GRADIENTS MUST EARN THEIR INFLUENCE: UNIFYING SFT WITH GENERALIZED ENTROPIC OBJECTIVES

Anonymous authors
Paper under double-blind review

ABSTRACT

Standard negative log-likelihood (NLL) for Supervised Fine-Tuning (SFT) applies uniform token-level weighting. This rigidity creates a two-fold failure mode: (i) overemphasizing low-probability targets can amplify gradients on noisy supervision and disrupt robust priors, and (ii) uniform weighting provides weak sharpening when the model is already confident. Existing methods fail to resolve the resulting plasticity–stability dilemma, often suppressing necessary learning signals alongside harmful ones. To address this issue, we unify token-level SFT objectives within a generalized deformed-log family and show that they all share a *gate* $\times$ *error* gradient structure, where the gate controls how much the model trusts its current prediction. We further propose two gate designs according to the predictive information they consume: *probability-level* and *distribution-level*. At the probability level, by employing the Cayley transform, we map the model's continuously evolving uncertainty onto a continuous focus trajectory, enabling a seamless interpolation between uncertain novel concepts and well-established knowledge. At the distribution level, we propose **Dynamic Entropy Fine-Tuning (DEFT)**, which gates on the concentration of the predictive distribution to distinguish diffuse uncertainty from concentrated disagreement. Across seven backbones and three capability regimes, both methods improve over the standard NLL objective without per-task tuning. DEFT achieves the highest mean task score in different task configurations, suggesting that distribution-level information is a substantive and underexplored dimension of token-level SFT.

1 INTRODUCTION

Under the standard negative log-likelihood (NLL) objective used in SFT of LLMs, low-probability targets can represent either useful new supervision or potential conflicts with current predictions Diao et al. (2026); Chen et al. (2025); Li et al. (2025). NLL uses a unit token gate, so it does not explicitly distinguish these cases, although its gradients still depend on the predictive error. Figure 1 makes this symptom concrete: in a standard math-SFT run, roughly 25% of tokens are forgotten over both steps 0–10 and 0–100. Moreover, as training proceeds, forgetting increasingly concentrates on *high*-confidence prior-knowledge tokens. The root cause is the uniform token weighting of NLL. At low target probability p, it indiscriminately amplifies two populations that demand opposite treatment (learnable knowledge gaps, which should be pursued, and *confident conflicts*, i.e., high-confidence predictions that disagree with the label, which should be suppressed), inducing large parameter updates that trigger catastrophic forgetting Diao et al. (2026); Wu et al. (2025a); Taheri et al. (2025).At high p, the same uniform gate decays gradients linearly with the error $1-p$, stalling entropy reduction on already-correct trajectories. These two regimes form a single plasticity–stability dilemma: NLL cannot sharpen the confident without also amplifying the conflicting.

Existing remedies address only one horn of this dilemma. Offline data selection and token masking (Pang et al., 2025; Liu et al., 2026; Ruan et al., 2025)raise the signal-to-noise ratio but rely on static heuristics that cannot track the model's state during training. Online methods adapt token weights during training: InfoSFT uses clipped current-probability weighting (Sabbaghi et al., 2026), EAFT uses distribution-level entropy (Diao et al., 2026), and RankTuner combines target rank with distributional uncertainty for probability–entropy calibration (Yu et al., 2026). However,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

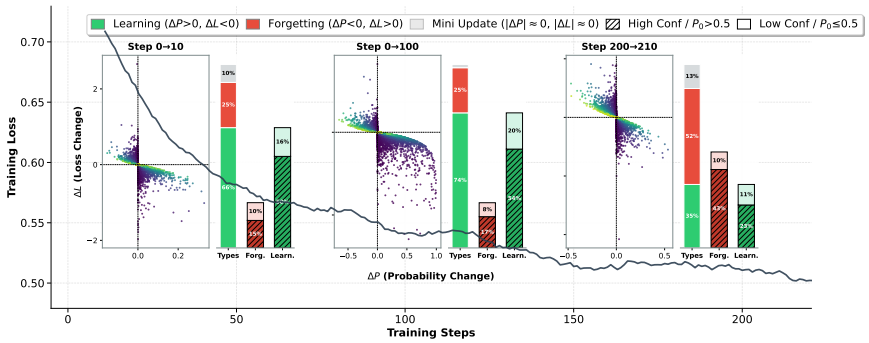

Figure 1: Learning conflicts of different tokens under NLL during training (token-wise learning and forgetting). The scatter plot shows token probability ($\triangle P$) and loss ($\triangle L$) changes. Bar plots report token proportions for overall learning and forgetting, forgotten tokens by confidence (Q2 in the scatter plot), and learned tokens by confidence (Q4 in the scatter plot). Hatched patterns indicate high-confidence tokens.

RankTuner requires task-specific tuning, while EAFT and InfoSFT still struggle to alleviate the learning–sharpening dilemma. Therefore, a question is: How can we design an SFT objective that preserves coverage for uncertain knowledge gaps, efficiently sharpens confident predictions, and suppresses confident conflicts, *without* expensive offline token filtering?

We address this question by unifying token-level SFT objectives within a deformed-log family based on the Tsallis q-logarithm (Tsallis, 1988), which exposes a universal *gate* $\times$ *error* gradient structure in which the gate decides how much the model trusts its current prediction. Unlike other online methods, our design places the adaptive focus index inside the loss as an exponent, rather than specifying only an external CE weight. With the focus index detached, the induced update also admits a weighted-CE representation; the distinction is the focus-rule construction, not a separate class of gradients. This view reveals that no single static gate simultaneously delivers coverage at low confidence and sharpening at high confidence. Within this unified view, we further classify trust gates by the predictive information they consume—*probability-level* versus *distribution-level*, and instantiate one parameter-free method on each axis. **Cayley-Trans** derives a state-dependent focusing trajectory using geometric principles and the Cayley transform, yielding a dynamic trust gate that smoothly transitions from NLL-like behavior when the model is uncertain to probability-loss-like behavior when the model is confident. **Dynamic Entropy Fine-Tuning (DEFT)** extends the gate to the full predictive distribution, using the distribution concentration as its trust signal to separate predictive states that any probability-level gate collapses into a single value (e.g., diffuse uncertainty versus concentrated disagreement).

Empirical validation. Following Li et al. (2025), we partition the training data into three regimes based on the model's domain-specific priors: Model-Strong, Model-Intermediate, and Model-Weak. Across seven backbones and multiple domain-specific benchmarks, both Cayley-Trans and DEFT deliver superior overall performance across all three regimes, adaptively matching the best regime-specific static-gate behavior without task-specific tuning. DEFT achieves the highest mean task score in all three regimes. **Our main contributions are:**

a. We unify token-level SFT objectives within a deformed-log family with a *gate* $\times$ *error* gradient structure, and classify gates by the information they consume—*probability and distribution-level*
b. We propose two parameter-free dynamic gating mthods: Cayley-Trans, a probability-level method that introduces a Cayley-induced focusing trajectory to balance coverage and sharpening; and DEFT, a distribution-level method that modulates the focus index via distribution concentration.
c. Across diverse settings, Cayley-Trans and DEFT consistently attain superior paradigm-level averages, with DEFT achieving the highest mean task score in all three regimes.

2 Learning Signal Allocation in SFT

In the post-pretraining stage, the model is initialized from a pre-trained base $p_{\boldsymbol{\theta}_0}$ that already encodes rich domain prior knowledge. Given that the supervision signals often contain noise or are suboptimal,

the core challenge of SFT lies in how the objective function addresses the mismatch between token-level optimization granularity and the model's intrinsic knowledge state. Let an SFT example be a pair $(x, \tilde{\mathbf{y}})$ with target sequence $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_N)$, unrolled into token-level pairs via the context

$$c_t \triangleq (x, \tilde{y}_{<t}), \qquad \tilde{y}_t \in \mathcal{V}, \tag{1}$$

inducing a token-level distribution T over $(c, \tilde{y})$. The model emits logits $z_\theta(c) \in \mathbb{R}^{|\mathcal{V}|}$ and a next-token distribution $p_\theta(\cdot \mid c) = \mathrm{softmax}(z_\theta(c))$; we write $p \triangleq p_\theta(\tilde{y} \mid c) \in (0, 1]$ for the supervised-target probability.

2.1 Token-Level Objectives and Gradient Structure

For any differentiable, nonincreasing function $f : [0, 1] \to \mathbb{R}$, we define the token-level SFT objective as:

$$\mathcal{L}_f(\boldsymbol{\theta}) = \mathbb{E}_{(c,\tilde{y})\sim T}[f(p_{\boldsymbol{\theta}}(\tilde{y} \mid c))]\,. \tag{2}$$

The standard NLL corresponds to $f(p) = -\log p$. For a single supervised token, the softmax-logit gradient is

$$\frac{\partial f(p)}{\partial z_i(c)} = s_f(p)\big(p_{\boldsymbol{\theta}}(i \mid c) - \delta_{i,\tilde{y}}\big), \qquad s_f(p) \triangleq -f'(p)p \geq 0. \tag{3}$$

The target-logit gradient magnitude is therefore

$$W_f(p) \triangleq \left|\frac{\partial f(p)}{\partial z_{\tilde{y}}(c)}\right| = s_f(p)(1 - p). \tag{4}$$

Eq. (4) exposes a universal gate $\times$ error structure: $(1 - p)$ measures prediction error, while $s_f(p)$ acts as a confidence-dependent *gate* that decides how much the model should trust (and learn from) the current token. Detailed proofs are provided in Appendix F.

2.2 The Dual Nature of Low Probability: Coverage vs. Sharpening

For NLL, $f(p) = -\log p$ implies $f'(p) = -1/p$ and therefore $s_f(p) = 1$, the target-logit gradient is

$$\frac{\partial\,(-\log p)}{\partial z_{\tilde{y}}} = -(1 - p). \tag{5}$$

This induces an intrinsic *uniform gating* property: the gate $s_f(p) = 1$ is constant across all confidence levels, maximizing update magnitude precisely when the model is least confident, while providing negligible signal for high-confidence predictions. However, low target probabilities during post-training are inherently ambiguous, arising either from a *learnable gap* (knowledge the model has yet to acquire) or an *adversarial conflict* (a confident prior that disagrees with the supervision target).

By allocating gradient magnitude inversely to confidence, NLL guarantees that no low-probability target is left behind, forcing the model to overcome prior ignorance and absorb domain-specific facts or reasoning steps that lie outside the pretrained support. However, this same uniform allocation becomes a liability in two regimes. (i) *Vanishing sharpening at high confidence*: once p approaches 1, the gradient decays linearly with the error $(1 - p)$, yielding diminishing returns and stalling entropy reduction even on already-correct trajectories. (ii) *Indiscriminate amplification at low confidence*: lacking any signal that separates learnable gaps from adversarial conflicts, NLL amplifies both equally; in regimes with strong priors, the resulting tail-noise gradients overwhelm high-quality tokens and corrode established knowledge. As shown in Figure 1, token-level conflicts persist throughout training (upper-left scatter cluster, "Forg" histogram column), the forgotten-token fraction grows from steps 0–10 to 0–100, and pronounced inter-step fluctuations (200–210) further amplify forgetting—stalling entropy reduction and impeding convergence.

3 Trust Gating with the Deformed-Log Family

Section 2 motivates comparing confidence-dependent gates. We use the deformed-log/GCE family to organize these rules and characterize the Bayes-optimal predictor when the focus parameter α is held constant.

3.1 TRUST GATING VIA THE DEFORMED-LOG FAMILY

We formalize a one-parameter family of token losses based on the Tsallis q-logarithm (Tsallis, 1988), which provides the cleanest parameterization of the gate term.

Definition 1 (q-logarithm) . *For $q \in \mathbb{R}$, $x > 0$, the q-logarithm is $\ln_q(x) \triangleq \frac{x^{1-q}-1}{1-q}$, with $\ln_q(x) \to \ln x$ as $q \to 1$.*

Negative q-log likelihood. For a target-token probability $p \in (0, 1]$, we instantiate Def. 1 with $q = 1 - \alpha$ for a single *focus index* $\alpha \geq 0$ and obtain the token loss in closed form:

$$\mathcal{L}_\alpha(p) = \frac{1 - p^\alpha}{\alpha}, \tag{6}$$

This loss recovers standard NLL as $\alpha \to 0$, the linear probability loss ($-p$) at $\alpha = 1$, for $0 < \alpha < 1$ it coincides with the generalized cross-entropy (GCE) family studied for label-noise robustness (Zhang & Sabuncu, 2018). The parameter α thus acts as a *focus index*, smoothly interpolating between NLL-like coverage at small α and high-confidence emphasis at large α.

Gradient decomposition. For the focus-indexed loss $\mathcal{L}_\alpha(p) = \frac{1-p^\alpha}{\alpha}$, the target-logit gradient is:

$$\frac{\partial \mathcal{L}_\alpha(p)}{\partial z_{\tilde{y}}} = -\, p^\alpha\, (1 - p). \tag{7}$$

This factorizes into a *trust gate* and a *prediction error*:

$$\underbrace{p^\alpha}_{\text{Gate: } \mathcal{G}(p)} \quad \times \quad \underbrace{(1 - p)}_{\text{Error}}. \tag{8}$$

The gate controls how much the model trusts its current prediction: The gate sweeps from nearly open at $\alpha \to 0$ (full gradient on low-confidence tokens) to progressively closed as α grows (linear at $\alpha = 1$, more aggressive at $\alpha > 1$), shifting signal toward high-confidence regions.

3.2 BAYES-OPTIMAL PREDICTION AND ENTROPY INTERPRETATION

The choice of α determines which generalized entropy the learning system preferentially reduces , thereby specifying the *geometric space* in which optimization occurs. To characterize how the focus parameter shapes the preferred prediction, we first hold α constant and minimize the population risk. The resulting optimum connects the loss's confidence preference to an entropy functional of the target distribution.

Theorem 1 (Bayes optimum for a constant focus parameter) . *At a fixed context, let r be a target distribution and minimize $R_\alpha(\hat{p}; r) = \mathbb{E}_{y\sim r}[\mathcal{L}_\alpha(\hat{p}_y)]$ over the closed probability simplex. For $0 < \alpha < 1$, set $\beta = (1 - \alpha)^{-1} > 1$ and $Z_\beta = \sum_y r_y^\beta$. The unique minimizer and Bayes risk are*

$$\hat{p}^\star_y = \frac{r_y^\beta}{Z_\beta}, \qquad B_\alpha(r) = \frac{1 - Z_\beta^{1/\beta}}{\alpha} = \frac{1 - e^{-\alpha H_\beta(r)}}{\alpha}, \tag{9}$$

where $H_\beta(r) = (1 - \beta)^{-1} \log Z_\beta$ is Rényi entropy. At $\alpha = 0$, NLL is strictly proper with $\hat{p}^\star = r$ and $B_0 = H(r)$. At $\alpha = 1$, any distribution supported on $\arg\max_y r_y$ is optimal, with risk $1 - \max_y r_y$. For $\alpha > 1$, only one-hot predictors at a mode are optimal, with risk $(1 - \max_y r_y)/\alpha$.

For $0 < \alpha < 1$, the escort optimum amplifies the relative mass of more probable targets: $\hat{p}^\star_i / \hat{p}^\star_j = (r_i/r_j)^\beta$ on the support of r. Increasing α increases β, strengthening this preference. The Bayes risk is a monotone transform of Rényi-β entropy, recovering the Shannon limit as $\alpha \to 0$ and approaching $1 - \max_y r_y$ as $\alpha \to 1^-$. This characterization connects a constant focus parameter to the preferred prediction distribution. We next use this perspective to design state-dependent focus rules, whose updates are specified in Section 4.3 . Proofs and boundary cases are in Appendix G .

4 UNIFYING COVERAGE AND SHARPENING

Section 3 reveals an inherent tension: any static α is brittle—coverage-oriented choices help learn hard positives but overfit conflicts, whereas sharpening-oriented choices stabilize updates but under-learn genuine gaps—motivating a *state-dependent* gate, classified by the information it consumes:

- **Probability-level trust gating**: the gate depends only on the supervised-target probability p.
- **Distribution-level trust gating**: the gate depends on statistics of the full predictive distribution $p_\theta(\cdot \mid c)$, exploiting information beyond p alone.

4.1 PROBABILITY-LEVEL STATE-DEPENDENT GATING: THE CAYLEY TRANSFORM

We derive $\alpha(p)$ by imposing geometric boundary conditions at the two extremes of confidence.

Boundary Conditions. To recover NLL-like coverage as $p \to 0$, we impose the *Coverage Anchor* $\alpha(0) = 0$ together with $\alpha(p)|\log p| \to 0$, which ensures that the gate $p^{\alpha(p)} \to 1$. At $p \to 1$, we impose the *Sharpening Anchor* $\alpha(1) = 1$, approaching the linear-probability member and placing greater relative emphasis on higher-probability targets.

Derivation via Conformal Mapping. To connect these anchors, we parameterize the model's uncertainty using the intrinsic Fisher–Rao radius $z \triangleq \sqrt{1-p} \in [0, 1]$. We seek a smooth mapping $\alpha(z)$ that transforms this uncertainty radius z to the focus index α, satisfying the boundary conditions: $z = 1 \implies \alpha = 0$ (max uncertainty $\to$ coverage) and $z = 0 \implies \alpha = 1$ (certainty $\to$ sharpening). We connect these endpoints using the Cayley transform:

$$\alpha(z) \;=\; \frac{1-z}{1+z}. \tag{10}$$

Substituting $z = \sqrt{1-p}$ yields the *Cayley-Trans* focus trajectory:

$$\alpha^*(p) \;=\; \frac{1-\sqrt{1-p}}{1+\sqrt{1-p}}. \tag{11}$$

This closed-form solution is parameter-free: near $p \approx 0$, $\alpha^*(p) \approx 0$, prioritizing coverage; as p increases, $\alpha^*(p)$ smoothly ascends to 1, prioritizing sharpening. Notably, $\alpha^*(p)$ admits an information-theoretic identity; hence, the coverage–sharpening trade-off can also be characterized by the information content of the current prediction error (Prop. 5, Appendix H). Plugging $\alpha^*(p)$ into the deformed-log loss in Eq. (6), with $\alpha^*(p)$ treated as a stop-gradient quantity, defines the Cayley-Trans objective. Unlike fixed-α/GCE, Cayley-Trans requires no task-tuned focus index; unlike $-p$, its gate approaches the NLL gate as $p \to 0$.

4.2 DISTRIBUTION-LEVEL TRUST GATING: DEFT

At the same target probability p, a prediction can concentrate its remaining mass on one competitor or spread it over many alternatives. A probability-only gate cannot distinguish these states, motivating a focus index based on the full predictive distribution.

Concentration-conditioned gate. Write $P_v = p_\theta(v \mid c)$ and $p = P_{\tilde{y}}$. With natural logarithms, $H_2(P) = -\log \sum_v P_v^2$. We use

$$\alpha_{\mathrm{DEFT}}(c) = C(P) = e^{-H_2(P)} = \sum_{v \in \mathcal{V}} P_v^2, \qquad \mathcal{G}_{\mathrm{DEFT}} = p^{C(P)}. \tag{12}$$

Accordingly, as $\alpha_{\mathrm{DEFT}} \to 0$, the deformed-log objective approaches NLL; as $\alpha_{\mathrm{DEFT}} \to 1$, it approaches the linear probability loss $1 - p$, which is equivalent to $-p$ up to an additive constant. For a finite vocabulary, $1/|\mathcal{V}| \le C(P) \le 1$. This statistic adds no trainable parameters or optimizer state; its actual runtime and memory depend on how the vocabulary reduction is implemented. We freeze $C(P)$ during differentiation. Since $p \le p^{C(P)} \le 1$, the target-logit magnitude satisfies $p(1-p) \le W_{\mathrm{DEFT}} \le 1 - p$: DEFT reallocates updates relative to NLL rather than increasing their individual magnitudes.

What information lies beyond the target probability? For $0 < p < 1$, normalize the non-target probabilities as $\bar{P}_v = P_v/(1-p)$ for $v \neq \tilde{y}$, and let $S_{\text{tail}} = \sum_{v \neq \tilde{y}} \bar{P}_v^2$. Then

$$C(P) = p^2 + (1-p)^2 S_{\text{tail}}, \qquad \left.\frac{\partial \log \mathcal{G}_{\text{DEFT}}}{\partial S_{\text{tail}}}\right|_p = (1-p)^2 \log p < 0. \tag{13}$$

Thus, holding p fixed, more concentrated competitors yield stronger suppression. This dependence is present throughout $0 < p < 1$, while its log-gate sensitivity smoothly diminishes and vanishes only in the limit $p \to 1$. Further properties are in Appendix H.3. This decomposition identifies the additional predictive information used by DEFT.

4.3 Stop-Gradient Implementation

For both methods, we compute a state-dependent focus index at iteration k, $a_k = \alpha^*(p_{\theta_k})$ for Cayley-Trans and $a_k = C(P_{\theta_k})$ for DEFT, and stop its gradient during the corresponding backward pass. With sg denoting stop-gradient, the per-token objective and logit gradient are

$$\ell_k(\theta) = \frac{1 - p_\theta(\tilde{y} \mid c)^{\text{sg}[a_k]}}{\text{sg}[a_k]}, \qquad \frac{\partial \ell_k}{\partial z_i} = p^{a_k}(P_i - \delta_{i,\tilde{y}}), \tag{14}$$

with the continuous NLL extension at $a_k = 0$. Thus, the adaptive focus rescales the token-level CE gradient while preserving its direction. Without detachment, the two methods are affected differently: for Cayley-Trans, differentiating through $\alpha^*(p)$ changes the probability-dependent gradient magnitude but preserves each token's CE direction; for DEFT, differentiating through $C(P)$ introduces an additional distribution-level term that can change the logit-gradient direction by coupling the target update to all vocabulary probabilities. The complete derivation is provided in Appendix H.4, and Table 4 evaluates both variants.

5 Main Experiments

We empirically evaluate Cayley-Trans and DEFT along the Model Capability Continuum of Li et al. (2025), spanning Model-Strong, Model-Intermediate, and Model-Weak regimes defined by the alignment between pretraining priors and target tasks.

5.1 Experimental Setup

We instantiate each capability regime with a domain-specific dataset: the Model-Strong (MS) regime uses NuminaMath (LI et al., 2024); the Model-Intermediate (MI) regime uses the medical reasoning corpus M1-23K (Huang et al., 2025); and the Model-Weak (MW) regime uses synthetic FigFont character-art puzzles (Stojanovski et al., 2025). We span seven backbones to cover both the math-specialized and general-purpose pre-training axes; the full backbone list, training datasets, and optimization hyper-parameters are deferred to Appendix J.3. All primary multi-seed experiments use five random seeds, and we report their mean and standard deviation. We compare Cayley-Trans and DEFT with seven representative objectives. $-\log p$ serves as a baseline for undifferentiated full-data learning. $-p$ is the linear probability scaling loss, which mitigates noise while preserving the model prior by down-weighting low-probability labels (Li et al., 2025; Wu et al., 2025b). *Hellinger* (α=0.5) and *GCE* (α=0.7) (Zhang & Sabuncu, 2018) are two static instances of the deformed-log family $\mathcal{L}_\alpha(p) = (1-p^\alpha)/\alpha$, lying between the two endpoints above. We additionally include EAFT (Diao et al., 2026), InfoSFT (Sabbaghi et al., 2026), and RankTuner (Yu et al., 2026), which use entropy, confidence, or probability/rank-based token weighting. Baseline implementation details and reproducibility requirements are summarized in Appendix J.4.

5.2 Main Results

Model-Strong Results. Table 1 (MS block) reports the regime-averaged Avg@16 on five MATH-style benchmarks (additional-backbone per-benchmark scores in Table 8). When the prior already covers the domain, the optimization burden shifts from acquisition to *prior preservation*: most down-weighting objectives improve on $-\log p$ in the primary comparison, and DEFT delivers the

Table 1: **Model-Strong (math) per-benchmark results**: multi-seed mean ± standard deviation of Avg@16 accuracy. The best Avg result is in **bold** and the second-best Avg result is underlined.

Model	Method	AIME24	AMC23	Math500	Minerva	Olympiad	Avg
LLaMA-3.1-8B	$-\log p$	0.25 ± 0.15	6.25 ± 0.40	17.23 ± 0.12	6.06 ± 0.13	3.28 ± 0.13	6.61 ± 0.04
	$-p$	0.76 ± 0.27	10.55 ± 1.18	23.87 ± 0.44	10.20 ± 0.52	5.57 ± 0.27	10.19 ± 0.29
	Hellinger	0.41 ± 0.23	11.20 ± 1.34	25.99 ± 0.54	10.17 ± 0.40	6.53 ± 0.02	$\underline{10.86 \pm 0.08}$
	GCE	0.41 ± 0.47	10.12 ± 0.85	26.12 ± 0.21	10.61 ± 0.21	6.44 ± 0.38	10.74 ± 0.17
	EAFT	0.17 ± 0.08	6.59 ± 0.59	17.45 ± 0.35	5.99 ± 0.35	3.34 ± 0.21	6.71 ± 0.17
	InfoSFT	0.74 ± 0.28	7.97 ± 0.68	22.91 ± 0.43	8.77 ± 0.66	5.41 ± 0.15	9.16 ± 0.26
	RankTuner	0.48 ± 0.36	10.42 ± 1.13	26.17 ± 0.50	10.54 ± 0.55	6.30 ± 0.06	10.78 ± 0.20
	Cayley-Trans	0.55 ± 0.26	10.00 ± 0.77	25.39 ± 0.63	10.25 ± 0.44	6.29 ± 0.10	10.50 ± 0.17
	DEFT	0.69 ± 0.48	12.24 ± 0.62	26.15 ± 0.59	9.88 ± 0.40	6.27 ± 0.24	$\mathbf{11.04 \pm 0.31}$
Qwen2.5-Math-7B	$-\log p$	2.45 ± 0.92	25.44 ± 0.42	53.43 ± 0.76	19.64 ± 0.64	18.06 ± 0.36	23.80 ± 0.20
	$-p$	8.96 ± 0.99	41.97 ± 2.54	68.52 ± 0.70	29.48 ± 1.22	31.66 ± 0.63	$\underline{36.12 \pm 0.52}$
	Hellinger	6.92 ± 0.78	39.84 ± 1.70	66.81 ± 0.46	33.07 ± 1.02	31.40 ± 0.24	35.61 ± 0.44
	GCE	7.55 ± 0.94	41.78 ± 1.31	68.33 ± 0.35	30.84 ± 1.01	32.10 ± 0.21	$\underline{36.12 \pm 0.45}$
	EAFT	2.32 ± 0.54	24.06 ± 1.08	52.12 ± 0.52	19.68 ± 0.61	17.63 ± 0.36	23.16 ± 0.34
	InfoSFT	5.62 ± 1.13	35.33 ± 1.36	62.16 ± 0.62	29.04 ± 0.87	27.29 ± 0.47	31.89 ± 0.40
	RankTuner	7.87 ± 0.81	40.23 ± 0.89	67.92 ± 0.25	31.97 ± 0.24	32.08 ± 0.08	36.01 ± 0.25
	Cayley-Trans	5.87 ± 0.48	39.66 ± 0.95	65.98 ± 0.43	31.97 ± 0.72	30.02 ± 0.38	34.70 ± 0.20
	DEFT	7.87 ± 0.77	41.88 ± 1.25	68.72 ± 0.77	32.95 ± 0.80	32.22 ± 0.21	$\mathbf{36.73 \pm 0.21}$

Table 2: **Model-Intermediate (medical) results**. Per-benchmark entries are multi-seed means; Avg reports mean ± standard deviation of accuracy. The best Avg result is in **bold** and the second-best Avg result is underlined.

Model	Method	GPQA	Lancet	MMLU-P	MedB(4)	MedB(5)	MedMC	MedQA	MedX	NEJM	PubMedQA	Avg
LLaMA-3.1-8B	$-\log p$	38.20	58.74	53.47	51.36	44.42	54.59	60.94	15.53	58.21	75.00	51.05 ± 0.36
	$-p$	34.92	57.96	53.38	48.18	43.70	53.00	57.38	12.99	59.37	76.68	49.76 ± 0.44
	Hellinger	37.13	57.77	53.78	51.88	46.88	55.03	59.94	14.73	58.27	76.86	51.23 ± 0.43
	GCE	39.64	58.49	53.58	51.82	46.04	54.82	59.47	14.99	58.34	76.60	$\underline{51.38 \pm 0.35}$
	EAFT	38.14	57.41	54.15	52.60	45.45	54.64	61.27	14.91	59.33	74.33	51.22 ± 0.22
	InfoSFT	38.61	58.64	53.55	50.65	47.08	54.89	61.07	14.66	57.08	76.80	51.30 ± 0.67
	RankTuner	39.30	57.71	53.26	50.49	44.40	53.83	60.49	15.13	58.09	75.53	50.82 ± 0.55
	Cayley-Trans	38.91	58.37	53.48	53.09	44.32	55.31	60.70	14.73	57.96	75.45	51.23 ± 0.43
	DEFT	42.39	58.90	54.18	53.46	47.62	54.98	60.33	15.04	58.71	76.17	$\mathbf{52.18 \pm 0.51}$
Qwen2.5-Math-7B	$-\log p$	34.72	37.47	36.30	29.35	25.91	36.29	33.43	10.29	27.30	71.84	34.29 ± 0.61
	$-p$	31.54	37.92	38.32	33.28	25.97	36.77	34.92	10.82	26.70	70.45	34.67 ± 0.45
	Hellinger	31.28	36.53	37.33	32.30	26.22	36.61	33.54	9.92	28.32	70.58	34.27 ± 0.57
	GCE	34.87	35.92	37.20	31.98	26.79	36.67	33.74	10.25	28.73	69.72	34.59 ± 0.25
	EAFT	33.59	37.38	36.82	30.84	25.65	36.20	34.03	10.51	27.16	72.47	34.46 ± 0.26
	InfoSFT	32.24	36.89	36.70	33.93	27.59	36.31	34.11	10.25	28.40	70.62	34.70 ± 0.72
	RankTuner	33.91	38.29	35.53	29.95	28.17	36.39	33.58	10.51	26.16	71.90	34.44 ± 0.52
	Cayley-Trans	34.81	36.47	36.64	34.90	28.33	36.55	33.37	10.54	26.99	70.97	$\underline{34.95 \pm 0.53}$
	DEFT	36.98	37.07	36.50	32.88	26.30	36.05	34.47	10.92	28.93	70.40	$\mathbf{35.05 \pm 0.48}$

highest backbone-averaged score. This pattern aligns with prior observations that in prior-driven regimes, suppressing the gradient of low-probability tokens while exploiting high-confidence ones is the dominant lever (Li et al., 2025; Wu et al., 2025b). Among the down-weighting gates, the static deformed-log instances already realize most of the available gain over $-\log p$, while DEFT achieves a higher backbone-averaged score than $-p$ under both model settings—all without inheriting the brittleness $-p$ exhibits in the other regimes (cf. MW below).

Model-Intermediate Results. Table 2 reports the average accuracy on 10 medical benchmarks (additional-backbone per-benchmark scores in Table 9). The performance differences among loss functions are relatively small. Here the prior is partially aligned with the target distribution, and the spread among objectives is small: neither pure prior sharpening nor undifferentiated coverage dominates, since each backbone calls for a different mix of high-confidence reinforcement and low-confidence acquisition. Adaptive gates remain advantageous on aggregate—DEFT attains the highest average on both backbones ($\mathbf{52.18}$ on LLaMA-3.1-8B and $\mathbf{35.05}$ on Qwen2.5-Math-7B),followed by Cayley-Trans (51.23 and 34.95).

Table 3: **Model-Weak (FigFont) results**: multi-seed mean ± standard deviation of Jaro–Winkler similarity.

Model	$-\log p$	$-p$	Hellinger	GCE	EAFT	InfoSFT	RankTuner	Cayley-Trans	DEFT
LLaMA-3.1-8B	36.05 ± 4.27	13.09 ± 2.50	32.52 ± 1.65	33.63 ± 2.48	$\underline{42.74 \pm 4.31}$	33.60 ± 2.75	33.16 ± 5.19	34.34 ± 1.91	$\mathbf{46.97 \pm 4.27}$
Qwen2.5-7B	52.14 ± 17.45	15.86 ± 9.05	39.53 ± 8.09	35.81 ± 3.77	$\underline{71.03 \pm 14.05}$	34.04 ± 0.78	63.12 ± 22.48	56.20 ± 15.32	$\mathbf{72.35 \pm 13.73}$

Table 4: **Stop-gradient ablation**: Detach on treats the focus index as a constant (stop-gradient), while detach off differentiates through it; the comparison isolates the effect of gradient flow through the focus term.

		Math		Medical		FigFont	
Method	Detach	Llama 3.1-8B	Qwen 2.5-MATH-7B	Llama 3.1-8B	Qwen 2.5-MATH-7B	Llama 3.1-8B	Qwen 2.5-7B
Cayley-Trans	on	10.50 ± 0.17	34.70 ± 0.20	51.23 ± 0.43	34.95 ± 0.53	34.34 ± 1.91	56.20 ± 15.32
Cayley-Trans	off	10.37 ± 0.03	34.94 ± 0.85	51.49 ± 0.41	34.95 ± 0.32	36.55 ± 4.57	66.27 ± 15.52
DEFT	on	11.04 ± 0.31	36.73 ± 0.21	52.18 ± 0.51	35.05 ± 0.48	46.97 ± 4.27	72.35 ± 13.73
DEFT	off	7.95 ± 0.73	36.37 ± 1.56	44.63 ± 1.04	34.41 ± 0.45	25.26 ± 1.56	24.97 ± 13.31

Model-Weak Results. Table 3 reports the Jaro–Winkler similarity on the FigFont task (additional-backbone scores in Table 10); since the prior is least aligned with the target, this regime is the most stringent test of *plasticity*. Gaps among objectives widen sharply relative to MS/MI: the linear $-p$ collapses to 13.09 on LLaMA-3.1-8B and 15.86 on Qwen2.5-7B, well below $-\log p$ (36.05 and 52.14), confirming that a statically sharpening gate suppresses precisely the hard-positive coverage that from-scratch knowledge injection requires, while the static deformed-log instances (Hellinger, GCE) likewise fail to improve over $-\log p$. DEFT and Cayley-Trans, by comparison, effectively overcome this limitation, demonstrating superior generalization and adaptability.

5.3 Ablation Studies

Stop-gradient sensitivity. In Table 4, Cayley-Trans responds to detachment differently across tasks, whereas DEFT's mean scores decrease in all six configurations when detachment is removed, most sharply on FigFont and LLaMA medical. This supports freezing concentration in DEFT under the evaluated settings.

Suppression intervention. To test whether the attenuation applied to concentrated, low-target-probability states is itself responsible for DEFT's gains, we fine-tune a controlled variant of the same objective: during training, tokens satisfying a fixed (H, p) threshold have their gate set to 1 (i.e., they receive the plain NLL gradient), while every other token keeps the DEFT gate; the data, schedule, and seed set are unchanged, so the two runs differ only in the treatment of the targeted region. Removing the attenuation there reduces the average score from 36.73 to 24.42 and 25.20, with decreases on all five benchmarks. This indicates that the gate on these predictive states is load-bearing rather than incidental; fully separating the attenuation from the resulting change in update magnitude would require an update-scale-matched control Appendix J.4. See Appendix R for 5% answer-corruption results, CBW frequencies, and interpretation limits.

Analysis of the ablation triptych. Beyond the seven objectives in the main comparison, we ablate two alternative probability-level gates (AdaptiveTargetP, FocalLoss (Lin et al., 2018)) and three alternative distribution-level proxies for DEFT (Shannon, Margin, EffSupportInv); definitions, hyper-parameters, and the fair-comparison protocol are deferred to Appendix J.1 and Appendix J.2. Within each adaptive branch of Figure 2, Cayley-Trans is the strongest probability-level instantiation on MS and MI and remains positive over $-\log p$ on MW (Panel a/c), and DEFT-Rényi-2 strictly dominates Shannon, Margin, and EffSupportInv on all three regimes (Panel b), with the largest gain again in MW. Gating on the full distribution rather than on p alone is thus the principal driver of further improvement beyond Cayley-Trans. We also conduct a model-scale ablation on the Qwen3 family (0.6B–14B),confirming that DEFT consistently leads in MS and MW while Cayley-Trans remains the strongest probability-level method across regimes; detailed results are reported in Appendix O.

Adaptive gating and cross-regime transfer. These experiments further demonstrate the effectiveness of state-dependent gating. Sweeping a static $\alpha \in \{0, 0.1, \ldots, 0.9, 1\}$ on LLaMA-3.1-8B shows that the best static α is regime-specific, so no fixed choice transfers across regimes. Our state-dependent gates require no tuning yet deliver the best overall performance: DEFT attains the highest value in MS (11.35) and MW and matches the best static oracle in MI, while Cayley-Trans surpasses

Table 5: **Intervention on concentrated low-target-probability states** using Qwen2.5-Math-7B. Each variant is fine-tuned from the same base model under the same schedule and seed set as DEFT, except that *training* the gate is fixed to 1 for tokens satisfying the indicated Shannon-entropy/target-probability thresholds.

Method	AIME24	AMC23	Math500	Minerva	Olympiad	Avg
DEFT	7.87 ± 0.77	41.88 ± 1.25	68.72 ± 0.77	32.95 ± 0.80	32.22 ± 0.21	$\mathbf{36.73 \pm 0.21}$
DEFT, gate$=1$ if $H \leq 0.3,\ p \leq 0.1$	2.17 ± 0.45	25.55 ± 1.22	53.98 ± 0.08	21.48 ± 0.66	18.92 ± 0.04	24.42 ± 0.49
DEFT, gate$=1$ if $H \leq 0.5,\ p \leq 0.3$	2.81 ± 0.15	26.72 ± 1.55	55.40 ± 0.04	21.21 ± 0.09	19.89 ± 0.11	$\underline{25.20 \pm 0.39}$

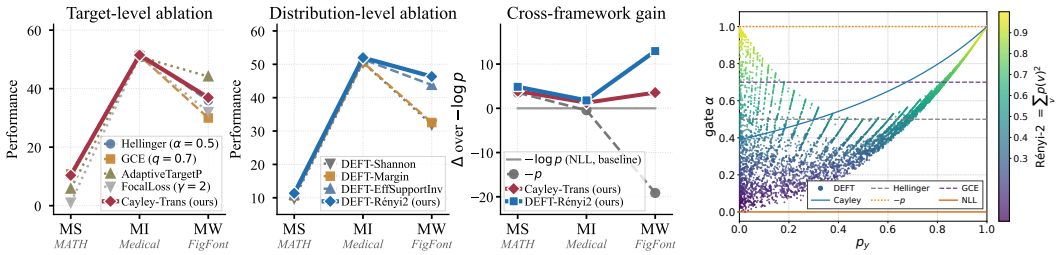

Figure 2: Ablation across MS, MI, MW (a–c) with the gate *space* view (far right). **(a)** Probability-level family; **(b)** four DEFT proxies (Renyi-2 default); **(c)** cross-framework Δ over $-\log p$; **Space**: mean α vs. p on MS, where DEFT gives different α at equal p. The *time* and *semantics* views are in Appendix P.

both untuned baselines ($-\log p$, $-p$) in all three regimes and the best static α in MW (Table 6; Appendix J.1). Cross-regime transfer (MS→MI, MI→MS, MW→MI) likewise shows that DEFT exceeds the reported static baselines in all three directions. Together, these results show that the gains stem from state-dependent adaptivity rather than any particular α, confirming the effectiveness of our objectives under the tested settings.

5.4 Mechanism Analysis

Adaptive Gate Analysis. We analyze Cayley-Trans and DEFT across *space*, *time*, and *semantics* in Figure 2 and Figure 10. **Space** (far-right panel of Figure 2): Cayley-Trans replaces a static scalar with a probability-dependent gate $\alpha(p)$, while DEFT further conditions on the full predictive distribution, exploiting structure unavailable to probability-only gates. **Time** (panel a of Figure 10): DEFT adapts both across regimes ($\alpha_0 \approx 0.65$ in MS versus 0.05 in MW) and over training, shifting automatically from coverage to sharpening. **Semantics** (panel b of Figure 10): At similar p, DEFT amplifies high-entropy reasoning tokens, dampens low-entropy format alternatives, and mutes mastered tokens—distinctions unavailable to probability-only gates. The token-level inventory of these regions is given in Appendix P, and post-training token distributions in Appendix K.

Gradient Landscape and Objective Analysis. We verify this picture with the token-level gradient distributions collected in Appendix L, covering the Model-Strong (Figure 6), Model-Intermediate (Figure 7), and Model-Weak (Figure 8) regimes. NLL emphasizes low-probability tail tokens, risking overfitting to rare or noisy patterns, whereas $-p$ emphasizes high-probability tokens, suppressing noise but underlearning hard positives. By also using model uncertainty, DEFT and Cayley-Trans avoid either extreme. In particular, DEFT suppresses low-probability, low-entropy updates, preventing them from dominating optimization while preserving learning capacity and stable representations.

6 Conclusion

We unify token-level SFT objectives within a deformed-log framework that exposes a universal *gate* $\times$ *error* gradient structure, and classify trust gates by the predictive information they consume: probability-level versus distribution-level. Under this taxonomy, we instantiate one parameter-free method on each axis: **Cayley-Trans**, derived from geometric boundary conditions via a Cayley conformal mapping; and **DEFT**, driven by the collision probability of the predictive distribution. Across seven backbones and three capability regimes, both consistently match or exceed strong baselines without task-specific tuning. A discussion of limitations is provided in Appendix Q.

REFERENCES

Yuzheng Cai, Siqi Cai, Yuchen Shi, Zihan Xu, Lichao Chen, Yulei Qin, Xiaoyu Tan, Gang Li, Zongyi Li, Haojia Lin, Yong Mao, Ke Li, and Xing Sun. Training-free group relative policy optimization, 2025. URL https://arxiv.org/abs/2510.08191.

Jack Chen, Fazhong Liu, Naruto Liu, Yuhan Luo, Erqu Qin, Harry Zheng, Tian Dong, Haojin Zhu, Yan Meng, and Xiao Wang. Step-wise adaptive integration of supervised fine-tuning and reinforcement learning for task-specific llms, 2025. URL https://arxiv.org/abs/2505.13026.

Muxi Diao, Lele Yang, Wuxuan Gong, Yutong Zhang, Zhonghao Yan, Yufei Han, Kongming Liang, Weiran Xu, and Zhanyu Ma. Entropy-adaptive fine-tuning: Resolving confident conflicts to mitigate forgetting, 2026. URL https://arxiv.org/abs/2601.02151.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, Danny Wyatt, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Francisco Guzmán, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Govind Thattai, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jack Zhang, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Karthik Prasad, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Kushal Lakhotia, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Maria Tsimpoukelli, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Ning Zhang, Olivier Duchenne, Onur Çelebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohan Maheswari, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gonguet, Virginie Do, Vish Vogeti, Vítor Albiero, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaofang Wang, Xiaoqing Ellen Tan, Xide Xia, Xinfeng Xie, Xuchao Jia, Xuewei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie Delpierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aayushi Srivastava, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Amos Teo, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Dong, Annie Franco, Anuj Goyal, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu,

Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Ce Liu, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Cynthia Gao, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Eric-Tuan Le, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Filippos Kokkinos, Firat Ozgenel, Francesco Caggioni, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hakan Inan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Hongyuan Zhan, Ibrahim Damlaj, Igor Molybog, Igor Tufanov, Ilias Leontiadis, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Janice Lam, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou U, Karan Saxena, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kiran Jagadeesh, Kun Huang, Kunal Chawla, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, Miao Liu, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikhil Mehta, Nikolay Pavlovich Laptev, Ning Dong, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Rangaprabhu Parthasarathy, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Russ Howes, Ruty Rinott, Sachin Mehta, Sachin Siby, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Mahajan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shishir Patil, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta, Summer Deng, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Koehler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaojian Wu, Xiaolan Wang, Xilun Wu, Xinbo Gao, Yaniv Kleinman, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yu Zhao, Yuchen Hao, Yundi Qian, Yunlu Li, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, Zhiwei Zhao, and Zhiyu Ma. The llama 3 herd of models, 2024. URL https://arxiv.org/abs/2407.21783.

Xiaoke Huang, Juncheng Wu, Hui Liu, Xianfeng Tang, and Yuyin Zhou. m1: Unleash the potential of test-time scaling for medical reasoning with large language models. *arXiv preprint arXiv:2504.00869*, 2025.

Gaotang Li, Ruizhong Qiu, Xiusi Chen, Heng Ji, and Hanghang Tong. Beyond log likelihood: Probability-based objectives for supervised fine-tuning across the model capability continuum. *arXiv preprint arXiv:2510.00526*, 2025.

Jia LI, Edward Beeching, Lewis Tunstall, Ben Lipkin, Roman Soletskyi, Shengyi Costa Huang, Kashif Rasul, Longhui Yu, Albert Jiang, Ziju Shen, Zihan Qin, Bin Dong, Li Zhou, Yann Fleureau, Guillaume Lample, and Stanislas Polu. Numinamath. https://huggingface.co/AI-MO/NuminaMath-CoT, 2024.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection, 2018. URL https://arxiv.org/abs/1708.02002.

Tao Liu, Taiqiang Wu, Runming Yang, Shaoning Sun, Junjie Wang, and Yujiu Yang. Profit: Leveraging high-value signals in sft via probability-guided token selection, 2026. URL https://arxiv.org/abs/2601.09195.

OpenAI, Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, Alex Iftimie, Alex Karpenko, Alex Tachard Passos, Alexander Neitz, Alexander Prokofiev, Alexander Wei, Allison Tam, Ally Bennett, Ananya Kumar, Andre Saraiva, Andrea Vallone, Andrew Duberstein, Andrew Kondrich, Andrey Mishchenko, Andy Applebaum, Angela Jiang, Ashvin Nair, Barret Zoph, Behrooz Ghorbani, Ben Rossen, Benjamin Sokolowsky, Boaz Barak, Bob McGrew, Borys Minaiev, Botao Hao, Bowen Baker, Brandon Houghton, Brandon McKinzie, Brydon Eastman, Camillo Lugaresi, Cary Bassin, Cary Hudson, Chak Ming Li, Charles de Bourcy, Chelsea Voss, Chen Shen, Chong Zhang, Chris Koch, Chris Orsinger, Christopher Hesse, Claudia Fischer, Clive Chan, Dan Roberts, Daniel Kappler, Daniel Levy, Daniel Selsam, David Dohan, David Farhi, David Mely, David Robinson, Dimitris Tsipras, Doug Li, Dragos Oprica, Eben Freeman, Eddie Zhang, Edmund Wong, Elizabeth Proehl, Enoch Cheung, Eric Mitchell, Eric Wallace, Erik Ritter, Evan Mays, Fan Wang, Felipe Petroski Such, Filippo Raso, Florencia Leoni, Foivos Tsimpourlas, Francis Song, Fred von Lohmann, Freddie Sulit, Geoff Salmon, Giambattista Parascandolo, Gildas Chabot, Grace Zhao, Greg Brockman, Guillaume Leclerc, Hadi Salman, Haiming Bao, Hao Sheng, Hart Andrin, Hessam Bagherinezhad, Hongyu Ren, Hunter Lightman, Hyung Won Chung, Ian Kivlichan, Ian O'Connell, Ian Osband, Ignasi Clavera Gilaberte, Ilge Akkaya, Ilya Kostrikov, Ilya Sutskever, Irina Kofman, Jakub Pachocki, James Lennon, Jason Wei, Jean Harb, Jerry Twore, Jiacheng Feng, Jiahui Yu, Jiayi Weng, Jie Tang, Jieqi Yu, Joaquin Quiñonero Candela, Joe Palermo, Joel Parish, Johannes Heidecke, John Hallman, John Rizzo, Jonathan Gordon, Jonathan Uesato, Jonathan Ward, Joost Huizinga, Julie Wang, Kai Chen, Kai Xiao, Karan Singhal, Karina Nguyen, Karl Cobbe, Katy Shi, Kayla Wood, Kendra Rimbach, Keren Gu-Lemberg, Kevin Liu, Kevin Lu, Kevin Stone, Kevin Yu, Lama Ahmad, Lauren Yang, Leo Liu, Leon Maksin, Leyton Ho, Liam Fedus, Lilian Weng, Linden Li, Lindsay McCallum, Lindsey Held, Lorenz Kuhn, Lukas Kondraciuk, Lukasz Kaiser, Luke Metz, Madelaine Boyd, Maja Trebacz, Manas Joglekar, Mark Chen, Marko Tintor, Mason Meyer, Matt Jones, Matt Kaufer, Max Schwarzer, Meghan Shah, Mehmet Yatbaz, Melody Y. Guan, Mengyuan Xu, Mengyuan Yan, Mia Glaese, Mianna Chen, Michael Lampe, Michael Malek, Michele Wang, Michelle Fradin, Mike McClay, Mikhail Pavlov, Miles Wang, Mingxuan Wang, Mira Murati, Mo Bavarian, Mostafa Rohaninejad, Nat McAleese, Neil Chowdhury, Neil Chowdhury, Nick Ryder, Nikolas Tezak, Noam Brown, Ofir Nachum, Oleg Boiko, Oleg Murk, Olivia Watkins, Patrick Chao, Paul Ashbourne, Pavel Izmailov, Peter Zhokhov, Rachel Dias, Rahul Arora, Randall Lin, Rapha Gontijo Lopes, Raz Gaon, Reah Miyara, Reimar Leike, Renny Hwang, Rhythm Garg, Robin Brown, Roshan James, Rui Shu, Ryan Cheu, Ryan Greene, Saachi Jain, Sam Altman, Sam Toizer, Sam Toyer, Samuel Miserendino, Sandhini Agarwal, Santiago Hernandez, Sasha Baker, Scott McKinney, Scottie Yan, Shengjia Zhao, Shengli Hu, Shibani Santurkar, Shraman Ray Chaudhuri, Shuyuan Zhang, Siyuan Fu, Spencer Papay, Steph Lin, Suchir Balaji, Suvansh Sanjeev, Szymon Sidor, Tal Broda, Aidan Clark, Tao Wang, Taylor Gordon, Ted Sanders, Tejal Patwardhan, Thibault Sottiaux, Thomas Degry, Thomas Dimson, Tianhao Zheng, Timur Garipov, Tom Stasi, Trapit Bansal, Trevor Creech, Troy Peterson, Tyna Eloundou, Valerie Qi, Vineet Kosaraju, Vinnie Monaco, Vitchyr Pong, Vlad Fomenko, Weiyi Zheng, Wenda Zhou, Wes McCabe, Wojciech Zaremba, Yann Dubois, Yinghai Lu, Yining Chen, Young Cha, Yu Bai, Yuchen He, Yuchen Zhang, Yunyun Wang, Zheng Shao, and Zhuohan Li. Openai o1 system card, 2024. URL https://arxiv.org/abs/2412.16720.

Jinlong Pang, Na Di, Zhaowei Zhu, Jiaheng Wei, Hao Cheng, Chen Qian, and Yang Liu. Token cleaning: Fine-grained data selection for llm supervised fine-tuning, 2025. URL https://arxiv.org/abs/2502.01968.

Qwen, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan,

Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 technical report, 2025. URL https://arxiv.org/abs/2412.15115.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model, 2024. URL https://arxiv.org/abs/2305.18290.

Zhiwen Ruan, Yixia Li, He Zhu, Yun Chen, Peng Li, Yang Liu, and Guanhua Chen. Enhancing large language model reasoning via selective critical token fine-tuning, 2025. URL https://arxiv.org/abs/2510.10974.

Mahdi Sabbaghi, George Pappas, Adel Javanmard, and Hamed Hassani. InfoSFT: Learn more and forget less with information-aware token weighting, 2026. URL https://arxiv.org/abs/2605.14967v1.

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, 2024. URL https://arxiv.org/abs/2402.03300.

Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. Hybridflow: A flexible and efficient rlhf framework. *arXiv preprint arXiv: 2409.19256*, 2024.

Zafir Stojanovski, Oliver Stanley, Joe Sharratt, Richard Jones, Abdulhakeem Adefioye, Jean Kaddour, and Andreas Köpf. Reasoning gym: Reasoning environments for reinforcement learning with verifiable rewards, 2025. URL https://arxiv.org/abs/2505.24760.

Ali Taheri, Alireza Taban, Shanshan Ye, Abdolreza Mirzaei, Tongliang Liu, and Bo Han. Forgetting: A new mechanism towards better large language model fine-tuning, 2025. URL https://arxiv.org/abs/2508.04329.

Constantino Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52(1–2):479–487, 1988.

Bingning Wang, Haizhou Zhao, Huozhi Zhou, Liang Song, Mingyu Xu, Wei Cheng, Xiangrong Zeng, Yupeng Zhang, Yuqi Huo, Zecheng Wang, Zhengyun Zhao, Da Pan, Fei Kou, Fei Li, Fuzhong Chen, Guosheng Dong, Han Liu, Hongda Zhang, Jin He, Jinjie Yang, Kangxi Wu, Kegeng Wu, Lei Su, Linlin Niu, Linzhuang Sun, Mang Wang, Pengcheng Fan, Qianli Shen, Rihui Xin, Shunya Dang, Songchi Zhou, Weipeng Chen, Wenjing Luo, Xin Chen, Xin Men, Xionghai Lin, Xuezhen Dong, Yan Zhang, Yifei Duan, Yuyan Zhou, Zhi Ma, and Zhiying Wu. Baichuan-m1: Pushing the medical capability of large language models, 2025a. URL https://arxiv.org/abs/2502.12671.

Zecheng Wang, Chunshan Li, Yupeng Zhang, Han Liu, Bingning Wang, Dianhui Chu, and Dianbo Sui. VPO: Reasoning preferences optimization based on $\mathcal{V}$-usable information. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025b. URL https://openreview.net/forum?id=LCZmI3iM8X.

Chao-Chung Wu, Zhi Rui Tam, Chieh-Yen Lin, Yun-Nung Chen, Shao-Hua Sun, and Hung yi Lee. Mitigating forgetting in llm fine-tuning via low-perplexity token learning, 2025a. URL https://arxiv.org/abs/2501.14315.

Yongliang Wu, Yizhou Zhou, Zhou Ziheng, Yingzhe Peng, Xinyu Ye, Xinting Hu, Wenbo Zhu, Lu Qi, Ming-Hsuan Yang, and Xu Yang. On the generalization of sft: A reinforcement learning perspective with reward rectification, 2025b. URL https://arxiv.org/abs/2508.05629.

An Yang, Beichen Zhang, Binyuan Hui, Bofei Gao, Bowen Yu, Chengpeng Li, Dayiheng Liu, Jianhong Tu, Jingren Zhou, Junyang Lin, Keming Lu, Mingfeng Xue, Runji Lin, Tianyu Liu, Xingzhang Ren, and Zhenru Zhang. Qwen2.5-math technical report: Toward mathematical expert model via self-improvement, 2024. URL https://arxiv.org/abs/2409.12122.

Wenhao Yu, Shaohang Wei, Jiahong Liu, Yifan Li, Minda Hu, Aiwei Liu, Hao Zhang, and Irwin King. Probability-entropy calibration: An elastic indicator for adaptive fine-tuning, 2026. URL https://arxiv.org/abs/2602.01745v1.

Shengyu Zhang, Linfeng Dong, Xiaoya Li, Sen Zhang, Xiaofei Sun, Shuhe Wang, Jiwei Li, Runyi Hu, Tianwei Zhang, Fei Wu, and Guoyin Wang. Instruction tuning for large language models: A survey, 2025. URL https://arxiv.org/abs/2308.10792.

Zhilu Zhang and Mert R. Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels, 2018. URL https://arxiv.org/abs/1805.07836.

A Related Works

In recent years, the rapid progress of LLMs has substantially advanced general-purpose AI capabilities Grattafiori et al. (2024); Qwen et al. (2025); Shao et al. (2024); Wang et al. (2025a); OpenAI et al. (2024). In the dominant post-training paradigm, aligning pretrained models with human preferences and domain needs—while enhancing reasoning—remains a central challenge Cai et al. (2025); Zhang et al. (2025); Rafailov et al. (2024); Wang et al. (2025b). SFT, as a core stage, serves as a key bridge from pretraining to downstream performance. However, standard NLL uses a unit token gate: its gradients reflect prediction error, but the objective does not explicitly weight tokens by data quality or distributional uncertainty. To improve the training signal-to-noise ratio, prior work mainly follows two lines: token-level valuation/filtering and loss reweighting based on uncertainty or predicted probability.

Token-level modeling and filtering. Recent studies advance data quality control from coarse, sample-level filtering to finer, token-level interventions by estimating the training value of each token. Specifically, Pang et al. (2025) evaluate token contributions to parameter updates or performance variations and perform separation/cleaning via thresholding. Taheri et al. (2025) classify tokens as beneficial vs. harmful and impose explicit forgetting/suppression on the latter. Moreover, Ruan et al. (2025) use counterfactual perturbations to identify critical reasoning tokens and selectively fine-tune them, improving stability and generalization on complex reasoning tasks. Liu et al. (2026) further show that high-probability tokens often encode core logical structure, whereas low-probability tokens are more interchangeable; selectively masking such tokens can improve reasoning performance. Despite their effectiveness, these approaches typically require costly offline computation or heuristic preprocessing, and their static/hard masking mechanisms are limited in adapting to the model's evolving knowledge boundaries during training.

Modifying SFT losses via probability/uncertainty-aware reweighting. Beyond explicit token filtering, a complementary line revises the SFT objective using model-confidence signals. The established *gate* $\times$ *error* representation for deformed losses and stop-gradient-weighted NLL, reviewed in Section 3, helps distinguish *probability-level* rules based only on target probability p from *distribution-level* rules using additional statistics of $p_\theta(\cdot \mid c)$. Probability-based objectives and detached probability weighting (Li et al., 2025; Wu et al., 2025b) can induce the gate $g = p$, reducing low-probability updates relative to NLL. Generalized cross-entropy (GCE) (Zhang & Sabuncu, 2018) supplies the fixed-power family $g = p^\alpha$. **InfoSFT** (Sabbaghi et al., 2026) emphasizes medium-confidence tokens with a detached, clipped logit-difference weight multiplied by the current target probability. Such probability-only rules assign the same multiplier to predictive distributions sharing the same target probability. Distribution-aware methods use additional features of the predictive state. **EAFT** (Diao et al., 2026) weights cross-entropy by predictive entropy and evaluates general-capability retention before and after adaptation, including MMLU, IFEval, and CLUEWSC. **RankTuner** (Yu et al., 2026) combines target rank and distributional uncertainty in a relative scale, with a probability–entropy approximation for training. Its final weighting also incorporates a task-dependent base: p for mathematical reasoning and 1 for general tasks. **DEFT** uses concentration as the exponent of the target probability: $g = p^C$, where $C = \sum_i P_i^2$. This rule couples target probability with the organization of non-target mass, as characterized by Eq. (13). An explicit frozen surrogate specifies its gradient, and **Cayley-Trans** provides a probability-only counterpart. We evaluate these rules across capability regimes using the same focus formulas across tasks, while retaining standard training hyperparameters. The detached weighted-CE representation in Section 4.3 provides a common basis for comparing their update rules.

B Ethics Statement

This work adheres to the ICLR Code of Ethics. No human subjects or animal experimentation was involved. All datasets used, including the NuminaMath corpus (LI et al., 2024), the M1-23K medical reasoning dataset (Huang et al., 2025), and the standard mathematical and medical benchmark suites (Math500, Minerva Math, OlympiadBench, AIME24, AMC23, MedMCQA, MedQA, PubMedQA, MMLU-Pro, GPQA, MedX, NEJM, and MedB), were sourced in compliance with their respective usage guidelines, ensuring no violation of privacy. The FigFont character-art puzzles are synthesized by us following Stojanovski et al. (2025) and contain no personal data. We have taken care to avoid any biases or discriminatory outcomes in our research process. No personally identifiable information

was used, and no experiments were conducted that could raise privacy or security concerns. We are committed to maintaining transparency and integrity throughout the research process.

C Reproducibility Statement

We have made every effort to ensure that the results presented in this paper are reproducible. The generalized deformed-log objective, the Cayley-Trans and DEFT focus indices, and their gradient structure are defined and derived analytically in the main text and in Appendix F and Appendix H, so that the optimization rules can be re-implemented from the equations alone. The experimental setup, including datasets, backbones, training steps, optimization details, and hardware, is described in Appendix J.3 and Appendix J.4. All code and evaluation scripts will be made publicly available in an anonymous repository to facilitate replication and verification. We also provide a full description of our contribution—the unifying objective family and the two state-dependent focus indices—to assist others in reproducing our experiments. The public datasets listed above, such as NuminaMath (LI et al., 2024), M1-23K (Huang et al., 2025), and the standard math and medical benchmarks, are publicly available, ensuring consistent and reproducible evaluation. As noted in Appendix J.4, some run-level details (the complete run manifest, proxy normalization constants, and per-seed records) remain to be finalized; these gaps currently limit statistical comparison and full replication and are being addressed. We believe these measures will enable other researchers to reproduce our work and further advance the field.

D LLM Usage

Large Language Models (LLMs) were used to aid in the writing and polishing of the manuscript. Specifically, we used an LLM to assist in refining the language, improving readability, and ensuring clarity in various sections of the paper. The model helped with tasks such as sentence rephrasing, grammar checking, and enhancing the overall flow of the text. It is important to note that the LLM was not involved in the ideation, research methodology, or experimental design. All research concepts, ideas, and analyses were developed and conducted by the authors. The contributions of the LLM were solely focused on improving the linguistic quality of the paper, with no involvement in the scientific content or data analysis. The authors take full responsibility for the content of the manuscript, including any text generated or polished by the LLM. We have ensured that the LLM-generated text adheres to ethical guidelines and does not contribute to plagiarism or scientific misconduct.

E Setup

For a prompt context $c \in \mathcal{C}$, let $\tilde{y} \in \mathcal{V}$ denote the supervised target token, where $\mathcal{V}$ is the vocabulary. The token-level training distribution is denoted by T, which generates pairs $(c, \tilde{y})$. We write $r(\cdot \mid c) \in \Delta^{|\mathcal{V}|-1}$ for the true conditional next-token distribution. Let $z_{\boldsymbol{\theta}}(c) \in \mathbb{R}^{|\mathcal{V}|}$ denote the pre-softmax logits of an autoregressive language model with parameters $\boldsymbol{\theta}$. The predicted next-token distribution is

$$p_{\boldsymbol{\theta}}(\cdot \mid c) = \mathrm{softmax}\big(z_{\boldsymbol{\theta}}(c)\big) \in \Delta^{|\mathcal{V}|-1}. \tag{15}$$

For brevity, we write $p_{\boldsymbol{\theta}}(y \mid c)$ as the probability assigned to token y under context c. We denote the base model by $p_{\boldsymbol{\theta}_0}$. For a differentiable function $f : [0, 1] \to \mathbb{R}$, the general SFT objective is

$$\mathcal{L}_f(\boldsymbol{\theta}) \;=\; \mathbb{E}_{(c,\tilde{y})\sim T}[f(p_{\boldsymbol{\theta}}(\tilde{y} \mid c))]\,. \tag{16}$$

The standard negative log-likelihood (NLL) objective corresponds to $f(p) = -\log p$. For a fixed context c and supervised target $\tilde{y}$, we abbreviate

$$p \;\triangleq\; p_{\boldsymbol{\theta}}(\tilde{y} \mid c) \in (0, 1]. \tag{17}$$

Throughout this appendix, we use $e_y \in \{0, 1\}^{|\mathcal{V}|}$ to denote the one-hot vector for token y, and $\mathbf{1} \in \mathbb{R}^{|\mathcal{V}|}$ to denote the all-ones vector.

F GRADIENT STRUCTURE

F.1 DETAILED DERIVATION OF THE SOFTMAX-LOGIT GRADIENT

Lemma 1 (Softmax Jacobian). *Let $z \in \mathbb{R}^{|\mathcal{V}|}$ be a vector of logits and $p = \mathrm{softmax}(z) \in \Delta^{|\mathcal{V}|-1}$ be the corresponding probabilities. Then for any $i, j \in \mathcal{V}$,*

$$\frac{\partial p_i}{\partial z_j} = p_i \left(\delta_{i,j} - p_j\right), \tag{18}$$

where $\delta_{i,j}$ is the Kronecker delta ($\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ otherwise).

Proof. **Step 0: Write the softmax formula explicitly.** By definition,

$$p_i = \frac{\exp(z_i)}{\sum_{k=1}^{|\mathcal{V}|} \exp(z_k)}. \tag{19}$$

Define the partition function $Z := \sum_{k=1}^{|\mathcal{V}|} \exp(z_k)$, so that $p_i = \exp(z_i)/Z$.

Step 1: Apply the quotient rule. To compute $\partial p_i / \partial z_j$, we use the quotient rule:

$$\frac{\partial p_i}{\partial z_j} = \frac{\partial}{\partial z_j}\left(\frac{\exp(z_i)}{Z}\right) = \frac{\left(\partial_j \exp(z_i)\right) Z - \exp(z_i)\left(\partial_j Z\right)}{Z^2}. \tag{20}$$

Step 2: Evaluate the numerator and denominator separately.

- $\partial_j \exp(z_i) = \delta_{i,j} \exp(z_i)$.
- $\partial_j Z = \partial_j \sum_k \exp(z_k) = \exp(z_j)$.

Substituting these into the quotient rule:

$$\frac{\partial p_i}{\partial z_j} = \frac{\delta_{i,j} \exp(z_i)\, Z - \exp(z_i)\, \exp(z_j)}{Z^2} = \frac{\delta_{i,j} \exp(z_i)}{Z} - \frac{\exp(z_i)\, \exp(z_j)}{Z^2}. \tag{21}$$

Step 3: Simplify using the softmax probabilities. Recall that $p_i = \exp(z_i)/Z$ and $p_j = \exp(z_j)/Z$. Thus,

$$\frac{\partial p_i}{\partial z_j} = \delta_{i,j}\, p_i - p_i p_j = p_i \left(\delta_{i,j} - p_j\right), \tag{22}$$

as claimed. □

Lemma 2 (General Objective Gradient). *Let $f : (0,1) \to \mathbb{R}$ be differentiable and nonincreasing. For the objective in Eq. 2, consider one supervised token's contribution $\ell_t = f(p_t)$, where $p_t \triangleq p_{\boldsymbol{\theta}}(\tilde{y} \mid c)$ is its target probability. The following derivatives are per-token derivatives, before any batch or population averaging. Then the gradient of ℓ_t with respect to the logits $z_{t,i}$ satisfies:*

$$\frac{\partial \ell_t}{\partial z_{t,i}} = s_f(p_t)\, \left(p_{t,i} - \delta_{i,\tilde{y}}\right), \qquad \textit{where} \quad s_f(p) \triangleq -f'(p)\, p \geq 0. \tag{23}$$

In particular, for the target token $i = \tilde{y}$,

$$\frac{\partial \ell_t}{\partial z_{t,\tilde{y}}} = -s_f(p_t)\,(1 - p_t) =: -W_f(p_t), \qquad W_f(p) \triangleq s_f(p)\,(1-p). \tag{24}$$

Proof. **Step 0: Express the loss contribution at step t.** The loss contribution for the pair $(c, \tilde{y})$ is $f(p_t)$, where $p_t = p_{\boldsymbol{\theta}}(\tilde{y} \mid c)$.

Step 1: Apply the chain rule through softmax. By the chain rule,

$$\frac{\partial \ell_t}{\partial z_{t,i}} = \frac{\partial f(p_t)}{\partial p_t} \cdot \frac{\partial p_t}{\partial z_{t,i}}. \tag{25}$$

Step 2: Compute each factor separately.

- The derivative of f with respect to p is simply $f'(p_t)$.
- From Lem. 1, setting the output neuron index to $\tilde{y}$,

$$\frac{\partial p_t}{\partial z_{t,i}} = \frac{\partial p_{\tilde{y}}}{\partial z_{t,i}} = p_{\tilde{y}} \left(\delta_{i,\tilde{y}} - p_i\right) = p_t \left(\delta_{i,\tilde{y}} - p_{t,i}\right). \tag{26}$$

Step 3: Combine and rearrange. Multiplying the two factors:

$$\frac{\partial \ell_t}{\partial z_{t,i}} = f'(p_t)\, p_t \left(\delta_{i,\tilde{y}} - p_{t,i}\right) = \left(-f'(p_t)\, p_t\right) \left(p_{t,i} - \delta_{i,\tilde{y}}\right). \tag{27}$$

Step 4: Define the gate term. Let $s_f(p) \triangleq -f'(p)\, p$. Since f is nonincreasing by assumption, $f'(p) \leq 0$ for all $p \in (0,1)$, hence $s_f(p) \geq 0$. This yields

$$\frac{\partial \ell_t}{\partial z_{t,i}} = s_f(p_t) \left(p_{t,i} - \delta_{i,\tilde{y}}\right). \tag{28}$$

Step 5: Specialize to the target logit $i = \tilde{y}$. Setting $i = \tilde{y}$ and using $\delta_{\tilde{y},\tilde{y}} = 1$:

$$\frac{\partial \ell_t}{\partial z_{t,\tilde{y}}} = s_f(p_t) \left(p_t - 1\right) = -s_f(p_t) \left(1 - p_t\right). \tag{29}$$

Define the learning signal magnitude $W_f(p) \triangleq s_f(p)(1-p) = -f'(p)\, p(1-p)$, which yields

$$\frac{\partial \ell_t}{\partial z_{t,\tilde{y}}} = -W_f(p_t), \tag{30}$$

as claimed. □

G Deformed-Log Losses and Bayes Risk

G.1 Properties of the q-Logarithm

Proposition 1 (Continuity and limit properties of $\ln_q$). *For the q-logarithm defined by*

$$\ln_q(x) \triangleq \frac{x^{1-q} - 1}{1-q}, \tag{31}$$

we have the following properties:

1. *As $q \to 1$, $\ln_q(x) \to \ln(x)$ for any $x > 0$.*
2. *For fixed $q \neq 1$, $\ln_q$ is strictly monotone increasing in $x > 0$.*
3. *The derivative is $\frac{\mathrm{d}}{\mathrm{d}x} \ln_q(x) = x^{-q}$.*

Proof. **(1) Limit as $q \to 1$ using L'Hôpital's rule.**

Fix $x > 0$. Write $\ln_q(x) = \frac{x^{1-q}-1}{1-q}$. As $q \to 1$, both numerator and denominator approach zero:

$$\lim_{q \to 1} (x^{1-q} - 1) = x^0 - 1 = 0, \qquad \lim_{q \to 1} (1 - q) = 0. \tag{32}$$

This is an indeterminate form $0/0$, so we apply L'Hôpital's rule.

Step 1: Differentiate numerator and denominator with respect to q.

- Numerator: $\frac{\mathrm{d}}{\mathrm{d}q}(x^{1-q}) = \frac{\mathrm{d}}{\mathrm{d}q} \exp\big((1-q)\ln x\big) = -\ln x \cdot x^{1-q}$.
- Denominator: $\frac{\mathrm{d}}{\mathrm{d}q}(1-q) = -1$.

Step 2: Apply L'Hôpital's rule.

$$\lim_{q\to 1} \ln_q(x) = \lim_{q\to 1} \frac{-\ln x \cdot x^{1-q}}{-1} = \ln x \cdot \lim_{q\to 1} x^{1-q} = \ln x \cdot x^0 = \ln x. \tag{33}$$

(2) Monotonicity.

Take the derivative with respect to x:

$$\frac{\mathrm{d}}{\mathrm{d}x} \ln_q(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x^{1-q}-1}{1-q}\right) = \frac{1}{1-q}(1-q)x^{(1-q)-1} = x^{-q}. \tag{34}$$

Since $x > 0$ and $x^{-q} > 0$ for any q, $\ln_q(x)$ is strictly increasing in x.

(3) Derivative formula.

This was computed in step (2):

$$\frac{\mathrm{d}}{\mathrm{d}x} \ln_q(x) = x^{-q}. \tag{35}$$

□

G.2 Bayes Optima and the Entropy Connection

Theorem 2 (Bayes optimum for a constant focus parameter; see Thm. 1) . *Fix a context c, a finite vocabulary, and a true distribution $r(\cdot \mid c)$; suppress c below. Optimize over the closed probability simplex, using $S_\alpha(\hat{p}, y) = (1-\hat{p}_y^\alpha)/\alpha$ for fixed $\alpha > 0$ and $S_0(\hat{p}, y) = -\ln \hat{p}_y$ (with the usual extended-value convention). For $\alpha > 0$, the conditional risk is*

$$\mathcal{R}_\alpha(\hat{p}; r) = \mathbb{E}_{y\sim r}[S_\alpha(\hat{p}, y)] = \frac{1 - \sum_y r_y \hat{p}_y^\alpha}{\alpha}. \tag{36}$$

For fixed $0 < \alpha < 1$, let $\beta = 1/(1-\alpha) > 1$ and $Z_\beta = \sum_y r_y^\beta$. The unique Bayes optimum and its risk are

$$\hat{p}_y^\star = \frac{r_y^\beta}{Z_\beta}, \qquad B_\alpha(r) = \min_{\hat{p}} \mathcal{R}_\alpha(\hat{p}; r) = \frac{1 - Z_\beta^{1/\beta}}{\alpha}. \tag{37}$$

This escort distribution raises the target probabilities to the power β and renormalizes them. With Rényi entropy $H_\beta(r) = \ln Z_\beta/(1-\beta)$,

$$B_\alpha(r) = \frac{1 - \exp(-\alpha H_\beta(r))}{\alpha}, \tag{38}$$

a strictly increasing transform of $H_\beta(r)$ for fixed α. At $\alpha = 0$, the score is strictly proper: $\hat{p}^\star = r$ and $B_0(r) = H(r)$. At $\alpha = 1$, every distribution supported on $\operatorname{argmax}_y r_y$ is optimal, including mixtures when modes tie. For $\alpha > 1$, only one-hot distributions at a mode are optimal; in both cases $B_\alpha(r) = (1 - \max_y r_y)/\alpha$.

Proof. **Step 1: Reduce the interior-index problem to the support of r.** For $0 < \alpha < 1$, minimizing

$$\mathcal{R}_\alpha(\hat{p}; r) = \frac{1 - F_\alpha(\hat{p})}{\alpha}, \qquad F_\alpha(\hat{p}) = \sum_y r_y \hat{p}_y^\alpha \tag{39}$$

is equivalent to maximizing F_α. No optimum places mass outside the support of r, since moving that mass onto its support strictly increases F_α. On this support the objective is strictly concave, and the infinite right derivative at zero gives an interior optimum (the one-point support case is immediate).

Step 2: Solve the stationarity equation. On the support of r, use

$$\mathcal{J}(\hat{p}, \lambda) = \sum_y r_y \hat{p}_y^\alpha - \lambda\left(\sum_y \hat{p}_y - 1\right). \tag{40}$$

Its first-order condition gives

$$\alpha r_y \hat{p}_y^{\alpha-1} = \lambda \quad \Longrightarrow \quad \hat{p}_y = (\alpha r_y/\lambda)^{1/(1-\alpha)} = \frac{r_y^\beta}{Z_\beta}. \tag{41}$$

Strict concavity proves uniqueness. Since $1 + \alpha\beta = \beta$ and $1 - \alpha = 1/\beta$,

$$F_\alpha(\hat{p}^\star) = \frac{\sum_y r_y^{1+\alpha\beta}}{Z_\beta^\alpha} = Z_\beta^{1/\beta}, \tag{42}$$

which proves Eqs. 37 and 38.

Step 3: Handle the endpoints and larger indices. At $\alpha = 0$, $\mathcal{R}_0(\hat{p}; r) = H(r) + \mathrm{KL}(r\|\hat{p})$, with equality only at $\hat{p} = r$. At $\alpha = 1$, maximizing $\sum_y r_y \hat{p}_y$ puts all mass on the mode set, without requiring a unique mode or a one-hot prediction. For $\alpha > 1$, writing $M = \max_y r_y > 0$ gives

$$\sum_y r_y \hat{p}_y^\alpha \le M \sum_y \hat{p}_y^\alpha \le M.$$

Equality in the second inequality requires a one-hot prediction, and equality in the first requires its support to be a mode. This proves the stated risks and optimal sets. □

Interpretation. For a constant $0 < \alpha < 1$, the escort exponent $\beta = 1/(1-\alpha)$ determines the concentration of the optimal prediction. On the support of r, $\hat{p}_i^\star/\hat{p}_j^\star = (r_i/r_j)^\beta$, so larger α amplifies probability ratios more strongly. The Bayes risk increases monotonically with Rényi-β entropy. Together, these properties describe the confidence preference induced by a constant focus parameter. State-dependent updates are defined by the surrogate in Section 4.3. Boundary optima are understood as limits when predictions are restricted to finite-logit softmax distributions.

H Unifying Coverage and Sharpening

H.1 Geometric Derivation of the State-Dependent Focus Trajectory

In this section, we provide a detailed geometric derivation of the state-dependent focus index $\alpha^*(p)$ that automatically transitions from coverage (NLL) to sharpening (linear loss) as the model's confidence increases.

Definition 2 (Fisher-Rao radius) . *For a probability $p \in (0, 1]$, define the intrinsic uncertainty radius as*

$$z \triangleq \sqrt{1-p} \in [0, 1]. \tag{43}$$

This quantity is a convenient monotone re-parameterization *of the Fisher-Rao distance-to-certainty for Bernoulli distributions (made explicit below).*

Proposition 2 (Fisher-Rao distance to certainty for Bernoulli) . *Consider the Bernoulli family* $\mathrm{Bern}(p)$ *with parameter $p \in (0, 1)$. The Fisher-Rao metric is*

$$\mathrm{d}s^2 = \frac{\mathrm{d}p^2}{p(1-p)}. \tag{44}$$

The Fisher-Rao geodesic distance from $\mathrm{Bern}(p)$ *to certainty* $\mathrm{Bern}(1)$ *is*

$$d_{\mathrm{FR}}(p, 1) \;=\; \int_p^1 \frac{\mathrm{d}u}{\sqrt{u(1-u)}} \;=\; 2\arccos\!\left(\sqrt{p}\right). \tag{45}$$

Moreover, the radius $z = \sqrt{1-p}$ equals the sine of the half-distance:

$$z = \sqrt{1-p} = \sin\left(\frac{d_{\mathrm{FR}}(p, 1)}{2}\right). \tag{46}$$

Proof. **Step 0: Fisher information and Fisher-Rao metric.** For Bernoulli, the Fisher information is $I(p) = \frac{1}{p(1-p)}$, which induces the 1D Fisher-Rao line element $\mathrm{d}s^2 = I(p)\,\mathrm{d}p^2$ and yields Eq. 44.

Step 1: Integrate the line element. The geodesic distance in 1D is the integral of the line element:

$$d_{\mathrm{FR}}(p, 1) = \int_p^1 \frac{\mathrm{d}u}{\sqrt{u(1-u)}}.$$

Substitute $u = \cos^2\theta$ with $\theta \in [0, \frac{\pi}{2}]$. Then $\mathrm{d}u = -2\cos\theta\sin\theta\,\mathrm{d}\theta$ and $\sqrt{u(1-u)} = \cos\theta\sin\theta$, hence

$$\int_p^1 \frac{\mathrm{d}u}{\sqrt{u(1-u)}} = \int_{\arccos(\sqrt{p})}^0 \frac{-2\cos\theta\sin\theta\,\mathrm{d}\theta}{\cos\theta\sin\theta} = 2\arccos(\sqrt{p}),$$

which gives Eq. 45.

Step 2: Relate z to d_{FR}. From Eq. 45, we have $d_{\mathrm{FR}}(p,1)/2 = \arccos(\sqrt{p})$. Thus,

$$\sin\left(\frac{d_{\mathrm{FR}}(p,1)}{2}\right) = \sin\big(\arccos(\sqrt{p})\big) = \sqrt{1-p} = z,$$

which is Eq. 46. □

From confidence p to intrinsic radius z. We will construct the focus index as a function of the Fisher-Rao radius z in Eq. 43, and then pull it back to p via $z = \sqrt{1-p}$. This is conceptually convenient because z is an *intrinsic* coordinate: by Eqs. 45–(46), it equals $\sin\big(d_{\mathrm{FR}}(p,1)/2\big)$, i.e., a monotone re-parameterization of the Fisher-Rao distance to certainty. Thus, specifying $\alpha(z)$ amounts to specifying how the objective should transition as the model approaches certainty along the information-geometric geodesic.

Proposition 3 (Boundary conditions for coverage and sharpening) . *To specify a coverage/sharpening interpolation in the probability coordinate, we impose a continuous mapping $\alpha : [0,1] \to [0,1]$ with:*

1. ***Coverage anchor:*** *$\alpha(0) = 0$ and $\alpha(p)|\ln p| \to 0$ as $p \to 0$, so that the gate approaches the NLL gate.*
2. ***Sharpening anchor:*** *$\alpha(1) = 1$, selecting the linear-probability endpoint as $p \to 1$.*

Justification. **Setup: an interpolating gate family.** Consider the detached-index *gate* ansatz

$$s(p) = p^\alpha, \qquad \alpha \in [0,1], \tag{47}$$

which connects the NLL gate ($s(p) = 1$ at $\alpha = 0$) and the linear-probability gate ($s(p) = p$ at $\alpha = 1$). We then choose $\alpha = \alpha(p)$ from the current confidence.

(1) Coverage anchor. For $f(p) = -\ln p$, the gate $s_f(p) = -pf'(p)$ in Eq. 23 is 1. Since $p^{\alpha(p)} = \exp(\alpha(p)\ln p)$, the condition $\alpha(p)|\ln p| \to 0$ recovers this gate as $p \to 0$, as formalized in Lemma 3.

(2) Sharpening anchor. The linear-probability endpoint $f(p) = 1 - p$ has $s_f(p) = p$, giving greater relative emphasis to higher-probability targets. Throughout the interpolation, $p(1-p) \le p^\alpha(1-p) \le 1-p$ for $\alpha \in [0,1]$, and the target-logit magnitude tends to zero as $p \to 1$. □

Lemma 3 (When does $p^{\alpha(p)} \to 1$ as $p \to 0$?) . *Let $\alpha : (0,1] \to [0,1]$ satisfy $\lim_{p\to 0}\alpha(p) = 0$ and $\lim_{p\to 0}\alpha(p)\,|\ln p| = 0$. Then*

$$\lim_{p\to 0} p^{\alpha(p)} = 1. \tag{48}$$

Proof. Take logs:

$$\ln\big(p^{\alpha(p)}\big) = \alpha(p)\ln p.$$

Since $\ln p \to -\infty$ as $p \to 0$, the assumption $\alpha(p)\,|\ln p| \to 0$ implies $\alpha(p)\ln p \to 0$, hence $p^{\alpha(p)} \to e^0 = 1$. □

Theorem 3 (The Cayley Transform Solution) . *Among Möbius (linear-fractional) maps $\alpha : [0,1] \to [0,1]$ that swap the endpoints*

$$z = 1 \;\Rightarrow\; \alpha = 0, \qquad z = 0 \;\Rightarrow\; \alpha = 1, \tag{49}$$

the general form is the one-parameter family $\alpha_\kappa(z) = \frac{1-z}{1+\kappa z}$ *with* $\kappa > -1$*, and every such map is* self-dual *in the sense of being an involution,*

$$\alpha(\alpha(z)) = z, \tag{50}$$

in particular including the affine reflection ($\kappa = 0$) *and a continuum of non-affine choices* ($\kappa \neq 0$)*. If we further impose the* surprisal linearization *requirement that* $\operatorname{arctanh}(\alpha_\kappa(z))$ *is affine in* $\ln z$ *(equivalently, compatibility with the normalized-surprisal form in Prop. 5 without any additive offset), this uniquely selects* $\kappa = 1$*, i.e. the Cayley transform:*

$$\alpha(z) \;=\; \frac{1-z}{1+z}. \tag{51}$$

Substituting $z = \sqrt{1-p}$ *yields the state-dependent focus trajectory:*

$$\alpha^*(p) \;=\; \frac{1-\sqrt{1-p}}{1+\sqrt{1-p}}. \tag{52}$$

The endpoint and surprisal-linearization conditions thus specify a unique member of the Möbius family.

Proof. **Step 0: Write the general Möbius form.**

A Möbius transformation has the form

$$\alpha(z) = \frac{az+b}{cz+d}, \qquad ad - bc \neq 0. \tag{53}$$

Step 1: Impose the endpoint constraints. We need to find a, b, c, d such that:

1. $\alpha(1) = 0$: $\frac{a(1)+b}{c(1)+d} = 0 \Rightarrow a + b = 0 \Rightarrow b = -a$.
2. $\alpha(0) = 1$: $\frac{a(0)+b}{c(0)+d} = 1 \Rightarrow \frac{b}{d} = 1 \Rightarrow b = d$.

Step 2: Reduce to a one-parameter family. From $b = -a$ and $b = d$, we have $d = -a$. Plugging into Eq. 53 gives

$$\alpha(z) = \frac{az-a}{cz-a} = \frac{z-1}{(\frac{c}{a})z-1}.$$

Let $\kappa := -\frac{c}{a}$ (this re-parameterization is without loss of generality since scaling (a, b, c, d) leaves α unchanged). Then

$$\alpha(z) = \frac{1-z}{1+\kappa z}. \tag{54}$$

For any $\kappa > -1$, this map is well-defined on $z \in [0, 1]$ and is monotone decreasing since

$$\alpha'(z) = -\frac{1+\kappa}{(1+\kappa z)^2} \leq 0. \tag{55}$$

Step 3: Impose self-duality (involution). We additionally require Eq. 50, i.e., $\alpha(\alpha(z)) = z$ for all $z \in [0, 1]$. Substituting Eq. 54 into itself yields

$$\alpha(\alpha(z)) = \frac{1-\frac{1-z}{1+\kappa z}}{1+\kappa\frac{1-z}{1+\kappa z}} = \frac{(1+\kappa)z/(1+\kappa z)}{(1+\kappa)/(1+\kappa z)} = z.$$

Thus, the involution constraint does *not* further restrict κ beyond the domain/range condition $\kappa > -1$. To select a canonical non-affine trajectory, we impose an additional information-theoretic regularity: $\operatorname{arctanh}(\alpha_\kappa(z))$ should be affine in $\ln z$ (equivalently, the normalized-surprisal representation in Prop. 5 should have no additive offset). Using Eq. 61 and $\alpha_\kappa(z) = \frac{1-z}{1+\kappa z}$, we obtain

$$\operatorname{arctanh}(\alpha_\kappa(z)) = \frac{1}{2}\ln\left(\frac{2+(\kappa-1)z}{(\kappa+1)z}\right). \tag{56}$$

This expression is affine in $\ln z$ if and only if $2 + (\kappa - 1)z$ is constant in z, i.e. $\kappa = 1$, which yields the Cayley transform in Eq. 51.

Step 4: Substitute $z = \sqrt{1-p}$ to obtain $\alpha^*(p)$.

Using the Fisher-Rao radius $z = \sqrt{1-p}$:

$$\alpha^*(p) = \alpha\big(\sqrt{1-p}\big) = \frac{1-\sqrt{1-p}}{1+\sqrt{1-p}}. \tag{57}$$

Step 5: Verify the endpoint behavior.

- As $p \to 0$: $z = \sqrt{1-0} = 1$, so $\alpha^*(0) = \frac{1-1}{1+1} = 0$.
- As $p \to 1$: $z = \sqrt{1-1} = 0$, so $\alpha^*(1) = \frac{1-0}{1+0} = 1$.

□

Proposition 4 (Asymptotics of $\alpha^*(p)$) . *The Cayley trajectory satisfies the following expansions:*

$$\alpha^*(p) = \frac{p}{4} + O(p^2) \quad \text{as } p \to 0, \qquad \alpha^*(p) = 1 - 2\sqrt{1-p} + O(1-p) \quad \text{as } p \to 1. \tag{58}$$

In particular, $\alpha^*(p) \in [0,1]$ and is strictly increasing in $p \in (0,1)$.

Proof. For $p \to 0$, use $\sqrt{1-p} = 1 - \frac{p}{2} - \frac{p^2}{8} + O(p^3)$ and substitute into Eq. 52. For $p \to 1$, let $\epsilon = 1 - p$ so $\sqrt{1-p} = \sqrt{\epsilon}$ and expand Eq. 52 as $\epsilon \to 0$. Monotonicity follows by differentiating Eq. 52:

$$\frac{\mathrm{d}\alpha^*}{\mathrm{d}p} = \frac{1}{\sqrt{1-p}\left(1+\sqrt{1-p}\right)^2} > 0.$$

□

Corollary 1 (Endpoint-consistent gate limits for $s(p) = p^{\alpha^*(p)}$) . *Define $s^*(p) \coloneqq p^{\alpha^*(p)}$. Then*

$$\lim_{p\to 0} s^*(p) = 1, \qquad \lim_{p\to 1} \frac{s^*(p)}{p} = 1. \tag{59}$$

Proof. As $p \to 0$, Prop. 4 gives $\alpha^*(p) = O(p)$, hence $\alpha^*(p)|\ln p| \to 0$ and Lemma 3 implies $p^{\alpha^*(p)} \to 1$. As $p \to 1$, we have $\alpha^*(p) \to 1$, so $s^*(p) = p^{\alpha^*(p)} = p \cdot p^{\alpha^*(p)-1}$ and $p^{\alpha^*(p)-1} \to 1$. □

H.2 Information-Theoretic Interpretation via Normalized Surprisal

Proposition 5 (Normalized Surprisal Gating) . *The Cayley trajectory $\alpha^*(p)$ admits an information-theoretic interpretation in terms of normalized error surprisal:*

$$\alpha^*(p) \;=\; \tanh\left(\frac{I_{\mathrm{err}}}{4}\right), \qquad I_{\mathrm{err}} \;\triangleq\; -\ln(1-p). \tag{60}$$

Proof. **Step 0: Recall the inverse hyperbolic tangent.**

The inverse hyperbolic tangent is defined by

$$\operatorname{arctanh}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right), \qquad x \in (-1,1). \tag{61}$$

Step 1: Express $\alpha^*(p)$ in terms of $\operatorname{arctanh}$.

From Thm. 3, we have

$$\alpha^*(p) = \frac{1-\sqrt{1-p}}{1+\sqrt{1-p}}. \tag{62}$$

Let $u \coloneqq \sqrt{1-p}$. Then $\alpha^*(p) = \frac{1-u}{1+u}$.

Taking arctanh of both sides:

$$\operatorname{arctanh}(\alpha^*) = \operatorname{arctanh}\left(\frac{1-u}{1+u}\right) = \frac{1}{2}\ln\left(\frac{1+(1-u)/(1+u)}{1-(1-u)/(1+u)}\right). \tag{63}$$

Step 2: Simplify the argument of ln**.**

Compute the numerator:

$$1 + \frac{1-u}{1+u} = \frac{(1+u)+(1-u)}{1+u} = \frac{2}{1+u}. \tag{64}$$

Compute the denominator:

$$1 - \frac{1-u}{1+u} = \frac{(1+u)-(1-u)}{1+u} = \frac{2u}{1+u}. \tag{65}$$

Thus,

$$\frac{1+(1-u)/(1+u)}{1-(1-u)/(1+u)} = \frac{2/(1+u)}{2u/(1+u)} = \frac{2}{2u} = \frac{1}{u}. \tag{66}$$

Step 3: Substitute back into arctanh**.**

$$\operatorname{arctanh}(\alpha^*) = \frac{1}{2}\ln\left(\frac{1}{u}\right) = -\frac{1}{2}\ln(u) = -\frac{1}{2}\ln\left(\sqrt{1-p}\right) = -\frac{1}{4}\ln(1-p). \tag{67}$$

Step 4: Define the error surprisal and invert.

Let $I_{\text{err}} \triangleq -\ln(1-p)$ be the information content of a prediction error. Then:

$$\operatorname{arctanh}(\alpha^*) = \frac{I_{\text{err}}}{4} \quad \Rightarrow \quad \alpha^*(p) = \tanh\left(\frac{I_{\text{err}}}{4}\right). \tag{68}$$

□

H.3 Formal Properties of DEFT

We recall that for a token context c, the model predictive distribution is $p_\theta(\cdot \mid c) \in \Delta^{|\mathcal{V}|-1}$, and DEFT defines the focus index

$$\alpha_{\text{DEFT}}(c) \triangleq \sum_{v\in\mathcal{V}} p_\theta(v \mid c)^2. \tag{69}$$

The DEFT gate for the supervised target-token probability $p \in (0,1]$ is $\mathcal{G}_{\text{DEFT}}(p,c) = p^{\alpha_{\text{DEFT}}(c)}$.

Proposition 6 (Range and extremizers of α_{DEFT}). *For any distribution $P \in \Delta^{|\mathcal{V}|-1}$,*

$$\frac{1}{|\mathcal{V}|} \le \sum_{v\in\mathcal{V}} P(v)^2 \le 1. \tag{70}$$

Moreover, $\sum_v P(v)^2 = 1$ if P is a point mass (one-hot), and $\sum_v P(v)^2 = 1/|\mathcal{V}|$ if P is uniform.

Proof. The upper bound follows from

$$\Big(\sum_v P(v)\Big)^2 = \sum_v P(v)^2 + 2\sum_{u<v} P(u)P(v) \ \ge \ \sum_v P(v)^2,$$

and since $\sum_v P(v) = 1$, we get $\sum_v P(v)^2 \le 1$. Equality holds iff all cross terms vanish, i.e., $P(u)P(v) = 0$ for all $u \neq v$, which means P is a point mass.

For the lower bound, apply Cauchy–Schwarz to vectors $(1, \dots, 1)$ and $(P(1), \dots, P(|\mathcal{V}|))$:

$$\Big(\sum_v 1 \cdot P(v)\Big)^2 \le \Big(\sum_v 1^2\Big)\Big(\sum_v P(v)^2\Big) \quad \Rightarrow \quad 1 \le |\mathcal{V}| \sum_v P(v)^2.$$

Thus $\sum_v P(v)^2 \ge 1/|\mathcal{V}|$, with equality iff P is proportional to the all-ones vector, i.e., uniform. □

Proposition 7 (Equivalence to Rényi-2 (collision) entropy) . *Let $H_2(P)$ denote the order-2 Rényi entropy:*

$$H_2(P) \triangleq -\ln \sum_{v \in \mathcal{V}} P(v)^2. \tag{71}$$

Then $\alpha_{\mathrm{DEFT}}(c) = \exp\big(-H_2(P_\theta(\cdot \mid c))\big)$. In particular, $\alpha_{\mathrm{DEFT}}(c)$ is a strictly decreasing (one-to-one) function of H_2 on $\Delta^{|\mathcal{V}|-1}$.

Proof. By definition,

$$\exp(-H_2(P)) = \exp\Big(\ln \sum_v P(v)^2\Big) = \sum_v P(v)^2.$$

Since the exponential is strictly monotone and $-\ln(\cdot)$ is strictly decreasing on $(0, \infty)$, the mapping $P \mapsto \alpha_{\mathrm{DEFT}}(P)$ is strictly decreasing in $H_2(P)$. □

Proposition 8 (Gate interpolation between coverage and sharpening) . *Fix any $p \in (0, 1]$ and define $g(\alpha) = p^\alpha$ for $\alpha \in [0, 1]$. Then:*

$$g(0) = 1, \qquad g(1) = p, \qquad \text{and } g(\alpha) \text{ is nonincreasing in } \alpha. \tag{72}$$

Consequently, for any context c,

$$p \;\le\; \mathcal{G}_{\mathrm{DEFT}}(p, c) = p^{\alpha_{\mathrm{DEFT}}(c)} \;\le\; 1, \tag{73}$$

so the DEFT gate continuously interpolates between an "open" gate (coverage) and a "sharpening" gate.

Proof. We have $g(0) = p^0 = 1$ and $g(1) = p^1 = p$. For monotonicity, for $p \in (0, 1)$,

$$\frac{d}{d\alpha} p^\alpha = p^\alpha \ln p \le 0,$$

since $\ln p \le 0$. Hence g is nonincreasing. Plugging $\alpha = \alpha_{\mathrm{DEFT}}(c) \in [1/|\mathcal{V}|, 1]$ from Prop. 6 yields $p \le p^\alpha \le 1$. □

Proposition 9 (General decomposition and bounds linking α_{DEFT} to target probability) . *Fix a context c and supervised target token $\tilde{y}$. Let $p \triangleq p_\theta(\tilde{y} \mid c) \in (0, 1]$. Define the normalized* tail distribution $\bar{p}_\theta(\cdot \mid c) \in \Delta^{|\mathcal{V}|-2}$ *over $\mathcal{V} \setminus \{\tilde{y}\}$ by*

$$\bar{p}_\theta(v \mid c) \triangleq \frac{p_\theta(v \mid c)}{1 - p}, \qquad v \neq \tilde{y} \quad (\textit{when } p < 1). \tag{74}$$

Let the tail concentration be

$$S_{\mathrm{tail}}(c) \triangleq \sum_{v \neq \tilde{y}} \bar{p}_\theta(v \mid c)^2 \in \left[\frac{1}{|\mathcal{V}| - 1}, 1\right]. \tag{75}$$

Then for all $p \in (0, 1)$,

$$\alpha_{\mathrm{DEFT}}(c) = \sum_{v \in \mathcal{V}} p_\theta(v \mid c)^2 = p^2 + (1 - p)^2 \, S_{\mathrm{tail}}(c). \tag{76}$$

Consequently,

$$p^2 + \frac{(1-p)^2}{|\mathcal{V}|-1} \;\le\; \alpha_{\mathrm{DEFT}}(c) \;\le\; p^2 + (1-p)^2. \tag{77}$$

Proof. For $p \in (0,1)$, split the sum into target and non-target parts:

$$\alpha_{\mathrm{DEFT}}(c) = p^2 + \sum_{v \neq \tilde{y}} p_\theta(v \mid c)^2 = p^2 + (1-p)^2 \sum_{v \neq \tilde{y}} \left(\frac{p_\theta(v \mid c)}{1-p} \right)^2 = p^2 + (1-p)^2 S_{\mathrm{tail}}(c),$$

which proves Eq. (76).

It remains to bound $S_{\mathrm{tail}}(c)$. Since $\bar{p}_\theta(\cdot \mid c)$ is a probability distribution over $|\mathcal{V}|-1$ tokens, by the same argument as in Prop. 6,

$$\frac{1}{|\mathcal{V}|-1} \le \sum_{v \neq \tilde{y}} \bar{p}_\theta(v \mid c)^2 \le 1.$$

Substituting these bounds into Eq. (76) yields Eq. (77). □

Remark 1 (Connection to the stylized "spike + uniform tail" case and Cayley) . *Prop. 9 implies that DEFT can be decomposed as*

$$\alpha_{\mathrm{DEFT}}(c) = p^2 + (1-p)^2 S_{\mathrm{tail}}(c),$$

where $S_{\mathrm{tail}}(c)$ captures the shape *(concentration) of the non-target mass. A commonly used stylized model is the "spike + uniform tail" assumption, i.e., $\bar{P}_\theta(\cdot \mid c)$ is uniform over $\mathcal{V} \setminus \{\tilde{y}\}$, which corresponds to $S_{\mathrm{tail}}(c) = 1/(|\mathcal{V}|-1)$. In this special case,*

$$\alpha_{\mathrm{DEFT}}(c) = p^2 + \frac{(1-p)^2}{|\mathcal{V}|-1},$$

which is increasing in p for $p \ge 1/|\mathcal{V}|$. Since the Cayley trajectory $\alpha^*(p)$ is strictly increasing in p on $(0,1)$, both focus indices move in the same direction as confidence grows (trend consistency) under this stylized regime.

Proposition 10 (Conflict suppression under confidently misaligned predictive states) . *Fix a context c and supervised target token $\tilde{y}$, and denote $p \triangleq P_\theta(\tilde{y} \mid c) \in (0,1)$. Assume the predictive distribution is* confidently misaligned *in the sense that there exists a non-target token $v^\star \neq \tilde{y}$ such that*

$$P_\theta(v^\star \mid c) \ge 1 - \varepsilon \qquad \text{for some } \varepsilon \in (0,1). \tag{78}$$

Then the DEFT focus index satisfies $\alpha_{\mathrm{DEFT}}(c) \ge (1-\varepsilon)^2$, and the corresponding target-logit learning-signal magnitude

$$W_{\mathrm{DEFT}}(p,c) \;\triangleq\; p^{\alpha_{\mathrm{DEFT}}(c)}(1-p) \tag{79}$$

is upper bounded as

$$W_{\mathrm{DEFT}}(p,c) \;\le\; p^{(1-\varepsilon)^2}(1-p), \tag{80}$$

which vanishes as $p \to 0$.

In contrast, for the Cayley trajectory $\alpha^*(p)$ in Theorem 3, the Cayley-gated signal $W_{\mathrm{Cayley}}(p) \triangleq p^{\alpha^*(p)}(1-p)$ satisfies

$$\lim_{p \to 0} W_{\mathrm{Cayley}}(p) = 1, \tag{81}$$

Proof. **Step 1: DEFT suppresses the learning signal under severe misalignment.** By definition, $\alpha_{\mathrm{DEFT}}(c) = \sum_v P_\theta(v \mid c)^2$. Since all terms are nonnegative,

$$\alpha_{\mathrm{DEFT}}(c) \ge P_\theta(v^\star \mid c)^2 \ge (1-\varepsilon)^2.$$

For any fixed $p \in (0,1)$, the map $a \mapsto p^a$ is strictly decreasing because $\frac{d}{da} p^a = p^a \ln p < 0$. Therefore,

$$p^{\alpha_{\mathrm{DEFT}}(c)} \le p^{(1-\varepsilon)^2},$$

and multiplying by $(1-p)$ yields

$$W_{\mathrm{DEFT}}(p,c) = p^{\alpha_{\mathrm{DEFT}}(c)}(1-p) \le p^{(1-\varepsilon)^2}(1-p) \xrightarrow[p\to 0]{} 0.$$

Step 2: Cayley gating remains unsuppressed as $p \to 0$. The Cayley trajectory has the closed form

$$\alpha^*(p) = \frac{1-\sqrt{1-p}}{1+\sqrt{1-p}}.$$

Using the expansion $\sqrt{1-p} = 1 - \frac{p}{2} + O(p^2)$ as $p \to 0$, we obtain

$$\alpha^*(p) = \frac{p}{4} + O(p^2).$$

Hence $\alpha^*(p)\ln p \to 0$ as $p \to 0$, and thus

$$\lim_{p\to 0} p^{\alpha^*(p)} = \lim_{p\to 0} \exp\big(\alpha^*(p)\ln p\big) = \exp(0) = 1.$$

Since $(1-p) \to 1$, it follows that $W_{\mathrm{Cayley}}(p) = p^{\alpha^*(p)}(1-p) \to 1$. □

H.4 Stop-Gradient and Differentiable-Focus Updates

Let

$$\ell(p,a) = \frac{1-p^a}{a}, \tag{82}$$

with the continuous extension $\ell(p,0) = -\log p$. For a fixed a, $\partial\ell/\partial p = -p^{a-1}$. Since $\partial p/\partial z_i = p(\delta_{i,\tilde{y}} - P_i)$, detaching the focus index gives

$$\left.\frac{\partial \ell}{\partial z_i}\right|_a = p^a(P_i - \delta_{i,\tilde{y}}). \tag{83}$$

This is also the gradient of $\mathrm{sg}[p^a](-\log p)$ under the same token reduction. Hence detachment preserves the per-token CE direction and changes only its magnitude.

Cayley-Trans. For $a = \alpha^*(p)$, the detached gradient field is generated by the fixed scalar loss

$$f_{\mathrm{CT}}(p) = \int_p^1 u^{\alpha^*(u)-1}\,du, \qquad f'_{\mathrm{CT}}(p) = -p^{\alpha^*(p)-1}. \tag{84}$$

To characterize the differentiable-focus variant, define

$$D(p,a) = \frac{\partial \ell}{\partial a} = \frac{p^a(1-a\log p)-1}{a^2} \le 0, \tag{85}$$

where the sign follows from $e^{-u}(1+u) \le 1$ for $u = -a\log p \ge 0$. Differentiating through the Cayley focus gives

$$\frac{d\ell(p,\alpha^*(p))}{dp} = -p^{\alpha^*(p)-1} + D(p,\alpha^*(p))\alpha^{*\prime}(p). \tag{86}$$

Because $D(p,a) \le 0$ and $\alpha^{*\prime}(p) > 0$, the derivative remains negative. The corresponding logit gradient is therefore still a positive rescaling of the CE direction for each token, although the rescaling changes and can alter the aggregate batch update.

DEFT. For $a = C(P) = \sum_j P_j^2$, the concentration derivative is

$$\frac{\partial C}{\partial z_i} = 2P_i(P_i - C). \tag{87}$$

Differentiating through concentration gives

$$\frac{d\ell}{dz_i} = p^C(P_i - \delta_{i,\tilde{y}}) + 2D(p,C)P_i(P_i - C). \tag{88}$$

The first term is the detached, label-directed update. The second term couples every vocabulary logit through the predictive concentration and is not generally parallel to the CE gradient. In particular, for a dominant non-target token with $P_i > C$, this term is negative because $D(p,C) \le 0$, so gradient descent can further increase that competing logit. Detachment removes this distribution-level feedback while retaining concentration as the scalar token weight. This difference explains why removing stop-gradient changes only the probability-dependent scaling for Cayley-Trans but can change the token-level update direction for DEFT.

I Additional Theoretical Results

I.1 Convexity Properties and Peak Location

Proposition 11 (Convex vs. Concave Objectives) . *Let $f \in C^2[0,1]$ with $f'(p) < 0$ for all $p \in (0,1)$, and define $W_f(p) = -f'(p)\,p(1-p)$.*

1. *If f is concave ($f'' \le 0$), then any maximizer of W_f lies in $[\frac{1}{2}, 1]$.*
2. *If f is convex ($f'' \ge 0$), then any maximizer of W_f lies in $[0, \frac{1}{2}]$.*

Proof. **Step 0: Set up the problem.**

Define $s(p) \coloneqq -f'(p) > 0$. Then

$$W_f(p) = s(p)\,p(1-p). \tag{89}$$

Step 1: Compute the first derivative of W_f.

Using the product rule:

$$W_f'(p) = s'(p)\,p(1-p) + s(p)\,\frac{\mathrm{d}}{\mathrm{d}p}\big[p(1-p)\big] = s'(p)\,p(1-p) + s(p)\,(1-2p). \tag{90}$$

Step 2: Analyze the concave case of $f'' \le 0$.

If $f'' \le 0$, then $s'(p) = -f''(p) \ge 0$. Consider two regions:

Region $p \in (0, \frac{1}{2})$: Here, $1 - 2p > 0$. Both terms in $W_f'(p)$ are nonnegative:

$$s'(p)\,p(1-p) \ge 0, \qquad s(p)\,(1-2p) > 0. \tag{91}$$

Since $s(p) > 0$, we have $W_f'(p) > 0$ on $(0, \frac{1}{2})$. Thus, W_f is strictly increasing on this interval.

Conclusion: Any interior maximizer must lie in $[\frac{1}{2}, 1]$ (since W_f is increasing on $(0, \frac{1}{2})$).

Step 3: Analyze the convex case of $f'' \ge 0$.

If $f'' \ge 0$, then $s'(p) = -f''(p) \le 0$. Consider the region $p \in (\frac{1}{2}, 1)$: Here, $1 - 2p < 0$. Both terms in $W_f'(p)$ are nonpositive:

$$s'(p)\,p(1-p) \le 0, \qquad s(p)\,(1-2p) < 0. \tag{92}$$

Thus, $W_f'(p) < 0$ on $(\frac{1}{2}, 1)$, meaning W_f is strictly decreasing on this interval.

Conclusion: Any interior maximizer must lie in $[0, \frac{1}{2}]$. □

I.2 Quantifying Loss Reduction via Gradient Flow

We characterize the initial change in evaluation risk through the alignment between training and evaluation gradients. This local analysis separates the effects of token gates and cross-context parameter coupling.

Definition 3 (Initial population-risk decrease) . *Let D be the evaluation distribution $c \sim \mu$, $y \sim r(\cdot \mid c)$, and let T be the possibly different training distribution. Define the evaluation risk and training loss by*

$$\mathcal{R}(\boldsymbol{\theta}) = \mathbb{E}_{(c,y)\sim D}[-\ln p_{\boldsymbol{\theta}}(y \mid c)], \qquad \mathcal{L}_f(\boldsymbol{\theta}) = \mathbb{E}_{(c',\tilde{y})\sim T}[f(p_{\boldsymbol{\theta}}(\tilde{y} \mid c'))].$$

Assume differentiability and integrability sufficient to exchange gradients and expectations. Along $\dot{\boldsymbol{\theta}}_t^{(f)} = -\nabla\mathcal{L}_f(\boldsymbol{\theta}_t^{(f)})$, starting at $\boldsymbol{\theta}_0$, define

$$D_f \coloneqq -\left.\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{R}(\boldsymbol{\theta}_t^{(f)})\right|_{t=0} = \lim_{\eta \searrow 0} \frac{\mathcal{R}(\boldsymbol{\theta}_0) - \mathcal{R}(\boldsymbol{\theta}_0 - \eta\nabla\mathcal{L}_f(\boldsymbol{\theta}_0))}{\eta} = \langle \nabla\mathcal{R}, \nabla\mathcal{L}_f \rangle. \tag{93}$$

All gradients below are evaluated at $\boldsymbol{\theta}_0$. Positive D_f means an initial decrease; negative D_f means an increase in evaluation risk.

Theorem 4 (Local first-order identity with cross-context coupling) . *Write $q_c = p_{\boldsymbol{\theta}_0}(\cdot \mid c)$, let $\Phi(c) \in \mathbb{R}^{\dim(\boldsymbol{\theta})\times|\mathcal{V}|}$ have columns $\nabla_{\boldsymbol{\theta}} z_{\boldsymbol{\theta}_0}(c, v)$, and define $K(c, c') = \Phi(c)^\top \Phi(c')$. For two differentiable losses, set $s_i(p) = -p f_i'(p)$ and $a(c, y) = e_y - q_c$. Then*

$$D_2 - D_1 = \mathbb{E}_{(c,y)\sim D}\mathbb{E}_{(c',\tilde{y})\sim T}\left[\big(s_2(q_{c',\tilde{y}}) - s_1(q_{c',\tilde{y}})\big)a(c, y)^\top K(c, c')a(c', \tilde{y})\right]. \tag{94}$$

The two draws are independent. The sign is determined jointly by the gate difference and the cross-context gradient alignment. The identity characterizes infinitesimal, unpreconditioned gradient descent at the shared initial state.

Proof. By the chain rule,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{R}(\boldsymbol{\theta}_t^{(i)})\Big|_{t=0} = -\langle \nabla\mathcal{R}, \nabla\mathcal{L}_{f_i}\rangle, \tag{95}$$

and hence

$$D_2 - D_1 = \langle \nabla\mathcal{R}, \nabla\mathcal{L}_{f_2} - \nabla\mathcal{L}_{f_1}\rangle. \tag{96}$$

The softmax Jacobian yields

$$\nabla\mathcal{R} = -\mathbb{E}_{(c,y)\sim D}[\Phi(c)a(c, y)], \tag{97}$$

$$\nabla\mathcal{L}_{f_i} = -\mathbb{E}_{(c',\tilde{y})\sim T}[s_i(q_{c',\tilde{y}})\Phi(c')a(c', \tilde{y})]. \tag{98}$$

Taking the inner product of these two expectations gives the double expectation

$$D_i = \mathbb{E}_D\mathbb{E}_T[s_i(q_{c',\tilde{y}})a(c, y)^\top K(c, c')a(c', \tilde{y})]. \tag{99}$$

Subtracting proves Eq. 94, with

$$s_2(p) - s_1(p) = -p\big(f_2'(p) - f_1'(p)\big). \tag{100}$$

Within-context orthonormality $K(c, c) = I$ does not remove cross-context terms or turn the product of expectations into a single expectation. □

Assumption 1 (Isolated-context example only) . *For the examples below only, both distributions are supported on the same single context c_0, with deterministic evaluation label y^* and training label $\tilde{y}$, and $\Phi(c_0)^\top\Phi(c_0) = I$. This holds, for example, when the parameters are the logits of a single categorical distribution.*

Two verifiable special cases. Under Assump. 1, write $q = q_{c_0}$ and $\Delta s = s_2(q_{\tilde{y}}) - s_1(q_{\tilde{y}}) \geq 0$. If the labels agree, then

$$\langle e_{y^*} - q, e_{\tilde{y}} - q\rangle = \|e_{y^*} - q\|^2, \tag{101}$$

$$D_2 - D_1 = \Delta s\,\|e_{y^*} - q\|^2 \geq 0. \tag{102}$$

If instead $q = \mathbf{1}/|\mathcal{V}|$ and $y^* \neq \tilde{y}$, then

$$\langle e_{y^*} - q, e_{\tilde{y}} - q\rangle = -1/|\mathcal{V}|, \tag{103}$$

$$D_2 - D_1 = -\Delta s/|\mathcal{V}| \leq 0. \tag{104}$$

These cases illustrate how label alignment changes the effect of a larger gate. A uniform model with correct supervision belongs to the first case; mixed labels and shared parameters are described by the full cross-context identity. Because D_i scales with the loss, comparisons of local decrease rates also depend on the chosen update scale.

J Experimental Details

J.1 Baseline and Proxy Definitions for the Ablation Study

We define the probability- and distribution-level gates used to organize the ablation study (Table 6 , Figure 2). Write $p = p_\theta(\tilde{y} \mid c)$ for the target probability and $q = p_\theta(\cdot \mid c)$ for the full predictive distribution. The mapping protocol is described in Appendix J.2 ; its correspondence to the recorded implementations remains a reproducibility requirement in Appendix J.4 .

Table 6: **Fixed-α ablation on LLaMA-3.1-8B**. Sensitivity sweep, evaluated separately from the primary multi-seed comparison. Within each regime, bold and underlining mark the highest and second-highest values, including ties. Run-level reconciliation requirements are noted in Appendix J.4.

Regime	-logp	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	$-p$	Cayley-Trans	DEFT
MS	6.49	7.80	9.10	9.90	10.30	10.70	<u>10.90</u>	<u>10.90</u>	10.70	<u>10.90</u>	10.03	10.38	**11.35**
MI	50.21	51.00	50.60	**52.00**	51.70	<u>51.90</u>	51.40	51.30	51.30	51.10	49.77	51.48	**52.00**
MW	33.40	32.15	33.24	33.42	33.92	35.93	36.47	29.97	34.67	14.26	14.26	<u>36.95</u>	**46.34**

Probability-level family. Panel (a) of Figure 2 uses the target-logit template $\partial\mathcal{L}/\partial z_{\tilde{y}} = -\mathcal{G}(p)(1-p)$. Adaptive exponents are detached for this template.

- **Hellinger ($\alpha = 0.5$).** $\mathcal{L} = 2(1-\sqrt{p})$, with gate $\mathcal{G}(p) = p^{0.5}$.
- **GCE ($\alpha = 0.7$) (Zhang & Sabuncu, 2018).** $\mathcal{L} = (1-p^{0.7})/0.7$, with gate $\mathcal{G}(p) = p^{0.7}$, a static deformed-log instance.
- **AdaptiveTargetP.** $\mathcal{G}(p) = p^{\alpha(p)}$, with the reported linear mapping $\alpha(p) = 0.1 + 0.8p$.
- **FocalLoss ($\gamma = 2$) (Lin et al., 2018).** For $\mathcal{L} = -(1-p)^{\gamma}\ln p$, the gate obtained by $-p\mathcal{L}'(p)$ is
$$\mathcal{G}(p) = (1-p)^{\gamma} - \gamma p(1-p)^{\gamma-1}\ln p.$$
For $\gamma > 0$ the gate approaches 1 as $p \to 0$ and 0 as $p \to 1$, down-weighting easy targets.
- **Cayley-Trans (ours).** $\mathcal{G}(p) = p^{\alpha^*(p)}$, where $\alpha^*(p) = (1-\sqrt{1-p})/(1+\sqrt{1-p})$ is the construction in Appendix H.

Distribution-level family. Panel (b) uses $\mathcal{G}(q) = p^{\alpha(q)}$. A larger concentration proxy maps to a larger α and hence a smaller gate at fixed $0 < p < 1$.

- **DEFT-Shannon.** Use $-H(q)$ as the concentration proxy, where $H(q) = -\sum_v q_v \ln q_v$.
- **DEFT-Margin.** Use the top-two gap $m(q) = q_{(1)} - q_{(2)}$, with an increasing mapping from margin to α.
- **DEFT-EffSupportInv.** Use the inverse entropy-based effective support $1/\mathrm{ESS}(q) = e^{-H(q)}$, where $\mathrm{ESS}(q) = e^{H(q)}$.
- **DEFT-Rényi-2 (ours, default).** $\alpha(q) = \sum_v q_v^2 = e^{-H_2(q)} \in [1/|\mathcal{V}|, 1]$, used directly without normalization.

Static-α sweep. The eleven columns of Table 6 use $\alpha \in \{0, 0.1, \ldots, 0.9, 1\}$ on LLaMA-3.1-8B. The loss is $(1-p^{\alpha})/\alpha$, with $-\ln p$ at $\alpha = 0$ and $1-p$ at $\alpha = 1$ (equivalent to $-p$ up to a constant).

J.2 Proxy-to-Focus Mapping

The non-default proxy protocol specifies batch-wise z-score normalization, followed by a sigmoid and an exponent clamp to $[\alpha_{\min}, \alpha_{\max}]$, increasing with the concentration proxies above. Exact normalization axes, stabilization constants, clamp bounds, and correspondence to the recorded implementations remain to be verified (Appendix J.4).

Endpoint behavior. For finite $0 < p < 1$ and finite α, the gate p^{α} and target-logit magnitude $p^{\alpha}(1-p)$ are positive. With $\alpha \geq \alpha_{\min} > 0$, the gate satisfies $p^{\alpha} \leq p^{\alpha_{\min}} \to 0$ as $p \to 0$; as $p \to 1$, the error factor tends to zero. Thus the update magnitude depends on both the mapped focus parameter and the target probability.

Comparison scope. Default DEFT uses the raw Rényi-2 concentration directly. The alternative designs additionally normalize and clamp their proxies, so the ablation evaluates complete statistic–mapping combinations. Separating these two factors requires matched mapping controls.

J.3 Experimental Setup

Datasets. For the Model-Strong (MS) regime we use the NuminaMath corpus (LI et al., 2024) of competition-level mathematical reasoning. For the Model-Intermediate (MI) regime we use the M1-23K medical reasoning dataset (Huang et al., 2025). For the Model-Weak (MW) regime we synthesize FigFont character-art puzzles following Stojanovski et al. (2025), where the model decodes glyphs rendered with FIGlet-style ASCII fonts. FigFont operationalizes a weak-prior setting here; it does not establish the absence of all relevant pretraining knowledge.

Backbones. The reported experiments cover seven open-source backbones, with general-purpose and math-specialized variants: LLaMA-3.2-3B and LLaMA-3.1-8B (Grattafiori et al., 2024); DeepSeekMath-7B (Shao et al., 2024); Qwen2.5-Math-1.5B and Qwen2.5-Math-7B (Yang et al., 2024); and Qwen2.5-1.5B and Qwen2.5-7B (Qwen et al., 2025). The evaluated subset differs by regime. Regime labels describe the model–task pairing rather than an intrinsic property of a backbone.

Training and optimization. The reported setup uses the **VERL toolkit** (Sheng et al., 2024) on a single node of $8\times$NVIDIA H20 GPUs, with `AdamW`, cosine learning-rate decay, and warm-up ratio 0.1. The reported base learning rate is 5×10^{-5} except for LLaMA-3.1-8B (2×10^{-5}), with 1 training epoch. Batch size and sequence length are described as matched within a regime, but their exact values and run-level matching require verification.

Objective configurations. We compare Cayley-Trans and DEFT with $-\log p$, $-p$, Hellinger ($\alpha = 0.5$), GCE ($\alpha = 0.7$), EAFT (Diao et al., 2026), InfoSFT (Sabbaghi et al., 2026), and RankTuner (Yu et al., 2026). The recent baselines use entropy-, confidence-, or probability/rank-based token weighting. InfoSFT uses the reported 0.93 setting. RankTuner uses its p variant for math and its one variant for medical and FigFont. Cayley-Trans and DEFT use state-dependent focus indices within the deformed-log objective. Implementation and reproducibility requirements are summarized in Appendix J.4.

J.4 Experimental Records and Reproducibility

The primary comparison in Table 1 – Table 3 reports per-benchmark means and task-level mean $\pm$ standard deviation for six task–backbone configurations, each evaluated with five random seeds. The broader backbone evaluation and sensitivity sweeps are reported separately; the added Qwen FigFont baseline block reports five-seed means, with per-seed intervals still to be tabulated. Complete paired records are needed for paired confidence intervals and effect-size estimation. Numerical differences between nominally matching configurations in the backbone evaluation and focus sweep also require run-level reconciliation.

A complete run manifest remains to be established: exact training/evaluation sample and token counts; dataset versions, preprocessing and splits; global and per-device batch sizes, accumulation, sequence lengths and packing; numerical precision; decoding prompts and settings; checkpoint identifiers and selection rules; and baseline code versions, hyperparameters, and stop-gradient conventions. Verification of budget matching and of the proxy formulas, normalization axes, stabilization constants, and clamp bounds against their implementations is pending.

For the differentiable-focus ablation, the sharp FigFont degradation under *detach off*—including the large seed-to-seed variance of the Qwen run (24.97 ± 13.31)—requires training and output-level diagnostics. Separating concentration feedback from numerical instability, parsing failures, and effective update-scale changes requires corresponding checks.

J.5 Training Cost

DEFT adds 0.24% wall-clock time and 4.36% peak memory over NLL in this configuration, with throughput changing from 306 to 305 tokens/s. The concentration reduction is $O(|\mathcal{V}|)$ per token, whereas Cayley-Trans uses scalar operations once the target probability is available. These measurements are implementation- and configuration-specific; profiling repetition counts, probability materialization, and kernel-fusion details remain to be documented.

Table 7: Training cost on math with LLaMA-3.1-8B.

Method	Time / 100 steps (s)	Throughput (tok/s)	Peak memory (GB)
NLL	4250.15 ± 4.60	306	48.90
Cayley-Trans	4241.84 ± 5.53	306	45.96
DEFT	4260.53 ± 5.70	305	51.03

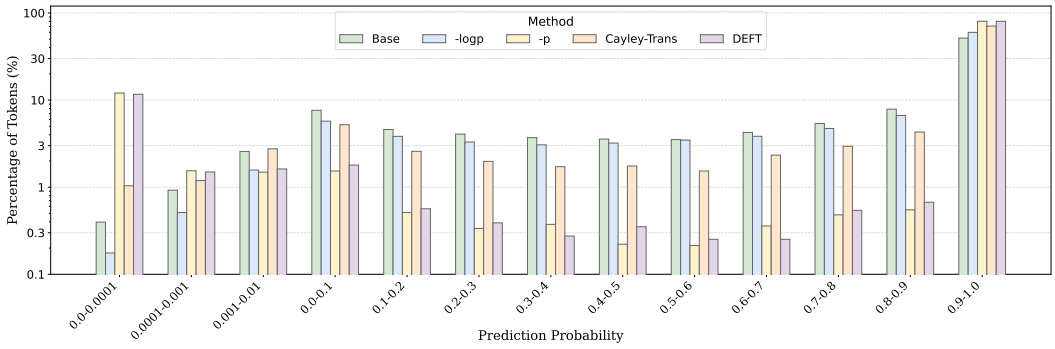

Figure 3: Token probability distributions on the training set under different objectives in the Model-Strong regime.

K Token Probability Distributions Across Regimes

This section collects the full token probability distributions on the training set under different objectives, across all three capability regimes (MS, MI, MW).

In the Model-Strong (MS) regime (Figure 3), the objectives produce different low- and high-probability target-token distributions. Cayley-Trans and DEFT shift mass toward high-confidence regions in the displayed runs. Inspected low-probability examples include formatting and function tokens, whose semantic roles depend on context. In the Model-Weak (MW) FigFont regime (Figure 4), many target tokens have low probability, where the linear-loss target-logit magnitude $p(1-p)$ is small. Cayley-Trans approaches the NLL gate as $p \to 0$, while DEFT approaches it when $\alpha|\ln p|$11. The two rules therefore provide different paths to retaining updates on low-probability targets. The Model-Intermediate (MI) distributions (Figure 5) also show objective-dependent probability shifts. Together, these plots characterize changes in supervised-token confidence across tasks and complement the benchmark scores and gradient-landscape analysis.

L Token-Level Gradient Distributions Across Regimes

This section collects the token-level gradient distributions across all three capability regimes: Model-Strong (MS, MATH; Figure 6), Model-Intermediate (MI, medical reasoning; Figure 7), and Model-Weak (MW, FigFont character-art; Figure 8). All panels share the same axes and color coding—the x-axis shows target-token probability, the y-axis predictive entropy, and color intensity the normalized gradient magnitude—with four sub-plots per regime, one per representative objective ($-\log p$, $-p$, Cayley-Trans, and DEFT). The qualitative patterns described in the main text—NLL concentrating on low-probability long-tail tokens, $-p$ exhibiting head bias, and the dynamic gating of Cayley-Trans and DEFT modulating gradient magnitude by model-internal uncertainty—persist across all three regimes.

M Additional-Backbone Evaluation

Table 8 , Table 9 , and Table 10 report results for additional model configurations in the Model-Strong (math), Model-Intermediate (medical), and Model-Weak (FigFont) settings. Together with the

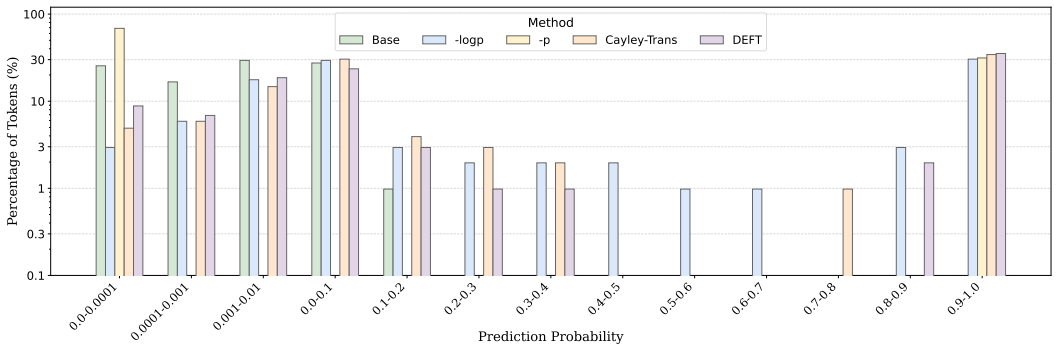

Figure 4: Token probability distributions on the training set under different objectives in the Model-Weak regime.

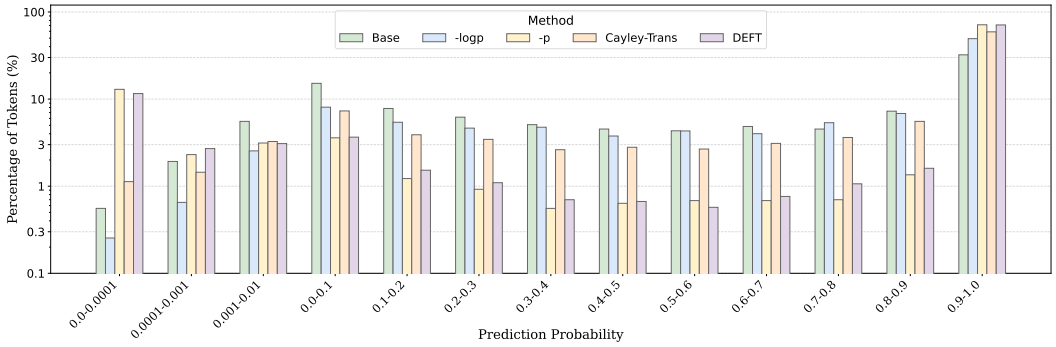

Figure 5: Token probability distributions on the training set under different objectives in the Model-Intermediate regime.

primary LLaMA and Qwen configurations in the main text, these tables span seven backbones and provide complementary model-coverage evidence.

N Per-Benchmark Cross-Regime OOD Results behind Table 11

This section reports the per-benchmark transfer scores underlying the regime-averaged entries of Table 11. All experiments use LLaMA-3.1-8B as the backbone; the source-regime corpus is used for fine-tuning and the target-regime benchmark suite for zero-shot evaluation. Table 12, Table 13, and Table 14 respectively cover MS $\rightarrow$ MI, MI $\rightarrow$ MS, and MW $\rightarrow$ MI; the row averages in each table match the corresponding rows of Table 11.

O Model-Size Ablation

Setup. To verify that the gains of Cayley-Trans and DEFT are not specific to a particular backbone scale, we re-run the three-regime evaluation on the Qwen3-base family at five sizes: 0.6B, 1.7B, 4B, 8B, and 14B. All other settings (datasets, optimizer, batch size, sequence length, learning-rate schedule, 1 epoch) follow Appendix J.3; only the loss function and the backbone size are varied. Figure 9 reports MATH Avg (MS), medical Avg (MI), and FigFont JW-Sim (MW) absolute scores across scales.

Regime-specific observations. The Qwen3 runs exhibit task-dependent objective sensitivity: $-p$ and DEFT are competitive on MATH, medical scores are more tightly grouped, and $-p$ has poor FigFont scores at smaller scales. The preferred loss varies with both task and model size.

Cross-regime takeaway. This sweep extends the evaluation from 0.6B to 14B parameters. MATH accuracy, medical accuracy, and FigFont similarity are reported separately to preserve their task-

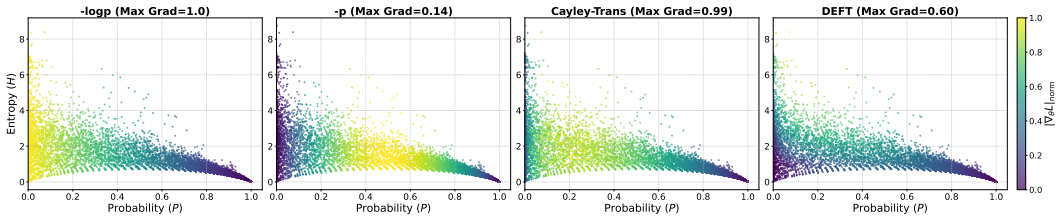

Figure 6: Token-level gradient distributions in the Model-Strong (MATH) regime for LLaMA-3.1-8B. The x-axis shows target-token probability, the y-axis predictive entropy, and color intensity the normalized gradient magnitude; the four sub-plots correspond to $-\log p$, $-p$, Cayley-Trans, and DEFT.

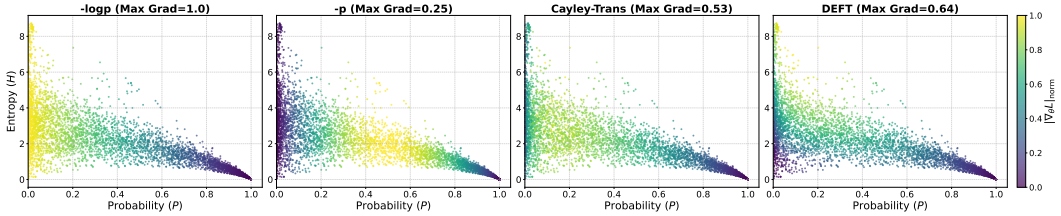

Figure 7: Token-level gradient distributions in the Model-Intermediate regime (M1-23K medical reasoning).

specific interpretation. These trends motivate repeated evaluations at these scales, particularly where method scores are closely grouped.

P Semantics: Qualitative Inventory of the Three Regions

The token inventory in Figure 10 (b) contains $n = 15{,}020$ target tokens. Three selected regions in the $(p, H(p_\theta(\cdot \mid c)))$ plane have mean concentration indices $\alpha \approx 0.08$ (UL), 0.81 (LL), and 0.99 (LR). The associated gradient multiplier is p^α. These regional summaries describe token examples at different predictive states; a controlled comparison of concentration effects would additionally match target probabilities within the low-p regions.

UL: diffuse low-target-probability predictions ($\alpha \approx 0.08$). This region has $p < 0.2$ and $H > 3$ nats. Examples include discourse connectives such as `Since`, `Thus`, and `Hence`; reasoning verbs such as `discuss`, `recognize`, and `derived`; and leading sub-tokens of domain terms such as `centroid`. At fixed p, a smaller α yields a larger gate, while the target-logit magnitude is determined jointly by p^α and $1 - p$.

LL: concentrated disagreement ($\alpha \approx 0.81$). This region has $p < 0.2$ and $H < 0.8$ nats. Examples include LaTeX formatting choices and function words such as `the`, `and`, and `is`. DEFT attenuates these targets relative to NLL through $p^\alpha < 1$. Whether this reduces fitting of interchangeable surface forms or suppresses necessary correction depends on the context and the validity of the model's competing prediction.

LR: high-target-probability predictions ($\alpha \approx 0.99$). The region $p > 0.7$ and $H < 0.3$ nats contains 60.7% of the sample (9,110 of 15,020 tokens). Examples include LaTeX commands (`frac`, `sqrt`, `cdot`), digit and variable continuations, and frequent mathematical collocations. With $\alpha \approx 1$, the gate satisfies $p^\alpha \approx p$ and approaches one as $p \to 1$; the target-logit gradient decreases through the shared residual factor $1 - p$.

Q Limitations

Task and model coverage. The weak-prior evaluation uses the synthetic FigFont task. The primary comparison covers six task–backbone configurations with five random seeds. Additional real-world weak-prior tasks and repeated evaluations across model sizes would strengthen external validity.

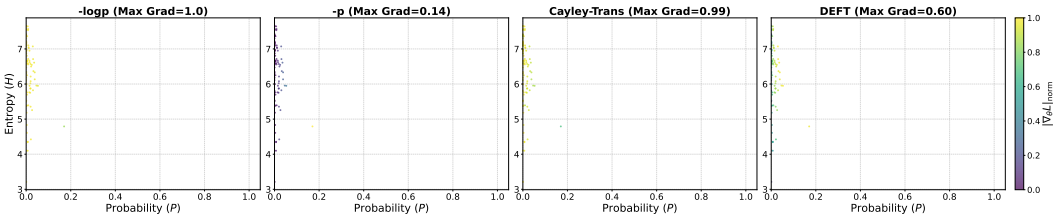

Figure 8: Token-level gradient distributions in the Model-Weak regime (FigFont character-art).

Table 8: Average@16 accuracy under the Model-Strong setting. Best results are in **bold** and the second-best are underlined.

Models	Methods	Math500	Minerva Math	Olympiad Bench	AIME24	AMC23	Avg
DeepseekMath-7B	Base	5.70	2.89	1.51	0.00	2.34	2.49
	-logp	28.48	9.91	6.65	0.21	8.75	10.80
	-p	39.52	19.28	14.51	1.03	17.18	**18.30**
	Hellinger (α=0.5)	39.80	14.50	12.70	0.80	17.70	17.10
	GCE (α=0.7)	41.40	15.50	13.00	1.00	18.80	17.90
	InfoSFT-0.93	37.24	15.09	10.77	0.83	17.19	16.22
	EAFT	28.06	10.33	6.31	0.63	9.84	11.03
	RankTuner-p	41.36	14.09	13.32	1.46	18.59	17.77
	Cayley-Trans	39.03	13.78	12.26	0.41	16.72	16.44
	DEFT	40.90	15.48	12.63	1.65	20.31	18.19
Qwen2.5-Math-1.5B	Base	30.71	8.81	14.88	2.49	17.97	14.97
	-logp	41.70	11.81	11.39	1.45	17.66	16.80
	-p	63.48	25.14	26.37	6.47	39.22	**32.14**
	Hellinger (α=0.5)	59.40	23.50	23.60	2.50	31.40	28.10
	GCE (α=0.7)	61.30	24.40	25.10	3.10	33.10	29.40
	InfoSFT-0.93	55.36	18.85	20.13	2.90	30.94	25.64
	EAFT	40.48	11.86	11.50	0.41	14.53	15.76
	RankTuner-p	62.80	23.88	25.58	5.64	37.19	30.02
	Cayley-Trans	56.04	21.75	21.96	1.86	28.44	26.01
	DEFT	61.41	24.13	24.96	5.63	35.63	30.35

Predictive confidence and correction. Cayley-Trans and DEFT derive their focus parameters from the evolving model state. Concentrated disagreement can reflect a valid alternative or a confidently wrong prior, so useful corrections may be attenuated. The current token-level diagnostics and transfer evaluations do not directly measure retained general capabilities or success at correcting verified wrong priors. Dedicated retention benchmarks and clean-label correction experiments are needed to characterize these tradeoffs. Matched mapping and update-scale controls would further isolate the contribution of non-target distributional information.

Experimental records and implementation. Per-seed outputs, full run configurations, and verification of proxy implementations remain outstanding as detailed in Appendix J.4 . These gaps limit statistical comparisons and reproducibility. Runtime and memory measurements apply to the tested implementation; their dependence on sequence length, vocabulary size, and kernel fusion requires further evaluation.

Table 9: Results for additional backbones in the Model-Intermediate setting (**bold**: best; underlined: second-best).

Method	MedMc	MedQA	PubMed	MMLU-P	GPQA	Lancet	MedB(4)	MedB(5)	MedX	NEJM	Avg
LLaMA-3.2-3B											
Base	21.30	21.92	22.60	11.40	23.08	25.00	23.05	15.26	10.35	23.22	19.72
-logp	42.22	48.86	66.60	35.44	32.82	44.90	45.13	37.34	12.56	44.44	**41.03**
-p	40.09	43.36	65.60	34.07	38.72	42.72	35.71	33.44	11.18	38.31	38.32
Hellinger (α=0.5)	43.40	45.20	67.30	37.30	32.30	42.20	42.90	34.40	10.10	42.10	39.70
GCE (α=0.7)	43.00	42.30	65.40	36.50	25.90	43.40	38.60	32.50	10.50	39.80	37.80
InfoSFT-0.93	44.42	42.89	66.80	35.64	34.62	44.90	38.64	34.09	10.56	41.63	39.42
EAFT	43.27	47.05	68.20	39.15	27.18	44.42	38.96	36.36	11.59	42.79	39.90
RankTuner-p	41.62	42.89	65.60	34.66	31.03	43.93	36.69	33.12	11.73	37.98	37.92
Cayley-Trans	42.67	46.43	66.80	37.65	32.56	45.63	41.56	35.71	10.63	39.30	39.89
DEFT	42.31	46.27	66.20	36.35	33.08	47.33	39.94	33.44	11.73	42.45	39.91
Qwen2.5-1.5B											
Base	22.21	21.84	18.50	11.21	24.36	22.57	24.03	17.53	10.84	18.74	19.18
-logp	38.85	40.06	68.80	35.24	32.56	36.89	36.69	28.25	9.94	32.50	35.98
-p	37.39	39.75	66.90	35.70	27.69	41.75	32.79	31.17	11.73	38.97	36.38
Hellinger (α=0.5)	38.80	38.90	65.80	34.00	26.20	43.40	36.70	29.90	11.20	35.30	36.00
GCE (α=0.7)	39.80	37.80	65.60	34.20	31.30	37.60	38.30	30.50	12.40	34.00	36.10
InfoSFT-0.93	40.07	37.31	67.50	35.18	26.92	39.56	30.52	32.47	11.11	37.48	35.81
EAFT	40.57	40.53	68.10	35.31	37.18	39.56	35.71	27.92	10.14	35.49	37.05
RankTuner-p	40.14	39.20	68.10	34.07	28.21	38.59	37.34	27.27	11.73	34.33	35.90
Cayley-Trans	39.23	39.75	67.10	35.44	28.72	39.56	34.42	26.30	11.25	32.17	35.39
DEFT	37.84	41.08	69.50	35.44	34.62	38.83	39.61	32.79	12.42	32.34	**37.45**

Table 10: Results for additional backbones in the Model-Weak (MW) setting. Performance is measured by Jaro–Winkler similarity; the added baselines (InfoSFT, EAFT, and RankTuner) are five-seed means ($\times 100$) from a separate evaluation protocol, reported without per-seed intervals, so rankings that mix them with the single-run entries are indicative. Bold and underlining mark the highest and second-highest scores among fine-tuned methods, excluding the Base reference.

Models	Base	-logp	-p	Hellinger (α=0.5)	GCE (α=0.7)	InfoSFT-0.93	EAFT	RankTuner	Cayley-Trans	DEFT
LLaMA-3.2-3B	41.89	40.37	7.68	36.59	36.74	35.80	48.59	31.88	46.55	**60.22**
Qwen2.5-1.5B	35.32	33.65	28.13	34.39	35.74	**37.86**	34.65	34.85	34.21	33.66

Table 11: **Cross-regime OOD evaluation on LLaMA-3.1-8B**: each row reports the benchmark-averaged transfer score after fine-tuning on the source-regime corpus and evaluating on the target-regime benchmarks. Per-benchmark scores are deferred to the tables below. Best in **bold**, second-best underlined.

Source → Target	$-\log p$ ($\alpha \to 0$)	$-p$ ($\alpha = 1$)	**Hellinger** ($\alpha = 0.5$)	**GCE** ($\alpha = 0.7$)	**Cayley-Trans** (probability-level)	**DEFT** (distribution-level)
MS → MI (NuminaMath → Medical Avg)	43.14	44.06	43.90	44.70	44.86	**45.35**
MI → MS (M1-23K → MATH Avg@16)	5.60	8.80	5.60	9.50	8.00	**10.20**
MW → MI (FigFont → Medical Avg)	35.60	22.10	27.60	30.30	**40.10**	38.70

Table 12: **MS → MI per-benchmark transfer scores** (LLaMA-3.1-8B fine-tuned on NuminaMath, evaluated on 10 medical benchmarks). Best in **bold**, second-best underlined.

Method	**MedMc**	**MedQA**	**PubMed**	**MedX**	**MMLU-P**	**GPQA**	**Lancet**	**MedB(4)**	**MedB(5)**	**NEJM**	**Avg**
$-\log p$	41.45	50.51	71.40	13.18	43.26	42.56	40.29	43.83	36.36	48.59	43.14
$-p$	46.31	54.05	74.20	12.63	43.06	35.90	47.82	41.88	33.12	51.58	44.06
Hellinger ($\alpha = 0.5$)	44.50	50.40	73.70	13.70	45.30	38.70	46.60	37.70	40.60	47.80	43.90
GCE ($\alpha = 0.7$)	45.10	51.50	75.00	13.80	43.60	39.50	50.20	36.40	43.80	48.30	44.70
Cayley-Trans	45.49	53.02	77.50	13.46	45.67	41.03	48.54	41.88	32.79	49.25	44.86
DEFT	45.90	53.57	75.60	13.94	46.06	43.59	45.87	45.13	36.04	47.76	**45.35**

Table 13: **MI → MS per-benchmark transfer scores** (LLaMA-3.1-8B fine-tuned on M1-23K medical reasoning, evaluated on 5 MATH-style benchmarks, Avg@16). Best in **bold**, second-best underlined.

Method	**Math500**	**Minerva**	**Olympiad**	**AIME24**	**AMC23**	**Avg**
$-\log p$	12.30	7.10	2.90	0.00	5.90	5.60
$-p$	18.70	11.60	4.50	0.00	9.40	8.80
Hellinger ($\alpha = 0.5$)	10.90	9.90	2.70	0.40	4.10	5.60
GCE ($\alpha = 0.7$)	20.50	13.00	4.50	0.20	9.20	9.50
Cayley-Trans	17.50	11.00	4.20	0.20	7.00	8.00
DEFT	21.30	15.90	4.70	0.60	8.40	**10.20**

Table 14: **MW → MI per-benchmark transfer scores** (LLaMA-3.1-8B fine-tuned on FigFont character-art, evaluated on 10 medical benchmarks). Best in **bold**, second-best underlined.

Method	MedMc	MedQA	PubMed	MMLU-P	GPQA	Lancet	MedB(4)	MedB(5)	MedX	NEJM	Avg
$-\log p$	33.40	39.40	72.20	30.90	21.80	39.10	34.10	31.80	12.30	40.60	35.60
$-p$	21.40	24.00	55.20	9.70	26.90	19.90	20.10	14.60	13.30	15.90	22.10
Hellinger ($\alpha=0.5$)	26.30	29.60	68.40	10.20	25.60	27.40	25.30	22.70	10.60	29.50	27.60
GCE ($\alpha=0.7$)	29.80	35.10	66.80	13.60	31.00	31.30	32.80	21.40	10.90	30.00	30.30
Cayley-Trans	37.80	43.40	72.20	36.60	27.40	45.60	39.90	36.00	11.50	50.10	**40.10**
DEFT	35.20	46.00	76.20	23.30	27.70	45.60	39.90	34.70	13.50	44.80	38.70

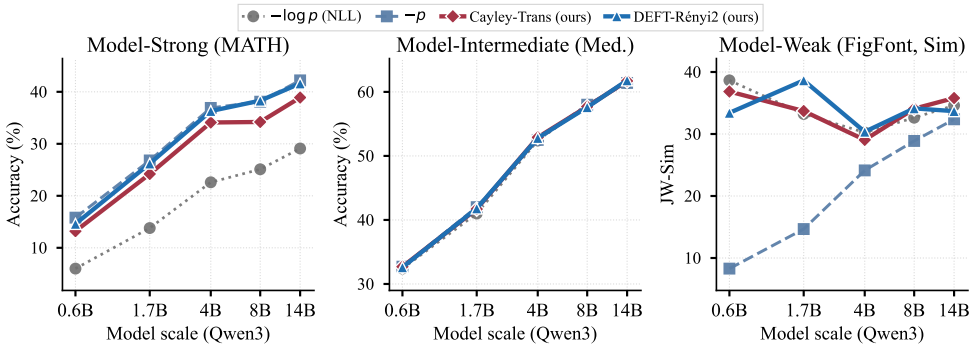

Figure 9: **Model-size ablation on Qwen3-base from 0.6B to 14B across the three capability regimes**: Model-Strong (MATH Avg, left), Model-Intermediate (Medical Avg, middle), and Model-Weak (FigFont JW-Sim, right). Four losses are compared: $-\log p$ (NLL), $-p$, Cayley-Trans (probability-level, ours), and DEFT-Rényi2 (distribution-level, ours).

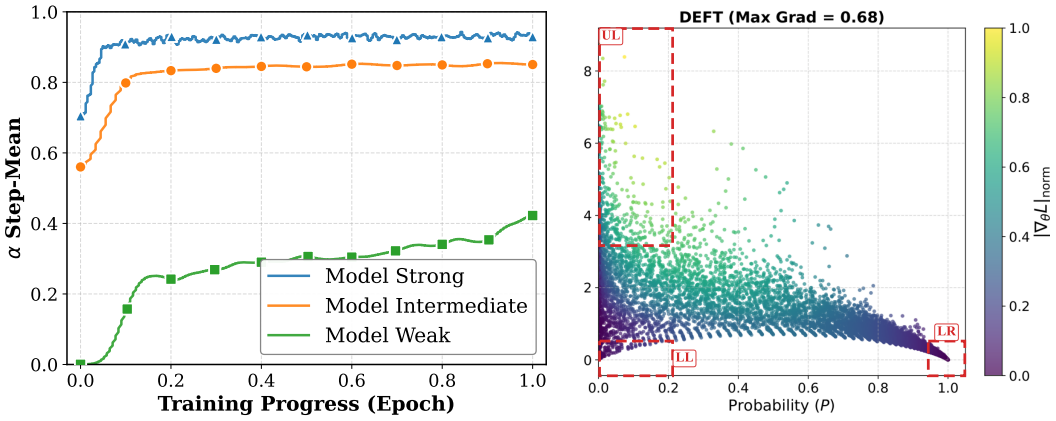

Figure 10: **Two views of the adaptive gate α. (a) Time.** Batch-averaged α along DEFT training of LLaMA-3.1-8B across three regimes. **(b) Semantics.** Three distinct semantic regions in the DEFT distribution. The *space* view (mean α vs. target probability p on the MS regime) is shown as the far-right panel of Figure 2 in the main text.

R Answer Corruption and Concentrated-Disagreement Frequency

Training with corrupted answers. On Qwen2.5-Math-7B, we replace the final answers in 5% of NuminaMath-CoT training examples with incorrect answers, fine-tune under the same reported training settings, and evaluate on the original mathematics benchmarks. Table 15 reports the benchmark-average accuracy and its standard deviation; clean-training means are included for context. This is answer corruption, not a verified test of correcting confidently wrong model priors.

Table 15: Qwen2.5-Math-7B math accuracy under 5% final-answer corruption. Corrupted scores are reported as mean $\pm$ SD; clean scores are means from the primary comparison.

Method	Clean Avg	Corrupted Avg
NLL	23.80	21.65 ± 0.12
EAFT	23.16	21.64 ± 0.25
InfoSFT	31.89	29.81 ± 0.32
RankTuner-p	36.01	35.60 ± 0.46
Cayley-Trans	34.70	30.34 ± 0.69
DEFT	36.73	34.66 ± 1.00

DEFT remains 13.01 points above NLL under corruption, but RankTuner-p has the highest corrupted-data mean. Relative to the reported clean means, DEFT drops 2.07 points, compared with 2.15 for NLL and 0.41 for RankTuner-p. Thus, these results support competitive performance under this corruption setting, not the smallest degradation or uniformly superior noise robustness. These are descriptive differences, not paired significance estimates or direct measurements of general-capability retention.

Frequency of concentrated disagreement on clean supervision. We perform one forward pass with Qwen2.5-Math-7B base on $5{,}000$ clean NuminaMath-CoT examples. Response tokens are assigned to answer, computation, reasoning, or format categories by rules. The "confident-but-wrong" (CBW) proxy requires both low Shannon entropy H (bits) and low supervised-target probability p: Loose uses $H \leq 0.5$, $p \leq 0.3$; Strict uses $H \leq 0.3$, $p \leq 0.1$. Table 16 gives the fraction of tokens satisfying each condition, overall and within each category.

Table 16: CBW-proxy frequency (%). Overall and answer columns also give matching token counts. Each category percentage uses its own token denominator.

Threshold	Overall (count)	Answer (count)	Computation	Reasoning	Format
Loose	1.10 (23,011)	2.55 (1,684)	0.93	1.14	1.28
Strict	0.62 (13,098)	1.75 (1,157)	0.54	0.61	0.73

The measured region is infrequent under both thresholds in this sample. Its frequency does not establish that suppressing it is harmless: correct supervision can still fall in this region, and a probability-based proxy does not independently verify semantic error. These observations quantify prevalence, while Table 5 tests a training intervention; neither replaces a dedicated clean-label correction or capability-retention evaluation.