1, \\ & \sum_j \gamma_{ij} = B^{-1} \sum_{J:\, i\in J} q_J, \quad \sum_i \gamma_{ij} = \alpha_j, \\ & \sum_{i,j} \gamma_{ij} \|x_i - x_j\|^2 \le \eta_{N_T}(\rho)^2 s^2_{x^+}. \end{aligned} \tag{40}$$

The objective uses all pairs, including pairs involving pilot responses. The variance in the constraint uses only the additional responses.

C.5 Proof of Proposition 3.2

Proof of Proposition 3.2. Put $b = mB/T$. Since $T \ge m + B$, $0 < b < \min(m, B)$. Choose an integer $K \in \{\lfloor b \rfloor, \lceil b \rceil\}$ with $\mathbb{E}K = b$, independently of the responses. Given K, select a uniform K-subset of the pilot and a uniform $(B-K)$-subset of the additional indices. The inclusion probabilities are $B\alpha_i$, so the construction is feasible for every ρ. The construction provides an initialization, not a restriction on the optimized pilot count. The compact-polytope and vertex arguments in Appendix C.1 give measurable maximizers and feasible transport matrices for (40), also with (39) at zero. There are finitely many possible values of N_T for each T.

Conditional on the pilot, the additional features are independent with distribution P_0. Hence

$$\mathbb{E}R_T = P_0, \qquad \mathbb{E}[s^2_{x^+} \mid Y_1, \dots, Y_m] = (1 - 1/N_T)s_0^2.$$

Joint convexity under mixtures gives

$$W_2(\bar{\Pi}_T, P_0)^2 \le \mathbb{E}\big[W_2(\bar{\Pi}_{x,q_{T,x}}, R_T)^2\big] \le \mathbb{E}[\eta_{N_T}(\rho)^2 s^2_{x^+}] = \rho^2 s_0^2.$$

Each output has finite second moment because at most T responses are generated. Thus $D(\Pi_T) \le \rho$ and, since $\Pi_T(E^B) = 1$,

$$\mathrm{Div}(\Pi_T) \le \mathrm{Div}^\star_E(\rho).$$

For a pool of n independent features, put

$$f(n) = \frac{\mathbb{E}[s^2_{x_n} \mathrm{Div}^\star_{x_n}(\eta_n(\rho))]}{s_0^2}.$$

Proposition 3.1, applied to exact maximizers, gives $f(n) \to \mathrm{Div}^\star_E(\rho)$ and $0 \le f(n) \le \mathrm{Div}^\star_E(\rho)$. For $T \ge mB$, construct a comparison distribution on the observed indices as follows. With probability $1 - b$, use an exact Release maximizer on the additional pool. With probability b, choose one uniform pilot index and $B - 1$ uniform distinct additional indices. Order the selected indices uniformly in either case. Denote the feature distribution from the first branch by Π^+ and the mixture by $\widetilde{\Pi}_x$. Then

$$\widetilde{\bar{\Pi}}_x = (1-b)\bar{\Pi}^+ + \frac{b}{B}P_{x_m} + \frac{b(B-1)}{B}P_{x^+}, \qquad R_T = (1-b)P_{x^+} + \frac{b}{B}P_{x_m} + \frac{b(B-1)}{B}P_{x^+}.$$

Therefore

$$W_2(\bar{\bar{\Pi}}_x, R_T)^2 \le (1-b) W_2(\bar{\Pi}^+, P_{x^+})^2 \le \eta_{N_T}(\rho)^2 s_{x^+}^2.$$

At $\rho = 0$, choose the Release maximizer with equal inclusion probabilities. The mixture then satisfies (39). Nonnegativity of pairwise squared distances and the objective-gap assumption yield

$$\mathrm{Div}(\Pi_T) \ge (1-b)\mathbb{E} f(N_T) - \varepsilon_T \frac{\mathbb{E} s_x^2}{s_0^2}.$$

The expectation of $f(N_T)$ is valid because N_T depends only on the pilot. Moreover,

$$\begin{aligned}\mathbb{E} s_x^2 &\le \mathbb{E}\left[\frac{\sum_{i=1}^{m+N_T} \|\xi_i - m_0\|^2}{m+N_T}\right] \\ &\le \frac{m}{m+B} s_0^2 + s_0^2 \le 2 s_0^2.\end{aligned}$$

Thus

$$(1 - mB/T)\mathbb{E} f(N_T) - 2\varepsilon_T \le \mathrm{Div}(\Pi_T) \le \mathrm{Div}_E^\star(\rho).$$

For fixed m, $N_T \to \infty$ almost surely gives $\mathbb{E} f(N_T) \to \mathrm{Div}_E^\star(\rho)$ by bounded convergence. Let $T \to \infty$ and $\varepsilon_T \to 0$. Independent solver seeds are handled by conditioning and averaging. □

The distribution R_T in (9) is generally not the uniform empirical distribution of all generated responses. A design that selects every generated response with the same probability therefore need not satisfy (39). The first construction in the proof supplies a feasible fallback. A fixed lower endpoint in the additional-sample grid does not by itself ensure $N_T \to \infty$. A grid whose smallest value tends to infinity suffices.

C.6 Distribution of each returned response

Proposition C.3. *Let $\rho = 0$, and let $(Y_{I_1}, \dots, Y_{I_B})$ be the batch returned by Algorithm 1 or Algorithm 2. Then $\mathcal{L}(Y_{I_j}) = \mathcal{L}(Y)$ for every $j \in \{1, \dots, B\}$.*

Proof of Proposition C.3. Fix $j \in \{1, \dots, B\}$ and $A \in \mathcal{A}$. Let $\mathcal{F}$ be the σ-algebra generated by $Y_1, Y_2, \dots$ and the solver seed. Uniform ordering gives $\Pr(I_j = i \mid \mathcal{F}) = B^{-1} \sum_{J:\, i \in J} \hat{q}(J)$ for every candidate index i.

For Algorithm 1, (30) gives

$$\Pr(Y_{I_j} \in A \mid \mathcal{F}) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{Y_i \in A\}}.$$

Each Y_i has distribution $\mathcal{L}(Y)$, so taking expectations gives $\Pr(Y_{I_j} \in A) = \Pr(Y \in A)$.

For Algorithm 2, (39) gives

$$\Pr(Y_{I_j} \in A \mid \mathcal{F}) = \frac{1}{T} \sum_{i=1}^{m} \mathbf{1}_{\{Y_i \in A\}} + \frac{T-m}{T N_T} \sum_{i=m+1}^{m+N_T} \mathbf{1}_{\{Y_i \in A\}}.$$

Since $N_T = g_{T,\rho}(Y_1, \dots, Y_m)$ and $Y_{m+1}, Y_{m+2}, \dots$ are independent of $(Y_1, \dots, Y_m)$,

$$\mathbb{E}\Big[\frac{1}{N_T} \sum_{i=m+1}^{m+N_T} \mathbf{1}_{\{Y_i \in A\}} \;\Big|\; Y_1, \dots, Y_m\Big] = \Pr(Y \in A).$$

Taking expectations gives

$$\Pr(Y_{I_j} \in A) = \frac{m}{T} \Pr(Y \in A) + \frac{T-m}{T} \Pr(Y \in A) = \Pr(Y \in A).$$

Since $A \in \mathcal{A}$ is arbitrary, $\mathcal{L}(Y_{I_j}) = \mathcal{L}(Y)$ in both cases. □

At zero distortion, Release and DAR keep the distribution of each returned response and change only the dependence among the returned responses. Consequently, $\mathbb{E} f(Y_{I_j}) = \mathbb{E} f(Y)$ for every measurable $f : \mathcal{Y} \to \mathbb{R}$ with $\mathbb{E}|f(Y)| < \infty$, such as a quality score or the indicator of a correct final answer.

D EXPERIMENTAL DETAILS

D.1 EXPERIMENTAL SETUP

Open-ended prompts are drawn from NoveltyBench curated and WildChat subsets (Zhang et al., 2025; Zhao et al., 2024), Infinity-Chat (Jiang et al., 2025), and WritingPrompts (Fan et al., 2018). Both models use the same 100 calibration and 100 held-out prompt identities, with source counts $20/40/20/20$ in each split. Each mathematical dataset has 100 calibration and 100 held-out problems, giving 200 per split. Calibration and held-out identities are disjoint.

Open-ended pools contain 512 generated responses per prompt. Mathematical pools contain 512 for Llama, and retain the first 320 held-out or 160 calibration records for Qwen before candidate filtering. Responses with no answer text or reaching the length cap are excluded, preserving generation order. The empirical distribution P_0 concerns usable answers under the base sampling configuration. Downstream Release, MMR, and k-DPP use the first 64 usable responses. Other pool sizes are stated explicitly. Distortion evaluation uses the last 128 usable responses. One Qwen mathematical prompt has only 252 usable responses, so a selection prefix of 128 overlaps the comparison set by four. For that prompt, distortion and accuracy changes relative to the comparison set are omitted only when the actual selection prefix overlaps. Task scores remain included.

Batch sizes are $B = 4$ for NoveltyBench, $B = 8$ for mathematical reasoning, and $B = 16$ for GRPO. A set of 16 outputs contains four, two, or one batches, respectively. MMR returns one batch per prompt and parameter setting. Repetition counts are specified below. No missing response is replaced by a newly generated response.

Generation settings. Llama uses temperature 1.0 and top-$p = 1.0$, without top-k truncation. Qwen uses temperature 0.6, top-$p = 0.95$, and top-$k = 20$ in thinking mode. The context limit is 32768 tokens, with the generation allowance set to the remaining context capacity minus one token. We use each model's chat template. For Qwen, the response text is the nonempty answer after the final `</think>` delimiter. Generated thinking tokens still count toward cost.

Features. The pool sweep and token comparison use `bge-base-en-v1.5` embeddings (Xiao et al., 2024). Text exceeding 512 wordpieces is split into windows, whose embeddings are averaged. NoveltyBench selection uses classes constructed in generation order with the classifier and matching rule specified in Appendix D.7. If response i belongs to class j of the observed pool, set $x_i = e_j$. Mathematical selection uses the same one-hot representation for the normalized final answer extracted from `\boxed{}` or `\fbox{}`, without correctness labels. Extraction failures form one class. MMR and k-DPP retain their embedding-based selection objectives on both tasks.

D.2 RELEASE

The embedding experiments use column generation with exact pricing for $B = 4$. The sweep uses $N \in \{16, 24, 32, 48, 64, 96, 128, 192, 256\}$ and $\rho \in \{0, 0.5, 1, 2\}$. All 72 model–size–distortion combinations use the same 100 prompts per model. The maximum certified gap is below 10^{-6} in reported diversity units, and feasibility residuals are below 10^{-7}. A gap g in Div_x units is converted to reporting units as $g s_x^2/\widehat{s}_0^2$. Release uses (31) with $\eta = \eta_N(\rho)$. Class features use Appendix C.1.

D.3 EMPIRICAL DIVERSITY AND DISTORTION

For a prompt with $n \geq 2$ usable responses and features $x_1, \ldots, x_n$, the reporting variance is

$$\widehat{s}_0^2 = \frac{1}{n-1} \sum_{i=1}^{n} \|x_i - \bar{x}\|^2, \qquad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

The full-pool estimate is used only for reporting. The Release selection constraint uses the observed pool variance. Figure 1, the Release rows of Table 1, Table 2, and the diversity rows of Release and DAR in Table 32 report $s_x^2 \mathrm{Div}_x(\Pi_{x,q})/\widehat{s}_0^2$, where x contains all responses used in selection. Other diversity entries, including the independent-sampling row of Table 1, use returned batches as

follows. When $\widehat{s}_0^2 > 0$, report mean batch diversity as

$$\widehat{\mathrm{Div}} = \frac{1}{R}\sum_{r=1}^{R}\frac{\sum_{i<j}\|X_{ri} - X_{rj}\|^2}{\widehat{s}_0^2 B(B-1)},$$

where R is the number of complete batches and X_{ri} is the feature vector at position i in batch r. For open-ended batches, form $\widehat{P}_a$ from the feature vectors in complete batches with equal weights. For mathematical batches, $\widehat{P}_a$ includes every returned response, including responses from incomplete batches, and omits unfilled positions. Let $\widehat{P}_{\mathrm{cmp}}$ be the empirical feature distribution of a disjoint comparison set of 128 responses. We report

$$\widehat{D} = \frac{W_2(\widehat{P}_a, \widehat{P}_{\mathrm{cmp}})}{\widehat{s}_0}.$$

Prompts with numerically zero full-pool feature variance are omitted from $\widehat{\mathrm{Div}}$ and $\widehat{D}$, and the number omitted is reported. Such prompts remain in pass, majority, and GRPO evaluation. Among the 200 mathematical problems, reporting variance is zero on no Llama prompts and on 170 Qwen prompts, comprising 88 GSM8K and 82 MATH-500 problems. Of the 170 single-class Qwen pools, 168 contain only the same correct final answer and two contain only the same incorrect final answer. Qwen diversity and distortion therefore describe 30 prompts, with one fewer for distortion when the maximal selection prefix overlaps the comparison set. Open-ended reporting variances are nonzero on all 100 prompts per model. Metric counts can be smaller when batches or the comparison set are unavailable. The cutoff for mathematical reporting is $\widehat{s}_0^2 > 10^{-12}$.

Transport is solved exactly. For one-hot features, $W_2^2 = \sum_j |\widehat{r}_j - \widehat{p}_j|$. Each eligible prompt receives equal weight in reported means. Standard errors are computed across prompt-level scores. Paired differences use common prompts. Intervals for paired differences draw 2,000 resamples of prompts with replacement, stratified by source. Confidence intervals are descriptive and are not adjusted for multiple comparisons. Repeated batches and Monte Carlo draws are averaged within prompts. Sixteen outputs per prompt give substantial nonzero distortion even for independent sampling. MMR supplies only B outputs, so its distortion requires independent sampling with the same number of outputs for comparison. The distortion of independent sampling is not subtracted from measured distortion.

D.4 DAR

We use $m = 8$ and $T \in \{64, 128\}$. The additional-count grids are

$$\mathcal{G}_{64} = \{16, 24, 32, 48, 56\}, \qquad \mathcal{G}_{128} = \{16, 24, 32, 48, 56, 64, 96, 120\}.$$

All $m + N_T$ responses remain candidates. Pilot mass is m/T, and the squared transport limit is $\rho^2 N_T s^2_{x^+_{N_T}}/(N_T - 1)$. At zero distortion, the solver enforces (39) at the index level, including repeated features.

For each model, feature choice, B, T, ρ, calibration is split $80/20$ within each source. At each $n \in \mathcal{G}_T$, ridge regressions predict the design value $\mathrm{Div}_{x_{m+n}}$ and total generation tokens. Predictors use only the pilot. Embedding predictors are its biased variance, maximum, minimum, and median pairwise squared distance, and a near-zero-variance indicator. Class predictors are the class fraction, singleton fraction, largest-class fraction, biased variance, and one-class indicator. Both add $\log(1 + z)$ transforms of the mean and standard deviation of usable-response lengths and pilot token cost, together with the rejected-attempt fraction. An intercept is included. Features are standardized by training means and population standard deviations, replacing scales below 10^{-10} by one. The intercept is not centered or penalized, and the ridge penalty is one. A zero-variance union has calibration target zero.

Writing the two predictions as $\widehat{d}_n$ and $\widehat{c}_n$, the policy chooses

$$N_T = \operatorname*{argmax}_{n \in \mathcal{G}_T}\left\{\widehat{d}_n - \alpha\frac{\max\{0, \widehat{c}_n\}}{c_{\mathrm{train}}}\right\},$$

where c_{train} is the median training cost of Release with $N = T$, and ties use the smaller n. The coefficient is selected from

$$\alpha \in \{0, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5\}$$

to minimize mean validation cost subject to a mean design-value loss of at most 0.01 relative to Release at $N = T$. Ties use smaller loss, then smaller α. If no coefficient satisfies the loss criterion, minimize loss, then cost, then α. The latter rule was used for 30 of 112 policies. The coefficient α controls the cost penalty and is distinct from the pilot mass m/T. The criterion is a calibration target, not a held-out performance guarantee. Policies are frozen before evaluation. Quality scores and correctness labels are excluded from predictors and policy selection. Union variance normalizes calibration design values, while full-pool variance is reserved for reporting.

D.5 Generation cost of DAR

The diversity, distortion, and token entries in Table 2 are means with standard errors across prompts. For paired token counts t_i and Release counts t_i^{Release}, the saving is $100(1-\bar{t}/\bar{t}^{\text{Release}})$. Its standard error is estimated by the delta method as

$$\frac{100}{\bar{t}^{\text{Release}}\sqrt{n}}\,\mathrm{sd}\left(t_i - \frac{\bar{t}}{\bar{t}^{\text{Release}}}t_i^{\text{Release}}\right),$$

where sd uses denominator $n-1$.

For method a and evaluated prompt u, let $t_{a,u}^{\text{think}}$ and $t_{a,u}^{\text{answer}}$ count all tokens generated in the thinking and answer stages, including pilot responses and responses not returned. Set

$$\bar{t}_a = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\left(t_{a,u}^{\text{think}} + t_{a,u}^{\text{answer}}\right), \qquad \text{Saving}_a = 100\left(1 - \frac{\bar{t}_a}{\bar{t}_{\text{Release},N}}\right)\%.$$

Methods are compared on the same prompt set $\mathcal{U}$. We report savings separately for $N = 64$ and $N = 128$. All generated records through the last used candidate are charged, including rejected or capped generations. Table 5 splits the generated tokens in Table 2 into thinking and answer tokens, and Table 6 reports the generation cost of calibration.

Model	Method	Thinking	Answer	Total
Llama	Release, $N = 64$	0.0±0.0	50.9±5.3	50.9±5.3
Llama	Release, $N = 128$	0.0±0.0	102.6±10.5	102.6±10.5
Llama	DAR, $T = 64$	0.0±0.0	19.4±2.0	19.4±2.0
Llama	DAR, $T = 128$	0.0±0.0	39.8±4.9	39.8±4.9
Qwen	Release, $N = 64$	53.4±3.0	32.1±2.3	85.5±4.0
Qwen	Release, $N = 128$	107.6±6.4	65.0±4.5	172.6±8.3
Qwen	DAR, $T = 64$	20.6±1.2	12.1±0.8	32.6±1.5
Qwen	DAR, $T = 128$	21.3±1.1	12.4±0.8	33.8±1.5

Table 5: Generated tokens per prompt in thousands, mean ± SE, for Table 2. All recorded attempts through the selected usable prefix are charged.

Model	Task	$T = 64$	$T = 128$	Shared total
Llama	Open	5.24	10.60	10.60
Llama	Math	11.23	22.70	22.70
Qwen	Open	11.56	23.29	23.29
Qwen	Math	46.37	93.49	93.49

Table 6: Calibration generation cost in millions of tokens. The same cached prefixes are reused across features, batch sizes, and distortion settings. Shared total counts the 128-response prefixes once, excluding held-out evaluation and comparison sets.

For Qwen baselines with precomputed thinking, charge each generated block once per use: one per output for Arithmetic and one per batch for stochastic beam search. Blocks reused across batch sizes are charged in each evaluated method. Recorded failed requests and retries also count.

D.6 BASELINES

Table 7 lists the configuration of each baseline.

Method	Configuration
In-context regeneration	NoveltyBench conversation protocol. Keep prior answers and request "Can you generate a different answer?" after each turn, with at most 10 retries per conversation. Qwen history omits thinking.
Arithmetic sampling	Codes $v_i = (i/(B+1) + V) \bmod 1, i = 1, \ldots, B$, with shared uniform V. Shuffle vocabulary order jointly at each step. Use float64 codes and the base sampling parameters.
Stochastic beam search	Gumbel top-B sampling without replacement from the sequence distribution after the base temperature and truncation operations. Beam size B.
Verbalized sampling	VS-Standard creativity and mathematical prompts, requesting $k = 4$ responses with explicit probabilities per call. Combine $B/4$ independent calls for each batch. Free decoding with LaTeX-preserving parsing, up to five retries for incomplete lists. Reported probabilities are not evaluation weights.
MMR	Greedy maximization of $\lambda \cos(x_i, z) - (1-\lambda) \max_{j \in S} \cos(x_i, x_j)$, where z embeds the prompt. The redundancy term is zero when S is empty. Use 64 candidates and $\lambda \in \{0.3, 0.5, 0.7\}$. Appendix E.5 adds $\lambda \in \{0.1, 0.9\}$ and 128 candidates.
k-DPP	Exact fixed-size sampling with $k = B$ and $L = ZZ^\top$ for unit-normalized candidate embeddings. No quality weights. Use the same 64 candidates.

Table 7: Configurations for the six main baselines. Sources appear in Section 4.3.

Arithmetic and stochastic beam search adapt the authors' algorithms to a Hugging Face decoding loop. For Qwen, Arithmetic uses an independent thinking block per output, and stochastic beam search uses a shared block per batch. A rank-deficient k-DPP kernel uses a diagonal ridge of 10^{-8} times its scale.

NoveltyBench pool batches are uniformly ordered, and generated sequences retain their output order.

D.7 DOWNSTREAM EVALUATION

Open-ended generation. For a batch of $B = 4$ responses in output order, let ℓ_i be the class assigned to the ith response by batch-level DeBERTa grouping and h_i its quality score. The reported scores are

$$\text{distinct} = |\{\ell_1, \ldots, \ell_B\}|, \qquad \text{utility} = \frac{\sum_{i=1}^{B} 0.8^{i-1} h_i \mathbf{1}_{\{\ell_i \notin \{\ell_1, \ldots, \ell_{i-1}\}\}}}{\sum_{i=1}^{B} 0.8^{i-1}}.$$

We use metric version 1.0 from the official NoveltyBench repository at commit `64a23be`. Responses join the first matching class in output order. Matching compares the new response with the class's first response, using exact string equality or class-1 probability above 0.102 from the DeBERTaV3 classifier `yimingzhang/deberta-v3-large-generation-similarity` (He et al., 2023). The classifier uses the first 128 tokens of each response. Quality scores use `Skywork/Skywork-Reward-Gemma-2-27B-v0.2` (Liu et al., 2024), with the published nine thresholds mapping rewards to integer scores from 1 to 10. The evaluation uses $k = B = 4$, rather than the leaderboard's 10 responses, so scores are not directly comparable with leaderboard values. Prompt scores average complete batches. Incomplete batches are excluded. Among planned open-ended baseline batches, five are incomplete for Llama verbalized sampling, five for Qwen Arithmetic, and fifteen for Qwen stochastic beam search. Every prompt retains at least one complete batch. Other open-ended baseline batches are complete. Independent sampling in Figure 2 and Table 22 uses four uniformly sampled batches from the 64-response pool. The batches are drawn with seed zero. Release and DAR have four complete batches per prompt.

Selection groups the observed pool, whereas evaluation groups each returned batch in output order. Nontransitive classifier decisions can produce different partitions. Distinct and utility evaluate the same returned batch sets after scoring is complete. Selection features exclude reward scores. Equation (24) applies to the fixed selection partition.

Mathematical reasoning. Set $B = 8$ for mathematical reasoning. Pass@B is the indicator that at least one response is correct. For majority@B, count extracted final answers and omit <none>. Score one only when the most frequent answer is unique and correct. Ties and batches with no extracted answers score zero.

For a pool containing N responses and K correct responses, uniform selection without replacement has

$$\text{pass@}B = 1 - \frac{\binom{N-K}{B}}{\binom{N}{B}}.$$

Independent sampling selects uniformly without replacement from the first 64 independently generated responses. Tables use the Monte Carlo evaluation of independent sampling at $\rho = 0$ for both models. Fixed-size Release and independent sampling average 50 repeated sets of 16 outputs. DAR also averages 50 sets of 16 outputs. Each set contributes a separate distortion estimate before prompt-level averaging, so distortion is never computed from 800 pooled outputs. All repetition averages are formed within each prompt before computing SE. The displayed formula gives the exact conditional pass probability. For majority@B under independent sampling, let n_j be the class counts, including the extraction-failure class. For nonnegative integer counts $k_j \le n_j$, the count vector k has probability

$$\Pr(k) = \frac{\prod_j \binom{n_j}{k_j}}{\binom{N}{B}}, \qquad \sum_j k_j = B.$$

Sum over vectors with a unique most frequent extracted class that is correct, ignoring the failure-class count when determining the most frequent answer. Average prompt-level scores over held-out prompts. Incomplete batches are not filled or replaced. For pass and GRPO group statistics, the requested B positions are retained and missing positions have reward zero. Majority ignores missing positions and extraction failures. Diversity uses complete batches only. Distortion and response accuracy use all returned responses, omitting unfilled positions. Undefined normalized metrics remain missing. At $B = 8$, Llama regeneration, Arithmetic, and stochastic beam each have one incomplete batch. At $B = 16$, Llama regeneration has five and Arithmetic has one. Qwen Arithmetic has two incomplete batches at each of $B = 8$ and $B = 16$, and Qwen stochastic beam has one at $B = 8$. The handling of incomplete batches in mathematical evaluation differs from the complete-batch NoveltyBench evaluation.

GRPO group statistics. For $B = 16$ binary correctness rewards $r_1, \dots, r_B$, set $\bar{r} = B^{-1} \sum_i r_i$. Report the prompt-averaged constant-reward indicator and variance

$$\mathbf{1}_{\{r_1 = \cdots = r_B\}}, \qquad \frac{1}{B} \sum_{i=1}^{B} (r_i - \bar{r})^2.$$

Correctness labels do not enter selection. Pools with constant correctness rewards are retained.

Table 4 reports mean response accuracy for $B = 16$. For the same group size, Table 38 reports mean response accuracy minus the correctness rate of the comparison set. The population difference is zero for a distribution-preserving method, while the finite-sample difference need not be zero.

E Additional Experiments

E.1 Diversity as the pool grows

Table 8 reports all 72 certified pool-sweep entries. Each model uses the same 100 prompts at every pool size and distortion level. The $\rho = 2$ curves continue to increase through $N = 256$.

Model	ρ	16	24	32	48	64	96	128	192	256
Llama	0	1.1009	1.1170	1.1306	1.1469	1.1562	1.1657	1.1733	1.1783	1.1805
Llama	0.5	1.2520	1.2791	1.3020	1.3312	1.3484	1.3701	1.3847	1.3953	1.4007
Llama	1	1.4747	1.5418	1.5889	1.6510	1.6954	1.7504	1.7799	1.8074	1.8257
Llama	2	1.5150	1.6285	1.7069	1.8281	1.9299	2.0526	2.1278	2.2043	2.2701
Qwen	0	1.0862	1.1090	1.1201	1.1296	1.1398	1.1520	1.1584	1.1700	1.1771
Qwen	0.5	1.2151	1.2513	1.2679	1.2869	1.3048	1.3262	1.3381	1.3568	1.3697
Qwen	1	1.4344	1.5092	1.5461	1.5943	1.6394	1.6895	1.7145	1.7561	1.7863
Qwen	2	1.4796	1.6076	1.6824	1.7831	1.8685	1.9965	2.0625	2.1730	2.2549

Table 8: Mean design diversity at $B = 4$, normalized by full-pool reporting variance.

E.2 Generation cost of DAR

Tables 9 and 10 compare mathematical generation costs at matched response limits. Tables 11–14 give paired differences in diversity, distortion, and task scores. The comparisons do not establish noninferiority. At $\rho = 2$, Llama DAR with $T = 128$ has a constant-reward fraction 14.3 percentage points above Release at $N = 128$, with paired 95% interval $[9.3, 19.4]$, despite similar response accuracy. The DAR fraction is 24.40% under 50-repeat evaluation and 24.44% under exact conditional calculation. All-incorrect and all-correct groups contribute 23.26% and 1.14%, respectively. Approximate agreement in diversity does not imply agreement in reward-group statistics.

Model	T	ρ	Mean N_T	Tokens, 10^3	Saving, %
Llama	64	0	40.0	37.0 ± 2.3	43.4 [39.0, 47.2]
Llama	64	1	35.5	32.4 ± 2.0	50.4 [47.1, 53.2]
Llama	64	2	33.0	30.1 ± 1.9	53.9 [50.9, 56.5]
Llama	128	0	26.0	26.9 ± 2.0	79.5 [78.3, 80.6]
Llama	128	1	41.6	33.9 ± 2.0	74.2 [72.3, 75.9]
Llama	128	2	54.0	40.5 ± 2.2	69.2 [66.4, 71.5]
Qwen	64	0	27.2	140.4 ± 9.8	43.5 [40.4, 46.5]
Qwen	64	1	27.0	139.7 ± 9.8	43.8 [40.6, 46.9]
Qwen	64	2	44.8	224.6 ± 15.8	9.7 [6.5, 13.5]
Qwen	128	0	28.0	142.6 ± 10.0	71.4 [69.6, 73.1]
Qwen	128	1	64.9	294.8 ± 20.5	40.9 [35.0, 46.5]
Qwen	128	2	71.9	320.2 ± 22.1	35.8 [30.9, 40.5]

Table 9: DAR generation cost for $B = 8$ on 200 mathematical prompts. Savings compare Release at $N = T$. Brackets give 95% paired intervals from resampling prompts. N_T excludes the eight pilot responses.

Model	T	ρ	Mean N_T	Tokens, 10^3	Saving, %
Llama	64	0	37.8	34.7 ± 2.1	46.9 [43.0, 50.1]
Llama	64	1	39.9	36.0 ± 2.2	44.8 [40.9, 48.0]
Llama	64	2	50.3	49.6 ± 3.4	24.0 [19.2, 28.2]
Llama	128	0	44.9	36.5 ± 2.1	72.2 [69.9, 74.2]
Llama	128	1	54.8	45.2 ± 2.7	65.6 [62.9, 67.9]
Llama	128	2	93.8	73.6 ± 4.4	44.0 [37.7, 49.7]
Qwen	64	0	27.0	140.1 ± 9.8	43.7 [40.4, 46.7]
Qwen	64	1	28.2	146.0 ± 10.2	41.3 [38.5, 44.2]
Qwen	64	2	33.7	189.2 ± 15.3	23.9 [19.2, 29.0]
Qwen	128	0	28.1	142.7 ± 10.0	71.4 [69.6, 73.1]
Qwen	128	1	67.9	309.1 ± 22.0	38.0 [33.0, 42.9]
Qwen	128	2	72.7	350.7 ± 25.5	29.7 [25.4, 34.3]

Table 10: DAR generation cost for $B = 16$ on 200 mathematical prompts. Savings compare Release at $N = T$. Brackets give 95% paired intervals from resampling prompts. N_T excludes the eight pilot responses.

Model	T	ρ	$\Delta\widehat{\mathrm{Div}}$	$\Delta\widehat{D}$
Llama	64	0	-0.039 $[-0.059, -0.020]$	$+0.014$ $[0.004, 0.025]$
Llama	64	1	-0.019 $[-0.033, -0.008]$	$+0.001$ $[-0.009, 0.011]$
Llama	64	2	-0.005 $[-0.011, -0.001]$	-0.002 $[-0.003, -0.000]$
Llama	128	0	-0.026 $[-0.042, -0.009]$	$+0.013$ $[0.003, 0.022]$
Llama	128	1	-0.005 $[-0.010, -0.001]$	-0.004 $[-0.016, 0.008]$
Llama	128	2	-0.007 $[-0.013, -0.002]$	-0.001 $[-0.006, 0.004]$
Qwen	64	0	-0.037 $[-0.050, -0.024]$	$+0.031$ $[0.023, 0.039]$
Qwen	64	1	-0.013 $[-0.024, -0.005]$	-0.001 $[-0.009, 0.008]$
Qwen	64	2	-0.036 $[-0.057, -0.018]$	-0.012 $[-0.019, -0.005]$
Qwen	128	0	-0.061 $[-0.080, -0.043]$	$+0.045$ $[0.034, 0.056]$
Qwen	128	1	-0.002 $[-0.004, 0.000]$	$+0.005$ $[-0.004, 0.015]$
Qwen	128	2	-0.000 $[-0.001, 0.000]$	$+0.000$ $[-0.000, 0.000]$

Table 11: Embedding measurements. DAR minus Release at $N = T$, on common prompts. Brackets are descriptive 95% paired intervals from resampling prompts.

Model	T	ρ	Δ distinct	Δ utility
Llama	64	0	-0.045 $[-0.140, 0.048]$	-0.124 $[-0.308, 0.060]$
Llama	64	1	$+0.122$ $[0.045, 0.195]$	-0.055 $[-0.269, 0.143]$
Llama	64	2	$+0.037$ $[-0.055, 0.128]$	-0.015 $[-0.265, 0.233]$
Llama	128	0	-0.035 $[-0.128, 0.052]$	-0.264 $[-0.479, -0.056]$
Llama	128	1	$+0.005$ $[-0.072, 0.080]$	-0.047 $[-0.275, 0.176]$
Llama	128	2	$+0.005$ $[-0.087, 0.100]$	-0.011 $[-0.231, 0.209]$
Qwen	64	0	-0.030 $[-0.100, 0.040]$	-0.019 $[-0.181, 0.139]$
Qwen	64	1	$+0.022$ $[-0.055, 0.098]$	$+0.044$ $[-0.129, 0.219]$
Qwen	64	2	-0.060 $[-0.163, 0.045]$	-0.163 $[-0.398, 0.070]$
Qwen	128	0	-0.045 $[-0.120, 0.028]$	-0.060 $[-0.227, 0.097]$
Qwen	128	1	$+0.003$ $[-0.085, 0.095]$	-0.047 $[-0.258, 0.161]$
Qwen	128	2	$+0.010$ $[-0.090, 0.105]$	-0.076 $[-0.288, 0.129]$

Table 12: NoveltyBench scores. DAR minus Release at $N = T$, on common prompts. Brackets are descriptive 95% paired intervals from resampling prompts.

Model	T	ρ	Δ pass@8	Δ maj@8
Llama	64	0	-0.3 $[-1.1, 0.5]$	-0.6 $[-1.8, 0.3]$
Llama	64	1	-1.2 $[-2.5, -0.0]$	-1.7 $[-4.5, 0.9]$
Llama	64	2	-1.0 $[-2.8, 0.7]$	$+4.0$ $[-0.6, 8.9]$
Llama	128	0	-1.5 $[-3.0, -0.2]$	-1.1 $[-3.2, 0.5]$
Llama	128	1	-2.5 $[-4.5, -0.7]$	-0.5 $[-3.2, 2.2]$
Llama	128	2	-2.8 $[-4.8, -0.8]$	$+4.5$ $[1.4, 7.8]$
Qwen	64	0	-0.1 $[-0.6, 0.3]$	-0.3 $[-0.8, 0.0]$
Qwen	64	1	-0.4 $[-1.1, 0.0]$	-0.2 $[-0.5, 0.0]$
Qwen	64	2	$+0.0$ $[0.0, 0.0]$	$+0.1$ $[0.0, 0.2]$
Qwen	128	0	-0.0 $[-0.7, 0.6]$	-0.3 $[-0.8, 0.0]$
Qwen	128	1	-0.2 $[-0.8, 0.0]$	-0.1 $[-0.3, 0.0]$
Qwen	128	2	$+0.0$ $[0.0, 0.0]$	$+0.1$ $[0.0, 0.3]$

Table 13: Mathematical scores at $B = 8$. DAR minus Release at $N = T$, on common prompts. Brackets are descriptive 95% paired intervals from resampling prompts. Differences are percentage points.

Table 15 reports additional-response counts, pilot usage, and solver times. The policy may choose fewer responses at the larger limit because the policy and the pilot mass m/T in R_T both change with T. Generation-token comparisons exclude feature computation, solver time, and calibration.

Controls matched to calibration cost. We compare DAR with four controls chosen using calibration prompts only. Fixed Release uses either the candidate count nearest to the calibration mean cost of DAR, or a mixture of adjacent counts that matches that mean. Two controls keep the con-

Model	T	ρ	Δ constant	Δ accuracy
Llama	64	0	$+0.1$ $[-1.4, 1.6]$	$+0.2$ $[-0.3, 0.6]$
Llama	64	1	$+1.5$ $[-0.5, 3.7]$	$+0.5$ $[-0.1, 1.0]$
Llama	64	2	$+5.3$ $[2.1, 8.9]$	-0.2 $[-0.7, 0.2]$
Llama	128	0	$+4.3$ $[1.7, 7.1]$	$+0.5$ $[-0.1, 1.1]$
Llama	128	1	$+5.2$ $[2.3, 8.4]$	$+0.7$ $[0.1, 1.3]$
Llama	128	2	$+14.3$ $[9.3, 19.4]$	-0.3 $[-1.0, 0.4]$
Qwen	64	0	$+0.7$ $[-0.4, 2.1]$	-0.1 $[-0.3, 0.1]$
Qwen	64	1	$+2.1$ $[0.4, 4.2]$	-0.0 $[-0.2, 0.2]$
Qwen	64	2	$+0.0$ $[0.0, 0.0]$	$+0.1$ $[-0.1, 0.3]$
Qwen	128	0	$+0.7$ $[-0.9, 2.7]$	-0.0 $[-0.2, 0.1]$
Qwen	128	1	-0.2 $[-1.2, 0.7]$	$+0.1$ $[-0.0, 0.2]$
Qwen	128	2	$+0.3$ $[-0.2, 1.1]$	$+0.1$ $[-0.1, 0.3]$

Table 14: Mathematical scores at $B = 16$. DAR minus Release at $N = T$, on common prompts. Brackets are descriptive 95% paired intervals from resampling prompts. Differences are percentage points.

Model	Task	T	Mean N_T	$N_T = 16$, %	$N_T = T - 8$, %	Pilot selected	Solver, s
Llama	Embedding	64	22.0	81	11	0.512	0.196
Llama	Embedding	128	54.2	24	19	0.230	1.256
Qwen	Embedding	64	16.6	97	0	0.487	0.129
Qwen	Embedding	128	18.7	89	1	0.237	0.189
Llama	NB	64	23.9	58	3	0.535	0.006
Llama	NB	128	21.8	73	0	0.230	0.004
Qwen	NB	64	17.2	85	0	0.500	0.002
Qwen	NB	128	25.8	54	3	0.253	0.005
Llama	Math 8	64	40.0	20	33	1.052	0.001
Llama	Math 8	128	26.0	72	4	0.490	0.001
Qwen	Math 8	64	27.2	17	2	0.990	0.001
Qwen	Math 8	128	28.0	16	1	0.465	0.000
Llama	GRPO 16	64	37.8	24	24	1.980	0.001
Llama	GRPO 16	128	44.9	26	10	1.025	0.001
Qwen	GRPO 16	64	27.0	18	2	2.005	0.001
Qwen	GRPO 16	128	28.1	16	1	1.000	0.000

Table 15: DAR diagnostics at $\rho = 0$. Pilot selected is the mean number of pilot responses per returned batch. Solver time excludes cached feature construction and calibration.

straint of DAR: one fixes the additional count by calibration cost matching, and the other draws the count independently of the pilot from the calibration distribution of counts. Held-out costs need not match. Tables 16–19 report all four controls at $T = 64$ and $\rho \in \{0, 2\}$. At $\rho = 0$, the fixed-size mixture differs from DAR in embedding diversity by 0.002 on Llama and 0.006 on Qwen, with both paired intervals containing zero. Table 19 shows the clearest difference. On Llama at $\rho = 0$, the Fixed control has 5.6 percentage points more constant-reward GRPO groups than DAR, with paired 95% interval $[3.3, 8.2]$, while its token difference has interval $[-3.7, 0.6]$ thousand. Across tasks, the controls do not show a uniform advantage for conditioning the response count on the pilot.

Model	ρ	Control	$\Delta\widehat{\text{Div}}$	Δ tokens, 10^3
Llama	0	Fixed	$+0.000$ $[-0.009, 0.010]$	-0.3 $[-0.4, -0.2]$
Llama	0	Mixture	$+0.002$ $[-0.007, 0.012]$	-0.0 $[-0.3, 0.3]$
Llama	0	Constant	-0.006 $[-0.012, -0.001]$	-0.3 $[-0.4, -0.1]$
Llama	0	Independent count	$+0.005$ $[-0.005, 0.015]$	$+7.0$ $[3.6, 10.8]$
Llama	2	Fixed	$+0.005$ $[0.001, 0.011]$	$+1.3$ $[0.4, 2.7]$
Llama	2	Mixture	$+0.001$ $[-0.009, 0.010]$	$+0.9$ $[-0.3, 2.3]$
Llama	2	Constant	$+0.005$ $[0.001, 0.011]$	$+1.3$ $[0.3, 2.6]$
Llama	2	Independent count	-0.003 $[-0.013, 0.006]$	$+0.5$ $[-1.0, 2.2]$
Qwen	0	Fixed	$+0.006$ $[-0.001, 0.014]$	-0.6 $[-1.5, 0.0]$
Qwen	0	Mixture	$+0.006$ $[-0.002, 0.015]$	-0.4 $[-1.4, 0.3]$
Qwen	0	Constant	$+0.001$ $[-0.000, 0.002]$	-0.6 $[-1.5, 0.0]$
Qwen	0	Independent count	-0.000 $[-0.002, 0.001]$	-0.5 $[-1.4, 0.2]$
Qwen	2	Fixed	-0.050 $[-0.081, -0.020]$	-10.7 $[-14.5, -6.6]$
Qwen	2	Mixture	-0.062 $[-0.094, -0.032]$	-13.8 $[-18.1, -9.2]$
Qwen	2	Constant	-0.089 $[-0.119, -0.058]$	-20.9 $[-25.1, -16.7]$
Qwen	2	Independent count	-0.038 $[-0.074, -0.003]$	-4.0 $[-10.3, 2.4]$

Table 16: Embedding controls minus DAR, with descriptive 95% paired intervals from resampling prompts. Fixed and Mixture use the constraint of Release. Constant and Independent count use the constraint of DAR. Positive token differences mean the control costs more.

Model	ρ	Control	Δ utility	Δ tokens, 10^3
Llama	0	Fixed	$+0.030$ $[-0.144, 0.219]$	-2.4 $[-3.5, -1.3]$
Llama	0	Mixture	$+0.082$ $[-0.103, 0.258]$	-1.4 $[-2.8, -0.2]$
Llama	0	Constant	$+0.061$ $[-0.109, 0.237]$	-2.4 $[-3.6, -1.3]$
Llama	0	Independent count	$+0.068$ $[-0.127, 0.255]$	$+6.7$ $[3.4, 10.5]$
Llama	2	Fixed	-0.096 $[-0.284, 0.103]$	-0.2 $[-0.3, -0.1]$
Llama	2	Mixture	-0.104 $[-0.289, 0.088]$	$+0.2$ $[-0.1, 0.6]$
Llama	2	Constant	$+0.033$ $[-0.126, 0.186]$	-0.2 $[-0.3, -0.1]$
Llama	2	Independent count	$+0.123$ $[-0.068, 0.334]$	$+4.1$ $[1.6, 7.2]$
Qwen	0	Fixed	$+0.063$ $[-0.107, 0.224]$	-2.0 $[-2.9, -1.1]$
Qwen	0	Mixture	$+0.140$ $[-0.036, 0.311]$	$+0.6$ $[-0.8, 1.9]$
Qwen	0	Constant	$+0.101$ $[-0.041, 0.234]$	-2.0 $[-3.0, -1.1]$
Qwen	0	Independent count	$+0.103$ $[-0.040, 0.249]$	$+0.6$ $[-0.6, 1.8]$
Qwen	2	Fixed	$+0.087$ $[-0.070, 0.250]$	-3.7 $[-4.9, -2.6]$
Qwen	2	Mixture	$+0.169$ $[-0.000, 0.353]$	$+1.6$ $[-0.3, 3.5]$
Qwen	2	Constant	$+0.083$ $[-0.050, 0.228]$	-3.7 $[-4.9, -2.6]$
Qwen	2	Independent count	$+0.011$ $[-0.171, 0.197]$	$+4.8$ $[1.7, 8.1]$

Table 17: NoveltyBench controls minus DAR, with descriptive 95% paired intervals from resampling prompts. Fixed and Mixture use the constraint of Release. Constant and Independent count use the constraint of DAR. Positive token differences mean the control costs more.

Model	ρ	Control	Δ pass, pp	Δ tokens, 10^3
Llama	0	Fixed	-0.6 $[-1.4, 0.1]$	-3.9 $[-6.4, -1.4]$
Llama	0	Mixture	-0.0 $[-1.0, 0.8]$	$+5.3$ $[1.5, 8.8]$
Llama	0	Constant	-0.2 $[-1.2, 0.8]$	$+4.1$ $[0.9, 7.4]$
Llama	0	Independent count	$+0.5$ $[-0.2, 1.2]$	$+16.0$ $[10.2, 22.1]$
Llama	2	Fixed	$+0.3$ $[-1.0, 1.6]$	$+3.0$ $[1.0, 4.9]$
Llama	2	Mixture	$+0.3$ $[-1.1, 1.8]$	$+2.4$ $[0.6, 4.5]$
Llama	2	Constant	-0.4 $[-2.2, 1.4]$	$+3.0$ $[1.1, 5.0]$
Llama	2	Independent count	-0.7 $[-2.8, 1.3]$	$+10.2$ $[6.1, 14.7]$
Qwen	0	Fixed	-0.0 $[-0.2, 0.1]$	-15.7 $[-23.6, -8.1]$
Qwen	0	Mixture	-0.0 $[-0.2, 0.0]$	-1.4 $[-10.4, 8.3]$
Qwen	0	Constant	$+0.0$ $[-0.3, 0.3]$	-15.7 $[-23.8, -8.3]$
Qwen	0	Independent count	$+0.2$ $[0.0, 0.6]$	-7.6 $[-16.2, 1.3]$
Qwen	2	Fixed	$+0.0$ $[0.0, 0.0]$	-7.0 $[-14.3, 1.8]$
Qwen	2	Mixture	$+0.0$ $[0.0, 0.0]$	$+2.2$ $[-5.5, 11.5]$
Qwen	2	Constant	$+0.0$ $[0.0, 0.0]$	-7.0 $[-15.0, 2.6]$
Qwen	2	Independent count	$+0.0$ $[0.0, 0.0]$	-30.6 $[-46.5, -13.1]$

Table 18: Pass@8 controls minus DAR, with descriptive 95% paired intervals from resampling prompts. Fixed and Mixture use the constraint of Release. Constant and Independent count use the constraint of DAR. Positive token differences mean the control costs more.

Model	ρ	Control	Δ constant, pp	Δ tokens, 10^3
Llama	0	Fixed	$+5.6$ $[3.3, 8.2]$	-1.6 $[-3.7, 0.6]$
Llama	0	Mixture	$+3.3$ $[1.2, 5.6]$	$+3.3$ $[0.6, 6.2]$
Llama	0	Constant	$+2.3$ $[0.5, 4.2]$	$+6.3$ $[3.4, 9.5]$
Llama	0	Independent count	$+1.0$ $[-1.0, 3.0]$	$+13.3$ $[8.8, 18.2]$
Llama	2	Fixed	-2.9 $[-6.7, 0.8]$	-0.7 $[-3.7, 2.3]$
Llama	2	Mixture	-3.2 $[-6.9, 0.5]$	$+2.4$ $[-1.0, 5.9]$
Llama	2	Constant	$+3.8$ $[1.4, 6.5]$	$+8.1$ $[4.3, 12.2]$
Llama	2	Independent count	$+2.5$ $[-0.4, 5.5]$	$+10.5$ $[5.9, 15.4]$
Qwen	0	Fixed	$+0.4$ $[-0.5, 1.6]$	-15.4 $[-23.5, -7.9]$
Qwen	0	Mixture	$+0.0$ $[-0.9, 1.2]$	$+2.2$ $[-7.8, 13.0]$
Qwen	0	Constant	$+0.9$ $[0.0, 2.1]$	-15.4 $[-23.4, -7.9]$
Qwen	0	Independent count	$+0.3$ $[-1.0, 1.8]$	-2.1 $[-10.6, 7.0]$
Qwen	2	Fixed	$+1.0$ $[0.0, 2.5]$	-2.9 $[-15.1, 10.1]$
Qwen	2	Mixture	$+1.0$ $[0.0, 2.5]$	-1.5 $[-13.6, 11.8]$
Qwen	2	Constant	$+0.0$ $[0.0, 0.0]$	$+28.4$ $[16.2, 42.3]$
Qwen	2	Independent count	$+2.8$ $[0.8, 5.0]$	-33.3 $[-48.7, -18.8]$

Table 19: GRPO controls minus DAR, with descriptive 95% paired intervals from resampling prompts. Fixed and Mixture use the constraint of Release. Constant and Independent count use the constraint of DAR. Positive token differences mean the control costs more.

Pilot size and calibration split. At $T = 64$ and $\rho \in \{0, 2\}$, we separately calibrate $m \in \{4, 8, 16\}$. The admissible grid contains the original candidate counts not exceeding $T - m$, together with $T - m$. We also retain three additional source-stratified $80/20$ calibration splits at $m = 8$. No held-out result is used to select the pilot size or split. Tables 20 and 21 report every task and model. The primary setting remains $m = 8$ with the original split.

Model	Task	ρ	Score, $m = 4$	Tokens	Score, $m = 8$	Tokens	Score, $m = 16$	Tokens
Llama	Embedding	0	1.112 ± 0.013	16.0	1.117 ± 0.012	19.4	1.131 ± 0.010	25.5
Llama	Embedding	2	1.925 ± 0.043	48.5	1.925 ± 0.042	49.6	1.917 ± 0.042	45.9
Qwen	Embedding	0	1.103 ± 0.011	29.8	1.103 ± 0.011	32.6	1.137 ± 0.008	77.4
Qwen	Embedding	2	1.862 ± 0.036	84.5	1.833 ± 0.036	74.9	1.859 ± 0.036	83.2
Llama	NoveltyBench	0	5.109 ± 0.230	17.2	5.021 ± 0.239	21.5	5.012 ± 0.220	25.5
Llama	NoveltyBench	2	5.642 ± 0.257	16.1	5.558 ± 0.259	19.3	5.668 ± 0.250	35.5
Qwen	NoveltyBench	0	5.351 ± 0.197	32.1	5.299 ± 0.210	34.0	5.242 ± 0.209	45.5
Qwen	NoveltyBench	2	5.959 ± 0.238	40.1	5.864 ± 0.238	35.7	5.934 ± 0.236	45.1
Llama	Pass@8	0	84.7 ± 2.5	22.4	85.3 ± 2.4	37.0	84.2 ± 2.5	35.3
Llama	Pass@8	2	85.8 ± 2.4	34.7	86.3 ± 2.4	30.1	85.9 ± 2.4	36.2
Qwen	Pass@8	0	96.7 ± 1.2	199.3	96.6 ± 1.2	140.4	96.6 ± 1.2	148.9
Qwen	Pass@8	2	97.5 ± 1.1	219.3	97.5 ± 1.1	224.6	97.5 ± 1.1	209.6
Llama	GRPO	0	26.0 ± 2.9	22.4	21.5 ± 2.7	34.7	24.4 ± 2.8	35.3
Llama	GRPO	2	20.4 ± 2.8	48.6	18.8 ± 2.8	49.6	17.9 ± 2.7	47.4
Qwen	GRPO	0	94.0 ± 1.5	199.6	94.9 ± 1.5	140.1	94.8 ± 1.5	148.9
Qwen	GRPO	2	91.5 ± 2.0	227.9	91.5 ± 2.0	189.2	91.5 ± 2.0	224.9

Table 20: Pilot-size comparison. Scores are design diversity for Embedding, utility for NoveltyBench, pass@8 for mathematics, and constant-reward percentage for GRPO. Score entries are means $\pm$ SE. Token means are in thousands. Smaller constant fractions are preferable.

Model	Task	ρ	Score range	Token range
Llama	Embedding	0	1.114–1.117	19.2–19.4
Llama	Embedding	2	1.896–1.925	39.9–49.6
Qwen	Embedding	0	1.103–1.109	32.6–37.9
Qwen	Embedding	2	1.833–1.865	74.9–85.3
Llama	NoveltyBench	0	4.978–5.066	19.2–30.1
Llama	NoveltyBench	2	5.441–5.619	19.2–22.6
Qwen	NoveltyBench	0	5.263–5.299	32.3–34.0
Qwen	NoveltyBench	2	5.863–5.928	35.2–54.2
Llama	Pass@8	0	84.6–85.3	27.7–37.0
Llama	Pass@8	2	85.8–86.3	30.1–36.8
Qwen	Pass@8	0	96.6–96.8	133.6–149.0
Qwen	Pass@8	2	97.0–97.5	184.4–224.6
Llama	GRPO	0	21.5–23.2	27.5–34.7
Llama	GRPO	2	17.8–19.9	46.0–52.4
Qwen	GRPO	0	94.6–95.1	134.2–149.0
Qwen	GRPO	2	91.5–91.7	189.2–211.7

Table 21: Ranges of held-out means across the original and three additional calibration splits. Score and token units match Table 20. Ranges describe split sensitivity, not confidence intervals.

E.3 NoveltyBench scores

Table 22 lists the means and standard errors of distinct and utility across prompts for the methods in Figure 2.

E.4 Diversity and distortion on open-ended tasks

Table 23 evaluates all returned batches in the same bge-base feature space. Release and DAR use NoveltyBench class features for selection. The measurements complement distinct and utility and do not measure the class-feature constraint.

	Llama		Qwen	
Method	Distinct	Utility	Distinct	Utility
Independent	2.725±0.100	4.955±0.229	1.970±0.093	5.204±0.201
Regeneration	2.843±0.098	4.639±0.212	2.473±0.100	5.802±0.215
Arithmetic	2.745±0.101	4.985±0.214	1.983±0.095	5.217±0.202
Stochastic beam	2.753±0.101	5.007±0.223	1.243±0.040	3.668±0.094
Verbalized	**3.327**±0.089	2.789±0.230	**3.068**±0.107	3.526±0.263
MMR, $\lambda = 0.3$	3.120±0.109	5.154±0.251	2.380±0.123	5.993±0.258
MMR, $\lambda = 0.5$	3.070±0.108	5.362±0.250	2.330±0.117	5.914±0.250
MMR, $\lambda = 0.7$	2.940±0.117	5.679±0.253	2.150±0.118	5.541±0.250
k-DPP	2.805±0.095	5.164±0.218	2.098±0.096	5.467±0.207
Release, $\rho = 0$	2.790±0.098	5.116±0.216	2.025±0.098	5.253±0.211
Release, $\rho = 1.0$	2.950±0.094	5.444±0.229	2.243±0.107	5.719±0.235
Release, $\rho = 2.0$	3.090±0.084	5.589±0.233	2.378±0.104	6.042±0.230
DAR, $\rho = 0$	2.745±0.102	4.992±0.231	1.995±0.095	5.234±0.208
DAR, $\rho = 1.0$	3.073±0.095	5.388±0.248	2.265±0.112	5.763±0.238
DAR, $\rho = 2.0$	3.128±0.093	5.574±0.251	2.318±0.113	5.879±0.240
Release, $\rho = 4.0$	3.143±0.079	**5.739**±0.223	2.438±0.100	**6.206**±0.223
DAR, $\rho = 4.0$	3.143±0.091	5.569±0.248	2.463±0.107	6.197±0.236

Table 22: NoveltyBench, $B = 4$, with 100 prompts per model. Entries are mean ± SE across prompts. Bold marks the highest observed mean.

		Llama		Qwen	
Method	Outputs	$\widehat{\mathrm{Div}}$	$\widehat{D}$	$\widehat{\mathrm{Div}}$	$\widehat{D}$
Independent	16	1.007±0.017	1.141±0.010	0.994±0.014	1.164±0.010
Regeneration	16	1.123±0.030	1.309±0.014	1.282±0.089	1.457±0.031
Arithmetic	16	0.980±0.011	1.120±0.009	1.012±0.019	1.161±0.010
Stochastic beam	16	0.999±0.015	1.131±0.009	0.661±0.023	1.205±0.009
Verbalized	16	2.244±0.100	1.926±0.043	3.711±0.239	2.472±0.069
MMR, $\lambda = 0.3$	4	1.669±0.046	1.460±0.019	1.613±0.035	1.455±0.015
MMR, $\lambda = 0.5$	4	1.324±0.045	1.389±0.017	1.277±0.035	1.384±0.014
MMR, $\lambda = 0.7$	4	0.877±0.026	1.287±0.011	1.002±0.029	1.326±0.011
k-DPP	16	1.154±0.021	1.184±0.011	1.124±0.016	1.192±0.010
Release, $\rho = 0$	16	1.040±0.015	1.144±0.010	0.998±0.013	1.164±0.010
Release, $\rho = 1$	16	1.013±0.017	1.140±0.009	1.052±0.014	1.177±0.010
Release, $\rho = 2$	16	1.054±0.018	1.162±0.010	1.043±0.014	1.178±0.010
DAR, $\rho = 0$	16	1.008±0.016	1.150±0.010	1.005±0.014	1.184±0.010
DAR, $\rho = 1$	16	1.044±0.022	1.174±0.010	1.033±0.017	1.201±0.008
DAR, $\rho = 2$	16	1.056±0.020	1.190±0.009	1.044±0.018	1.202±0.010

Table 23: Common BGE measurements, mean ± SE, on 100 prompts per model. Outputs is the planned count per prompt. Release uses $N = 64$ and DAR uses $T = 64$. MMR has fewer outputs, so its empirical distortion is not directly comparable with the 16-output rows.

Class-space distortion requires a common class assignment for selected responses and the comparison set and remains unreported. Because NoveltyBench classes depend on the observed pool, the positive-distortion guarantee for a fixed feature map is not asserted for that experiment. Table 24 reports independent controls matched to output count. With four outputs per prompt, independent sampling has empirical distortion 1.242 on Llama and 1.267 on Qwen. MMR has higher empirical distortion with the same number of outputs. The MMR values range from 1.287 to 1.460 on Llama and from 1.326 to 1.455 on Qwen. With 16 outputs per prompt, Release at $\rho = 0$ has empirical distortion 1.144 on Llama and 1.164 on Qwen in Table 23, and independent sampling has 1.141 and 1.164.

We retain all $\rho \in \{0, 0.5, 1, 2, 3, 4, 5\}$ at both size limits. Tables 25 and 26 report the NoveltyBench sweep for Llama and Qwen. Release uses $N = 64$ or 128, and DAR uses $T = 64$ or 128. Entries are means ± SE across prompts.

Task	B	Outputs	Llama	Qwen
Open	4	4	1.242 ± 0.007	1.267 ± 0.005
Open	4	16	1.137 ± 0.008	1.162 ± 0.009
Math	8	8	1.066 ± 0.010	1.059 ± 0.095[30]
Math	8	16	0.990 ± 0.010	0.975 ± 0.094[30]
Math	16	16	0.978 ± 0.011	0.990 ± 0.097[30]

Table 24: Independent empirical distortion with matched output counts, mean $\pm$ SE. Each prompt averages 50 separate evaluations. Superscripts give the number of problems when fewer than 200 are used. Qwen pools have zero final-answer variance on 170 of the 200 mathematical problems, so Qwen distortion uses the remaining 30.

NoveltyBench selection uses one-hot class features, for which $W_2(\bar{\Pi}_{x,q}, P_x)^2 = \sum_j |r_j - p_j| \leq 2$. Appendix C.1 derives the identity for W_2^2. The constraint in (8) is therefore inactive once $\eta_N(\rho) s_x \geq \sqrt{2}$, and Release then maximizes diversity over the pool without a distortion constraint. The classes present in the pool then limit the returned batch, and a larger pool contains more classes. On Llama, Release utility at $\rho = 5$ is lower than at $\rho = 4$ for $N = 64$ and higher for $N = 128$. Qwen does not show this difference. Between consecutive values of $\rho \geq 3$, utility changes by less than one standard error.

Method	ρ	Distinct	Utility	BGE $\widehat{\mathrm{Div}}$	BGE $\widehat{D}$
Release 64	0	2.790 ± 0.098	5.116 ± 0.216	1.040 ± 0.015	1.144 ± 0.010
Release 64	0.5	2.855 ± 0.099	5.189 ± 0.222	1.023 ± 0.014	1.138 ± 0.011
Release 64	1	2.950 ± 0.094	5.444 ± 0.229	1.013 ± 0.017	1.140 ± 0.009
Release 64	2	3.090 ± 0.084	5.589 ± 0.233	1.054 ± 0.018	1.162 ± 0.010
Release 64	3	3.143 ± 0.083	5.631 ± 0.240	1.055 ± 0.017	1.160 ± 0.010
Release 64	4	3.143 ± 0.079	5.739 ± 0.223	1.075 ± 0.022	1.167 ± 0.011
Release 64	5	3.133 ± 0.078	5.648 ± 0.229	1.039 ± 0.019	1.158 ± 0.010
Release 128	0	2.745 ± 0.099	5.111 ± 0.231	1.009 ± 0.015	1.131 ± 0.009
Release 128	0.5	2.848 ± 0.098	5.182 ± 0.222	1.022 ± 0.015	1.129 ± 0.010
Release 128	1	2.998 ± 0.091	5.439 ± 0.231	1.056 ± 0.015	1.140 ± 0.010
Release 128	2	3.118 ± 0.084	5.630 ± 0.232	1.070 ± 0.017	1.154 ± 0.010
Release 128	3	3.160 ± 0.079	5.659 ± 0.232	1.068 ± 0.016	1.153 ± 0.009
Release 128	4	3.150 ± 0.078	5.670 ± 0.238	1.050 ± 0.015	1.146 ± 0.010
Release 128	5	3.188 ± 0.074	5.751 ± 0.230	1.068 ± 0.017	1.161 ± 0.009
DAR 64	0	2.745 ± 0.102	4.992 ± 0.231	1.008 ± 0.016	1.150 ± 0.010
DAR 64	0.5	2.845 ± 0.097	5.263 ± 0.224	1.032 ± 0.016	1.156 ± 0.010
DAR 64	1	3.073 ± 0.095	5.388 ± 0.248	1.044 ± 0.022	1.174 ± 0.010
DAR 64	2	3.128 ± 0.093	5.574 ± 0.251	1.056 ± 0.020	1.190 ± 0.009
DAR 64	3	3.150 ± 0.092	5.556 ± 0.253	1.059 ± 0.020	1.189 ± 0.010
DAR 64	4	3.143 ± 0.091	5.569 ± 0.248	1.067 ± 0.020	1.193 ± 0.010
DAR 64	5	3.165 ± 0.090	5.651 ± 0.233	1.056 ± 0.021	1.191 ± 0.010
DAR 128	0	2.710 ± 0.103	4.847 ± 0.239	1.023 ± 0.017	1.157 ± 0.009
DAR 128	0.5	2.858 ± 0.104	5.222 ± 0.243	1.000 ± 0.017	1.156 ± 0.010
DAR 128	1	3.003 ± 0.100	5.392 ± 0.239	1.040 ± 0.017	1.179 ± 0.010
DAR 128	2	3.123 ± 0.094	5.619 ± 0.248	1.065 ± 0.025	1.192 ± 0.011
DAR 128	3	3.098 ± 0.092	5.456 ± 0.240	1.053 ± 0.021	1.192 ± 0.010
DAR 128	4	3.170 ± 0.095	5.537 ± 0.246	1.059 ± 0.019	1.196 ± 0.010
DAR 128	5	3.185 ± 0.092	5.608 ± 0.244	1.061 ± 0.021	1.192 ± 0.009

Table 25: Llama NoveltyBench sweep, $B = 4$, on 100 prompts. BGE measurements use the returned batches and do not estimate the class-space selection constraint.

Method	ρ	Distinct	Utility	BGE $\widehat{\text{Div}}$	BGE $\widehat{D}$
Release 64	0	2.025 ± 0.098	5.253 ± 0.211	0.998 ± 0.013	1.164 ± 0.010
Release 64	0.5	2.098 ± 0.100	5.456 ± 0.217	1.008 ± 0.011	1.161 ± 0.011
Release 64	1	2.243 ± 0.107	5.719 ± 0.235	1.052 ± 0.014	1.177 ± 0.010
Release 64	2	2.378 ± 0.104	6.042 ± 0.230	1.043 ± 0.014	1.178 ± 0.010
Release 64	3	2.398 ± 0.101	6.067 ± 0.222	1.067 ± 0.019	1.194 ± 0.010
Release 64	4	2.438 ± 0.100	6.206 ± 0.223	1.064 ± 0.018	1.192 ± 0.011
Release 64	5	2.450 ± 0.099	6.168 ± 0.222	1.066 ± 0.017	1.198 ± 0.010
Release 128	0	2.010 ± 0.094	5.230 ± 0.201	1.011 ± 0.012	1.155 ± 0.011
Release 128	0.5	2.043 ± 0.094	5.353 ± 0.210	1.015 ± 0.011	1.155 ± 0.009
Release 128	1	2.248 ± 0.106	5.783 ± 0.233	1.048 ± 0.014	1.170 ± 0.009
Release 128	2	2.340 ± 0.101	5.988 ± 0.219	1.065 ± 0.016	1.180 ± 0.010
Release 128	3	2.473 ± 0.104	6.253 ± 0.233	1.078 ± 0.018	1.193 ± 0.012
Release 128	4	2.480 ± 0.104	6.285 ± 0.229	1.080 ± 0.017	1.184 ± 0.011
Release 128	5	2.430 ± 0.099	6.197 ± 0.224	1.086 ± 0.021	1.194 ± 0.011
DAR 64	0	1.995 ± 0.095	5.234 ± 0.208	1.005 ± 0.014	1.184 ± 0.010
DAR 64	0.5	2.150 ± 0.105	5.564 ± 0.226	1.007 ± 0.013	1.184 ± 0.009
DAR 64	1	2.265 ± 0.112	5.763 ± 0.238	1.033 ± 0.017	1.201 ± 0.008
DAR 64	2	2.318 ± 0.113	5.879 ± 0.240	1.044 ± 0.018	1.202 ± 0.010
DAR 64	3	2.413 ± 0.112	6.119 ± 0.242	1.059 ± 0.017	1.207 ± 0.010
DAR 64	4	2.463 ± 0.107	6.197 ± 0.236	1.058 ± 0.018	1.204 ± 0.010
DAR 64	5	2.448 ± 0.110	6.159 ± 0.242	1.087 ± 0.018	1.212 ± 0.009
DAR 128	0	1.965 ± 0.094	5.170 ± 0.203	1.008 ± 0.013	1.182 ± 0.010
DAR 128	0.5	2.120 ± 0.104	5.459 ± 0.219	0.997 ± 0.014	1.182 ± 0.009
DAR 128	1	2.250 ± 0.107	5.736 ± 0.230	1.024 ± 0.013	1.180 ± 0.008
DAR 128	2	2.350 ± 0.111	5.912 ± 0.234	1.063 ± 0.019	1.204 ± 0.010
DAR 128	3	2.415 ± 0.109	6.075 ± 0.234	1.075 ± 0.017	1.209 ± 0.010
DAR 128	4	2.465 ± 0.108	6.156 ± 0.236	1.069 ± 0.018	1.217 ± 0.009
DAR 128	5	2.400 ± 0.113	6.062 ± 0.242	1.055 ± 0.019	1.208 ± 0.010

Table 26: Qwen NoveltyBench sweep, $B = 4$, on 100 prompts. BGE measurements use the returned batches and do not estimate the class-space selection constraint.

Feature choice. We compare Release at $N = 64$ using class indicators and bge-base embeddings on the same prompts, with $\rho \in \{0, 2\}$. Both feature choices use the official NoveltyBench evaluator. Table 27 reports both task metrics.

Model	ρ	Features	Distinct	Utility
Llama	0	Classes	2.790 ± 0.098	5.116 ± 0.216
Llama	0	Embedding	2.840 ± 0.101	5.157 ± 0.231
Llama	2	Classes	3.090 ± 0.084	5.589 ± 0.233
Llama	2	Embedding	3.193 ± 0.092	5.064 ± 0.243
Qwen	0	Classes	2.025 ± 0.098	5.253 ± 0.211
Qwen	0	Embedding	2.088 ± 0.100	5.418 ± 0.209
Qwen	2	Classes	2.378 ± 0.104	6.042 ± 0.230
Qwen	2	Embedding	2.400 ± 0.120	5.952 ± 0.252

Table 27: Feature comparison, $B = 4$, on 100 prompts per model. Entries are means $\pm$ SE.

Model	ρ	Different partition, %
Llama	0	45.8 ± 3.5
Llama	2	47.0 ± 3.6
Qwen	0	41.0 ± 3.8
Qwen	2	46.5 ± 4.0

Table 28: Fraction of returned class-feature batches whose pool-induced partition differs from the official batch partition, mean $\pm$ SE across prompts.

Pool-level classes and batch-level evaluation need not agree because classifier matching is order-dependent and need not be transitive. Table 28 reports how often the pool partition and the batch partition differ. The fixed-partition results in Appendix B.3 therefore do not assert optimality for the official distinct or utility scores.

E.5 MMR PARAMETER RANGE

Table 29 extends MMR to $\lambda \in \{0.1, 0.9\}$ on the same 100 held-out prompts per model, with 64 and 128 candidates. Utility is highest at $\lambda = 0.7$ on Llama and at $\lambda = 0.3$ on Qwen, so the largest MMR utility lies inside the range reported in Table 22. With 128 candidates, MMR utility does not exceed its value with 64 candidates at the same λ.

		Llama		Qwen	
Candidates	λ	Distinct	Utility	Distinct	Utility
64	0.1	3.120 ± 0.107	5.081 ± 0.245	2.380 ± 0.123	5.970 ± 0.257
64	0.3	3.120 ± 0.109	5.154 ± 0.251	2.380 ± 0.123	5.993 ± 0.258
64	0.5	3.070 ± 0.108	5.362 ± 0.250	2.330 ± 0.117	5.914 ± 0.250
64	0.7	2.940 ± 0.117	5.679 ± 0.253	2.150 ± 0.118	5.541 ± 0.250
64	0.9	2.740 ± 0.115	5.330 ± 0.252	1.990 ± 0.110	5.233 ± 0.239
128	0.1	3.040 ± 0.108	4.875 ± 0.234	2.380 ± 0.121	5.931 ± 0.258
128	0.9	2.660 ± 0.110	5.140 ± 0.256	1.940 ± 0.105	5.111 ± 0.229

Table 29: MMR on NoveltyBench, $B = 4$, with 100 held-out prompts per model. Entries are mean $\pm$ SE across prompts.

E.6 SAMPLER SETTINGS ON LLAMA

Table 30 compares independent sampling with six sampler settings of Llama on the 100 held-out open-ended prompts. Each setting changes the temperature, the top-p threshold (Holtzman et al., 2020), or the min-p threshold (Nguyen et al., 2025) of the base configuration in Appendix D.1 and keeps the other parameters. Each setting generates 16 responses per prompt. For every setting, four batches of size $B = 4$ are drawn uniformly from the 16 responses with seed zero, and all settings use the same batch indices. Responses that reach the length limit are kept. The control applies the same procedure to the first 16 responses of the base pool.

The 16 outputs of four batches contain 11.0 distinct responses on average, whereas four batches from 64 responses contain about 14.6. Repeated outputs can raise empirical distortion. The control has empirical distortion 1.160, and independent sampling with 64 candidates has 1.141. Each setting is therefore compared with the control, and Table 31 reports paired differences with descriptive 95% intervals computed as in Appendix D.3.

Temperature 0.8, top-p, and min-p have lower distinct, lower diversity, and higher utility than the control. Temperature 1.2 has higher distinct and lower utility than the control, and 9.5% of its responses reach the length limit. Every setting has higher empirical distortion than the control. The difference is between 0.058 and 0.069 for the five settings other than temperature 1.2 and is 0.507 for temperature 1.2. The paired intervals exclude zero for every difference except the diversity in the two rows of temperature 1.2. Excluding responses that reach the length limit changes each mean of temperature 1.2 by less than 0.04.

On the same prompts, Release at $\rho = 4$ has distinct 3.143 and utility 5.739 in Table 22. Release at $\rho = 4$ has higher distinct and higher utility than every sampler setting except temperature 1.2, which has higher distinct and lower utility. The utility of top-p 0.95 is 5.694, within one standard error of Release at $\rho = 4$. At $\rho = 0$, Release has distinct 2.790 and utility 5.116. Release at $\rho = 0$ has higher distinct and lower utility than every sampler setting except temperature 1.2. In Table 23, Release at $\rho = 0$ has empirical distortion 1.144 and independent sampling has 1.141.

Setting	Candidates	Distinct	Utility	$\widehat{\mathrm{Div}}$	$\widehat{D}$
Independent	64	2.725 ± 0.100	4.955 ± 0.229	1.007 ± 0.017	1.141 ± 0.010
Control	16	2.708 ± 0.101	4.917 ± 0.222	1.004 ± 0.019	1.160 ± 0.010
Temperature 0.8	16	2.430 ± 0.103	5.647 ± 0.211	0.826 ± 0.018	1.228 ± 0.017
Temperature 1.2	16	3.278 ± 0.083	2.573 ± 0.194	0.916 ± 0.077	1.667 ± 0.039
Temperature 1.2, uncapped	14.5	3.310 ± 0.082	2.595 ± 0.190	0.932 ± 0.076	1.652 ± 0.038
Top-p 0.9	16	2.360 ± 0.099	5.462 ± 0.205	0.832 ± 0.018	1.225 ± 0.017
Top-p 0.95	16	2.493 ± 0.104	5.694 ± 0.206	0.879 ± 0.019	1.222 ± 0.015
Min-p 0.05	16	2.325 ± 0.100	5.511 ± 0.209	0.817 ± 0.016	1.228 ± 0.017
Min-p 0.1	16	2.240 ± 0.100	5.301 ± 0.205	0.792 ± 0.017	1.217 ± 0.017

Table 30: Sampler settings of Llama on 100 held-out open-ended prompts at $B = 4$. Entries are mean $\pm$ SE across prompts. Candidates is the number of responses per prompt from which batches are drawn. The uncapped row of temperature 1.2 excludes responses that reach the length limit and redraws batches from the remaining responses. Diversity and distortion use bge-base features.

Setting	Distinct	Utility	$\widehat{\mathrm{Div}}$	$\widehat{D}$
Temperature 0.8	-0.278 [-0.380, -0.173]	0.730 [0.388, 1.062]	-0.178 [-0.232, -0.130]	0.068 [0.038, 0.098]
Temperature 1.2	0.570 [0.450, 0.688]	-2.343 [-2.721, -1.969]	-0.087 [-0.202, 0.033]	0.507 [0.443, 0.574]
Temperature 1.2, uncapped	0.603 [0.478, 0.730]	-2.322 [-2.717, -1.933]	-0.072 [-0.196, 0.056]	0.493 [0.431, 0.556]
Top-p 0.9	-0.348 [-0.465, -0.230]	0.545 [0.236, 0.854]	-0.172 [-0.221, -0.126]	0.066 [0.038, 0.094]
Top-p 0.95	-0.215 [-0.328, -0.100]	0.777 [0.462, 1.094]	-0.124 [-0.179, -0.073]	0.063 [0.036, 0.089]
Min-p 0.05	-0.383 [-0.498, -0.273]	0.594 [0.294, 0.910]	-0.187 [-0.231, -0.143]	0.069 [0.041, 0.097]
Min-p 0.1	-0.468 [-0.585, -0.348]	0.384 [0.056, 0.725]	-0.212 [-0.256, -0.171]	0.058 [0.028, 0.087]

Table 31: Paired differences between each sampler setting and the control of Table 30, computed as the setting minus the control, with descriptive 95% intervals over 100 prompts.

E.7 Paired comparisons with independent sampling

Table 32 reports paired differences between each method at $\rho = 0$ and independent sampling, with descriptive 95% intervals computed as in Appendix D.3. The same resamples give a two-sided p-value for each difference, equal to twice the smaller of the fractions of resampled mean differences at most zero and at least zero, capped at one. After Holm correction (Holm, 1979) over the 64 differences in Table 32, Release with $N = 64$ differs significantly from independent sampling in embedding diversity and the constant-reward fraction on both models, in embedding distortion on Qwen, and in pass@8 and majority@8 on Llama.

Model	Metric	Release, N=64	Release, N=128	DAR, T=64	DAR, T=128
Llama	Diversity	$+0.150$ [0.119, 0.180]	$+0.167$ [0.132, 0.199]	$+0.110$ [0.071, 0.147]	$+0.141$ [0.104, 0.176]
	Distortion	-0.018 [-0.031, -0.005]	-0.025 [-0.038, -0.012]	-0.003 [-0.019, 0.013]	-0.012 [-0.028, 0.004]
	Distinct	$+0.065$ [-0.022, 0.155]	$+0.020$ [-0.068, 0.113]	$+0.020$ [-0.075, 0.115]	-0.015 [-0.113, 0.093]
	Utility	$+0.161$ [-0.010, 0.338]	$+0.156$ [-0.068, 0.397]	$+0.038$ [-0.188, 0.264]	-0.107 [-0.338, 0.125]
	Pass@8	$+2.13$ [1.33, 3.01]	$+2.81$ [1.78, 3.90]	$+1.86$ [0.85, 3.01]	$+1.30$ [-0.04, 2.60]
	Majority@8	$+3.31$ [1.80, 4.80]	$+2.54$ [0.95, 4.11]	$+2.72$ [1.01, 4.30]	$+1.41$ [-0.43, 3.13]
	Constant	-4.83 [-6.17, -3.58]	-9.12 [-11.25, -7.05]	-4.74 [-6.52, -3.08]	-4.80 [-6.96, -2.66]
	Accuracy	-0.13 [-0.28, 0.00]	-0.41 [-0.86, 0.07]	$+0.02$ [-0.43, 0.46]	$+0.08$ [-0.43, 0.60]
Qwen	Diversity	$+0.145$ [0.123, 0.168]	$+0.164$ [0.140, 0.189]	$+0.108$ [0.084, 0.133]	$+0.103$ [0.079, 0.129]
	Distortion	-0.021 [-0.030, -0.012]	-0.024 [-0.033, -0.015]	$+0.010$ [0.001, 0.019]	$+0.021$ [0.010, 0.032]
	Distinct	$+0.055$ [-0.043, 0.150]	$+0.040$ [-0.043, 0.125]	$+0.025$ [-0.060, 0.113]	-0.005 [-0.087, 0.072]
	Utility	$+0.049$ [-0.176, 0.267]	$+0.026$ [-0.148, 0.219]	$+0.030$ [-0.161, 0.221]	-0.034 [-0.228, 0.150]
	Pass@8	$+0.09$ [-0.03, 0.26]	$+0.05$ [-0.26, 0.40]	$+0.01$ [-0.33, 0.36]	$+0.02$ [-0.33, 0.38]
	Majority@8	$+0.08$ [0.00, 0.24]	$+0.08$ [0.00, 0.24]	-0.19 [-0.60, 0.02]	-0.18 [-0.55, 0.02]
	Constant	-1.10 [-1.93, -0.40]	-1.04 [-2.14, -0.18]	-0.37 [-1.74, 0.96]	-0.30 [-1.81, 1.17]
	Accuracy	-0.02 [-0.04, 0.00]	-0.08 [-0.21, 0.01]	-0.10 [-0.32, 0.06]	-0.11 [-0.37, 0.06]

Table 32: Paired differences from independent sampling at $\rho = 0$, computed as the method minus independent sampling, with descriptive 95% intervals over 100 prompts per task and model. Diversity and distortion use embedding features on open-ended prompts. Pass@8, majority@8, the constant-reward fraction, and response accuracy average GSM8K and MATH-500 and are in percentage points. Independent sampling uses the estimators of Tables 1, 3, 4, and 22.

E.8 DIVERSITY AND DISTORTION ON MATHEMATICAL TASKS

Tables 33 and 34 report sampled diversity and empirical distortion in final-answer class space. Zero-variance prompts remain in task scores but cannot contribute to diversity and distortion, which divide by the feature variance. Qwen diversity and distortion therefore use far fewer problems than Qwen task scores.

Method	Llama $\widehat{\mathrm{Div}}$	Llama $\widehat{D}$	Qwen $\widehat{\mathrm{Div}}$	Qwen $\widehat{D}$
Independent	0.974 ± 0.023	0.988 ± 0.011	$1.053 \pm 0.230^{[30]}$	$0.998 \pm 0.093^{[30]}$
Regeneration	3.108 ± 0.370	2.190 ± 0.103	$18.371 \pm 7.517^{[30]}$	$3.656 \pm 0.796^{[30]}$
Arithmetic	1.103 ± 0.061	0.997 ± 0.021	$1.144 \pm 0.402^{[30]}$	$1.018 \pm 0.130^{[30]}$
Stochastic beam	1.051 ± 0.054	0.984 ± 0.020	$0.061 \pm 0.039^{[30]}$	$0.955 \pm 0.106^{[30]}$
Verbalized	1.698 ± 0.196	1.420 ± 0.049	$16.050 \pm 5.244^{[30]}$	$5.452 \pm 1.065^{[30]}$
MMR, $\lambda = 0.3$	1.314 ± 0.085	1.162 ± 0.025	$2.202 \pm 0.730^{[30]}$	$1.356 \pm 0.183^{[30]}$
MMR, $\lambda = 0.5$	1.096 ± 0.055	1.097 ± 0.019	$2.049 \pm 0.721^{[30]}$	$1.276 \pm 0.183^{[30]}$
MMR, $\lambda = 0.7$	0.902 ± 0.060	1.064 ± 0.019	$1.015 \pm 0.376^{[30]}$	$1.074 \pm 0.127^{[30]}$
k-DPP	1.066 ± 0.049	1.005 ± 0.019	$1.533 \pm 0.674^{[30]}$	$1.079 \pm 0.161^{[30]}$
Release, $\rho = 0$	1.033 ± 0.024	0.927 ± 0.013	$1.042 \pm 0.205^{[30]}$	$0.960 \pm 0.095^{[30]}$
Release, $\rho = 1$	1.559 ± 0.049	1.208 ± 0.015	$1.899 \pm 0.403^{[30]}$	$1.179 \pm 0.119^{[30]}$
Release, $\rho = 2$	2.307 ± 0.124	1.576 ± 0.033	$4.341 \pm 1.142^{[30]}$	$1.814 \pm 0.222^{[30]}$
DAR, $\rho = 0$	1.043 ± 0.027	0.946 ± 0.013	$0.820 \pm 0.229^{[30]}$	$0.945 \pm 0.103^{[30]}$
DAR, $\rho = 1$	1.577 ± 0.056	1.234 ± 0.017	$1.475 \pm 0.457^{[30]}$	$1.135 \pm 0.135^{[30]}$
DAR, $\rho = 2$	2.312 ± 0.144	1.574 ± 0.038	$4.348 \pm 1.092^{[30]}$	$1.831 \pm 0.213^{[30]}$

Table 33: Final-answer class measurements at $B = 8$, mean $\pm$ SE over eligible prompts. Superscripts give the number of problems when fewer than 200 are used. Qwen pools have zero final-answer variance on 170 of the 200 problems, so Qwen diversity and distortion use the remaining 30. Release uses $N = 64$, DAR uses $T = 64$.

Method	Llama $\widehat{\mathrm{Div}}$	Llama $\widehat{D}$	Qwen $\widehat{\mathrm{Div}}$	Qwen $\widehat{D}$
Independent	0.974 ± 0.024	0.977 ± 0.011	$0.936 \pm 0.186^{[30]}$	$0.969 \pm 0.092^{[30]}$
Regeneration	$2.950 \pm 0.363^{[195]}$	2.333 ± 0.106	$22.363 \pm 7.938^{[30]}$	$4.276 \pm 0.894^{[30]}$
Arithmetic	$0.985 \pm 0.042^{[199]}$	0.967 ± 0.016	$0.794 \pm 0.357^{[29]}$	$0.912 \pm 0.119^{[30]}$
Stochastic beam	1.019 ± 0.044	0.970 ± 0.019	$0.119 \pm 0.067^{[30]}$	$0.992 \pm 0.109^{[30]}$
Verbalized	1.910 ± 0.174	1.429 ± 0.049	$22.656 \pm 7.092^{[30]}$	$5.541 \pm 1.058^{[30]}$
MMR, $\lambda = 0.3$	1.173 ± 0.051	1.054 ± 0.018	$2.534 \pm 0.813^{[30]}$	$1.380 \pm 0.189^{[30]}$
MMR, $\lambda = 0.5$	1.042 ± 0.042	1.003 ± 0.018	$1.357 \pm 0.397^{[30]}$	$1.049 \pm 0.141^{[30]}$
MMR, $\lambda = 0.7$	0.872 ± 0.040	0.993 ± 0.016	$1.223 \pm 0.386^{[30]}$	$1.073 \pm 0.137^{[30]}$
k-DPP	1.008 ± 0.038	0.985 ± 0.016	$1.174 \pm 0.302^{[30]}$	$1.007 \pm 0.124^{[30]}$
Release, $\rho = 0$	1.003 ± 0.025	0.903 ± 0.014	$1.043 \pm 0.229^{[30]}$	$0.942 \pm 0.103^{[30]}$
Release, $\rho = 1$	1.512 ± 0.047	1.201 ± 0.017	$1.992 \pm 0.479^{[30]}$	$1.198 \pm 0.132^{[30]}$
Release, $\rho = 2$	2.023 ± 0.097	1.494 ± 0.028	$3.489 \pm 0.867^{[30]}$	$1.679 \pm 0.200^{[30]}$
DAR, $\rho = 0$	1.012 ± 0.028	0.925 ± 0.014	$0.834 \pm 0.227^{[30]}$	$0.932 \pm 0.106^{[30]}$
DAR, $\rho = 1$	1.514 ± 0.054	1.197 ± 0.019	$1.526 \pm 0.459^{[30]}$	$1.122 \pm 0.141^{[30]}$
DAR, $\rho = 2$	2.005 ± 0.097	1.495 ± 0.029	$3.242 \pm 0.850^{[30]}$	$1.605 \pm 0.195^{[30]}$

Table 34: Final-answer class measurements at $B = 16$, mean $\pm$ SE over eligible prompts. Superscripts give the number of problems when fewer than 200 are used. Qwen pools have zero final-answer variance on 170 of the 200 problems, so Qwen diversity and distortion use at most the remaining 30. Regeneration and Arithmetic sampling have no complete batch on some problems. Release uses $N = 64$, DAR uses $T = 64$.

Tables 35 and 36 report all distortion settings at $B = 8$. Mathematical Release and DAR scores average 50 sets of 16 outputs within each prompt. Distortion is computed separately for each set before averaging.

Method	ρ	Pass@8	Maj@8	$\widehat{\mathrm{Div}}$	$\widehat{D}$
Release 64	0	85.4±2.3	77.0±2.9	1.033±0.024	0.927±0.013
Release 64	0.5	86.2±2.3	74.8±2.9	1.196±0.031	0.981±0.014
Release 64	1	87.2±2.3	62.8±3.2	1.559±0.049	1.208±0.015
Release 64	2	87.2±2.3	42.0±3.4	2.307±0.124	1.576±0.033
Release 64	3	87.2±2.3	39.4±3.4	2.643±0.177	1.708±0.045
Release 64	4	87.2±2.3	39.3±3.4	2.643±0.177	1.707±0.045
Release 64	5	87.2±2.3	39.4±3.4	2.643±0.177	1.708±0.045
Release 128	0	86.1±2.2	76.2±3.0	1.059±0.015	0.907±0.013
Release 128	0.5	87.2±2.2	74.1±3.0	1.227±0.019	0.966±0.013
Release 128	1	88.3±2.1	62.9±3.2	1.616±0.037	1.199±0.013
Release 128	2	88.8±2.1	35.4±3.3	2.465±0.105	1.626±0.028
Release 128	3	88.9±2.1	28.2±3.1	3.071±0.186	1.855±0.049
Release 128	4	88.8±2.1	27.3±3.1	3.341±0.238	1.939±0.061
Release 128	5	88.9±2.1	27.1±3.1	3.341±0.238	1.939±0.061
DAR 64	0	85.1±2.4	76.4±2.9	1.043±0.027	0.946±0.013
DAR 64	0.5	85.2±2.4	74.0±3.0	1.198±0.033	1.012±0.014
DAR 64	1	86.0±2.4	61.1±3.3	1.577±0.056	1.234±0.017
DAR 64	2	86.3±2.4	46.0±3.5	2.312±0.144	1.574±0.038
DAR 64	3	86.1±2.4	41.7±3.5	2.593±0.178	1.688±0.046
DAR 64	4	86.1±2.4	40.6±3.5	2.595±0.178	1.690±0.046
DAR 64	5	86.2±2.4	40.6±3.5	2.594±0.178	1.690±0.046
DAR 128	0	84.6±2.4	75.1±3.0	1.001±0.044	0.976±0.015
DAR 128	0.5	85.0±2.4	74.7±2.9	1.204±0.036	1.016±0.014
DAR 128	1	85.7±2.4	62.4±3.3	1.572±0.056	1.233±0.017
DAR 128	2	85.9±2.4	39.9±3.4	2.366±0.144	1.612±0.035
DAR 128	3	85.8±2.4	28.2±3.2	3.045±0.190	1.861±0.051
DAR 128	4	86.1±2.4	29.9±3.2	3.299±0.239	1.936±0.061
DAR 128	5	86.1±2.4	30.0±3.2	3.299±0.239	1.936±0.061

Table 35: Llama mathematical sweep at $B = 8$. Task scores are percentages, with SE in percentage points, averaged over GSM8K and MATH-500 with equal weight.

Method	ρ	Pass@8	Maj@8	$\widehat{\mathrm{Div}}$	$\widehat{D}$
Release 64	0	96.7±1.2	96.0±1.4	1.042±0.205[30]	0.960±0.095[30]
Release 64	0.5	97.0±1.2	96.0±1.4	1.376±0.330[30]	1.021±0.107[30]
Release 64	1	97.4±1.1	95.7±1.4	1.899±0.403[30]	1.179±0.119[30]
Release 64	2	97.5±1.1	94.8±1.6	4.341±1.142[30]	1.814±0.222[30]
Release 64	3	97.5±1.1	94.5±1.6	6.033±1.661[30]	2.200±0.294[30]
Release 64	4	97.5±1.1	94.5±1.6	6.033±1.661[30]	2.201±0.294[30]
Release 64	5	97.5±1.1	94.5±1.6	6.033±1.661[30]	2.201±0.294[30]
Release 128	0	96.7±1.2	96.0±1.4	1.010±0.147[30]	0.958±0.083[29]
Release 128	0.5	96.9±1.2	95.8±1.4	1.054±0.143[30]	0.951±0.088[29]
Release 128	1	97.5±1.1	95.6±1.4	1.903±0.329[30]	1.188±0.087[29]
Release 128	2	97.5±1.1	94.6±1.6	3.846±0.654[30]	1.797±0.137[29]
Release 128	3	97.5±1.1	94.0±1.7	6.472±1.268[30]	2.442±0.238[29]
Release 128	4	97.5±1.1	92.5±1.9	8.929±2.055[30]	2.899±0.335[29]
Release 128	5	97.5±1.1	92.5±1.9	8.929±2.055[30]	2.899±0.335[29]
DAR 64	0	96.6±1.2	95.7±1.4	0.820±0.229[30]	0.945±0.103[30]
DAR 64	0.5	96.9±1.2	95.7±1.4	0.882±0.209[30]	0.947±0.100[30]
DAR 64	1	97.0±1.2	95.5±1.5	1.475±0.457[30]	1.135±0.135[30]
DAR 64	2	97.5±1.1	94.8±1.6	4.348±1.092[30]	1.831±0.213[30]
DAR 64	3	97.5±1.1	94.5±1.6	5.357±1.575[30]	2.053±0.269[30]
DAR 64	4	97.5±1.1	94.5±1.6	5.404±1.572[30]	2.068±0.268[30]
DAR 64	5	97.5±1.1	94.5±1.6	5.370±1.574[30]	2.058±0.269[30]
DAR 128	0	96.6±1.2	95.7±1.4	0.846±0.260[30]	0.949±0.107[30]
DAR 128	0.5	96.9±1.2	95.6±1.4	1.202±0.261[30]	1.005±0.099[30]
DAR 128	1	97.2±1.1	95.5±1.5	1.878±0.378[30]	1.198±0.106[30]
DAR 128	2	97.5±1.1	94.7±1.6	4.692±1.213[30]	1.911±0.218[30]
DAR 128	3	97.5±1.1	94.4±1.6	6.119±1.414[30]	2.301±0.260[30]
DAR 128	4	97.5±1.1	93.0±1.8	8.748±2.057[30]	2.782±0.325[30]
DAR 128	5	97.5±1.1	92.5±1.9	8.846±2.052[30]	2.828±0.325[30]

Table 36: Qwen mathematical sweep at $B = 8$. Task scores are percentages, with SE in percentage points, averaged over GSM8K and MATH-500 with equal weight. Superscripts give the number of problems when fewer than 200 are used. Pools have zero final-answer variance on 170 of the 200 problems, so diversity and distortion use the remaining 30. For Release with $N = 128$, distortion excludes one further problem whose first 128 usable responses overlap the comparison set.

E.9 GRPO GROUP STATISTICS

Table 38 reports changes in response accuracy relative to the comparison set.

Table 37 reports the reward variance of the groups in Table 4. For binary rewards, the variance is $\bar{r}(1-\bar{r})$. Larger variance indicates a more balanced mixture of correct and incorrect responses, without implying a larger normalized GRPO advantage.

	Llama		Qwen	
Method	GSM8K	MATH-500	GSM8K	MATH-500
Independent	0.121±0.008	0.120±0.009	0.004±0.002	0.005±0.003
Regeneration	0.067±0.006	0.059±0.007	**0.088**±0.009	0.029±0.007
Arithmetic	0.121±0.009	0.116±0.009	0.005±0.003	0.006±0.003
Stochastic beam	0.124±0.009	0.116±0.009	0.001±0.001	0.000±0.000
Verbalized	**0.158**±0.009	0.109±0.010	0.012±0.005	**0.032**±0.007
MMR, $\lambda = 0.3$	0.137±0.009	**0.130**±0.010	0.007±0.003	0.009±0.003
MMR, $\lambda = 0.5$	0.127±0.010	0.126±0.009	0.006±0.003	0.006±0.003
MMR, $\lambda = 0.7$	0.111±0.010	0.120±0.010	0.006±0.003	0.005±0.002
k-DPP	0.122±0.009	0.121±0.009	0.009±0.004	0.006±0.003
Release, $\rho = 0$	0.125±0.009	0.125±0.009	0.004±0.002	0.005±0.003
Release, $\rho = 1.0$	0.138±0.008	0.118±0.009	0.008±0.003	0.007±0.003
Release, $\rho = 2.0$	0.153±0.007	0.108±0.008	0.012±0.004	0.011±0.004
DAR, $\rho = 0$	0.128±0.009	0.124±0.009	0.005±0.002	0.005±0.003
DAR, $\rho = 1.0$	0.143±0.008	0.124±0.009	0.008±0.003	0.006±0.003
DAR, $\rho = 2.0$	0.149±0.008	0.105±0.009	0.012±0.004	0.012±0.004

Table 37: Within-group reward variance at $B = 16$, as mean ± SE. Counts follow Table 4. Bold marks the largest mean in each column, including ties at the displayed precision.

	Llama		Qwen	
Method	GSM8K	MATH-500	GSM8K	MATH-500
Independent	0.1±0.6	0.1±0.6	0.1±0.1	0.1±0.1
Regeneration	−63.8±2.6	−34.5±3.2	−41.1±3.9	−9.5±2.4
Arithmetic	0.4±0.8	−0.2±1.0	0.0±0.2	−0.1±0.2
Stochastic beam	0.2±0.9	0.5±0.9	−0.7±0.8	−0.1±0.5
Verbalized	−11.2±2.1	−16.2±2.1	0.2±1.9	−19.5±3.5
MMR, $\lambda = 0.3$	−5.3±1.1	−5.3±1.1	−0.6±0.4	−0.4±0.2
MMR, $\lambda = 0.5$	−0.2±1.0	−1.0±1.2	−0.4±0.3	−0.1±0.2
MMR, $\lambda = 0.7$	3.7±1.0	4.3±1.2	−0.3±0.4	0.3±0.3
k-DPP	−2.1±1.0	−2.8±1.1	0.0±0.3	0.0±0.2
Release, $\rho = 0$	−0.1±0.6	−0.0±0.5	0.1±0.1	0.1±0.1
Release, $\rho = 1.0$	−12.5±1.3	−12.0±1.3	0.0±0.2	−0.5±0.3
Release, $\rho = 2.0$	−23.7±1.8	−21.5±2.0	−0.2±0.5	−1.3±0.5
DAR, $\rho = 0$	−0.3±0.7	0.5±0.6	0.0±0.1	0.0±0.1
DAR, $\rho = 1.0$	−12.5±1.4	−11.1±1.4	−0.2±0.2	−0.3±0.3
DAR, $\rho = 2.0$	−23.9±1.9	−21.8±2.1	−0.2±0.5	−1.0±0.4

Table 38: GRPO groups, $B = 16$. Mean response accuracy minus the correctness rate of the comparison set, in percentage points ± SE. Positive values indicate higher response accuracy than the comparison set.

Tables 39 and 40 report all distortion settings at $B = 16$.

Method	ρ	Constant	Accuracy	$\widehat{\mathrm{Div}}$	$\widehat{D}$
Release 64	0	21.3±2.6	56.9±2.4	1.003±0.025	0.903±0.014
Release 64	0.5	19.4±2.5	53.7±2.4	1.152±0.029	0.966±0.014
Release 64	1	16.2±2.4	44.7±2.4	1.512±0.047	1.201±0.017
Release 64	2	13.4±2.3	34.3±2.2	2.023±0.097	1.494±0.028
Release 64	3	13.4±2.3	34.4±2.2	2.023±0.097	1.494±0.028
Release 64	4	13.6±2.4	34.3±2.2	2.023±0.097	1.494±0.028
Release 64	5	13.4±2.3	34.3±2.2	2.023±0.097	1.494±0.028
Release 128	0	17.1±2.3	56.6±2.4	1.041±0.016	0.879±0.014
Release 128	0.5	15.3±2.3	53.4±2.4	1.189±0.019	0.945±0.013
Release 128	1	12.2±2.1	44.2±2.4	1.563±0.032	1.201±0.013
Release 128	2	10.1±2.0	28.8±2.1	2.357±0.101	1.647±0.027
Release 128	3	9.2±2.0	24.4±1.7	2.717±0.147	1.810±0.042
Release 128	4	9.4±2.0	24.4±1.7	2.717±0.147	1.810±0.042
Release 128	5	9.3±2.0	24.4±1.7	2.717±0.147	1.809±0.042
DAR 64	0	21.4±2.7	57.1±2.4	1.012±0.028	0.925±0.014
DAR 64	0.5	20.0±2.7	53.8±2.4	1.167±0.034	0.990±0.015
DAR 64	1	17.7±2.6	45.2±2.4	1.514±0.054	1.197±0.019
DAR 64	2	18.7±2.7	34.1±2.2	2.005±0.097	1.495±0.029
DAR 64	3	17.4±2.7	35.0±2.2	1.992±0.098	1.468±0.030
DAR 64	4	17.4±2.7	35.1±2.2	1.991±0.098	1.468±0.030
DAR 64	5	17.4±2.7	35.1±2.2	1.991±0.098	1.468±0.030
DAR 128	0	21.4±2.7	57.1±2.4	1.014±0.029	0.924±0.015
DAR 128	0.5	19.9±2.7	53.9±2.4	1.163±0.035	0.988±0.015
DAR 128	1	17.4±2.6	44.9±2.4	1.519±0.053	1.208±0.017
DAR 128	2	24.4±3.0	28.5±2.1	2.337±0.102	1.655±0.029
DAR 128	3	21.0±2.9	24.8±1.7	2.685±0.148	1.798±0.044
DAR 128	4	21.0±2.9	24.8±1.7	2.685±0.148	1.798±0.044
DAR 128	5	21.0±2.9	24.8±1.7	2.685±0.148	1.798±0.044

Table 39: Llama mathematical sweep at $B = 16$. Task scores are percentages, with SE in percentage points, averaged over GSM8K and MATH-500 with equal weight.

Method	ρ	Constant	Accuracy	$\widehat{\mathrm{Div}}$	$\widehat{D}$
Release 64	0	94.2±1.5	95.6±1.4	1.043±0.229[30]	0.942±0.103[30]
Release 64	0.5	93.4±1.6	95.5±1.4	1.278±0.267[30]	0.977±0.110[30]
Release 64	1	92.2±1.8	95.2±1.4	1.992±0.479[30]	1.198±0.132[30]
Release 64	2	91.5±2.0	94.7±1.4	3.489±0.867[30]	1.679±0.200[30]
Release 64	3	91.5±2.0	94.7±1.4	3.489±0.867[30]	1.679±0.200[30]
Release 64	4	91.5±2.0	94.7±1.4	3.489±0.867[30]	1.679±0.200[30]
Release 64	5	91.5±2.0	94.7±1.4	3.489±0.867[30]	1.679±0.200[30]
Release 128	0	94.3±1.4	95.5±1.4	0.986±0.139[30]	0.924±0.087[29]
Release 128	0.5	93.6±1.6	95.4±1.4	1.243±0.192[30]	0.964±0.085[29]
Release 128	1	92.5±1.8	95.2±1.4	1.813±0.288[30]	1.169±0.088[29]
Release 128	2	91.0±2.0	94.5±1.4	3.868±0.692[30]	1.862±0.148[29]
Release 128	3	90.5±2.1	94.1±1.4	5.434±1.057[30]	2.318±0.220[29]
Release 128	4	90.5±2.1	94.1±1.4	5.434±1.057[30]	2.318±0.220[29]
Release 128	5	90.5±2.1	94.1±1.4	5.434±1.057[30]	2.318±0.220[29]
DAR 64	0	95.0±1.5	95.5±1.4	0.834±0.227[30]	0.932±0.106[30]
DAR 64	0.5	94.4±1.5	95.4±1.4	0.957±0.214[30]	0.938±0.103[30]
DAR 64	1	94.3±1.6	95.2±1.4	1.526±0.459[30]	1.122±0.141[30]
DAR 64	2	91.5±2.0	94.8±1.4	3.242±0.850[30]	1.605±0.195[30]
DAR 64	3	91.5±2.0	94.8±1.4	3.308±0.875[30]	1.618±0.200[30]
DAR 64	4	91.5±2.0	94.8±1.4	3.308±0.875[30]	1.618±0.200[30]
DAR 64	5	91.5±2.0	94.8±1.4	3.308±0.875[30]	1.618±0.200[30]
DAR 128	0	95.0±1.5	95.5±1.4	0.809±0.239[30]	0.928±0.107[30]
DAR 128	0.5	94.7±1.5	95.4±1.4	1.100±0.338[30]	1.018±0.122[30]
DAR 128	1	92.2±1.8	95.2±1.4	2.028±0.489[30]	1.235±0.122[30]
DAR 128	2	91.3±2.0	94.7±1.4	3.666±0.753[30]	1.782±0.173[30]
DAR 128	3	90.5±2.1	94.1±1.4	5.421±1.058[30]	2.278±0.214[30]
DAR 128	4	90.5±2.1	94.1±1.4	5.421±1.058[30]	2.278±0.214[30]
DAR 128	5	90.5±2.1	94.1±1.4	5.421±1.058[30]	2.278±0.214[30]

Table 40: Qwen mathematical sweep at $B = 16$. Task scores are percentages, with SE in percentage points, averaged over GSM8K and MATH-500 with equal weight. Superscripts give the number of problems when fewer than 200 are used. Pools have zero final-answer variance on 170 of the 200 problems, so diversity and distortion use the remaining 30. For Release with $N = 128$, distortion excludes one further problem whose first 128 usable responses overlap the comparison set.