Watch Your Norms: On Test-Time Expert Reduction in Multilingual MoEs

Anonymous ACL submission

Abstract

Test-time expert reduction (top-k reduction) is a popular way to cut inference cost in Mixture-of-Experts (MoE) language models, avoiding the expensive model retraining. Existing work, however, analyzes the effects of top-k reduction almost exclusively on English benchmarks. In this work, we provide the first systematic multilingual study of test-time top-k reduction across three contemporary MoE models, 100 languages, and 5 downstream tasks. We show that down to 75% of retained experts, the models degrade gracefully, but overall retention collapses sharply at low k and widens a structural *multilingual gap*: low-resource languages degrade much faster than high-resource ones. We trace the gap to a renormalization-induced norm overshoot in the MoE block, where the renormalization over the surviving expert weights inflates the MoE contribution to the residual stream. This inflation harms low-resource languages disproportionately due to their smaller residual norms. As a remedy, we propose two training-free and latency-free interventions: (i) a grid-searched global scalar to rescale the representations output by the MoE layer (global α) and (ii) a parameter-free per-token correction (token α) of expert weights derived from the routing. The proposed interventions substantially improve overall retention and reduce the multilingual gap.

1 Introduction

Mixture-of-Experts (MoE) has become a standard mechanism for scaling Transformer-based language models, replacing dense feed-forward layers with a set of small expert networks and a gating function that activates only the top-k experts per token (Shazeer et al., 2017; Lepikhin et al., 2020; Fedus et al., 2022). This sparse activation pattern decouples the total parameter count of a model from actual per-token compute, enabling the training and deployment of ever larger MoE models

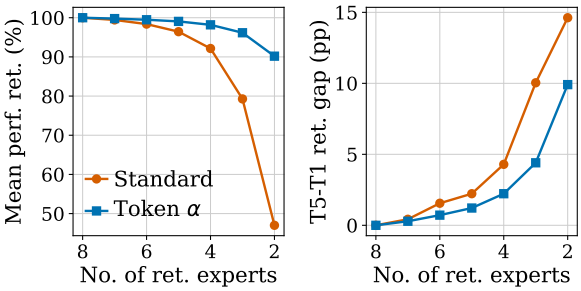


Figure 1: Top-k reduction on Qwen3.5-35B-A3B. **Left**: *mean performance retention* as the number of activated experts is reduced from k=8; **Right**: *retention gap* as the difference in mean retention between the high-resource (T5) and low-resource languages (T1), over the same axis. Standard top-k reduction (orange) collapses retention and widens the multilingual gap; our *token* α intervention (blue) substantially mitigates both.

(Jiang et al., 2024; Liu et al., 2024; Yang et al., 2025). However, the large cumulative number of activated experts across layers and tokens remains a critical latency bottleneck at inference time, particularly in resource-constrained deployment scenarios. A growing line of work addresses this bottleneck through test-time expert reduction (top-k reduction): reducing the number of activated experts per token without retraining the model (Lu et al., 2024; Zhong et al., 2024; Chitty-Venkata et al., 2025; Liu et al., 2026). Because these methods operate at test-time and require only lightweight calibration, they are appealing for practical deployment.

Existing work analyzes top-k reduction almost exclusively through the lens of English benchmarks, leaving open the question of how it affects the multilingual capabilities of MoE models. This question is orthogonal to the methodological refinements that distinguish recent reduction methods—whether k is allocated adaptively across layers (Zhong et al., 2024; Chitty-Venkata et al., 2025) or across tokens (Lu et al., 2024; Liu et al., 2026). To isolate the effect of reducing k itself from the confound of *which* experts an adaptive

scheme selects, we focus on uniform top-k reduction, and study how a model's multilingual behavior responds to a reduced activation budget.

We analyze three recent multilingual MoE models—Qwen3-30B-A3B (Yang et al., 2025), Qwen3.5-35B-A3B (Qwen Team, 2026), and Gemma4-26B-A4B (Gemma Team, 2026)—across 100 typologically diverse languages covering a spectrum from high- to low-resource. Figure 1 previews our findings for Qwen3.5: as the activation budget is reduced, the mean retention—the proportion of performance on downstream tasks that is retained w.r.t the full activation budget (i.e., $k = 8$)—collapses, and the gap in mean retention (i.e., relative performance retention) between high- and low-resource languages (T5 and T1, respectively) sharply widens, reaching 14.6pp for $k = 2$. Put simply, a larger reduction of the activation budget not only drastically reduces the performance for all languages, but does more so (i.e., disproportionately) for low-resource languages. We observe a similar pattern, with varying strength, across all three models, indicating a structural rather than model-specific phenomenon. We refer to this widening as the *multilingual retention gap*.

Our further analysis reveals that, even at full k, the magnitude of the MoE block output relative to the residual stream is larger for low-resource languages than for high-resource ones. We then show that top-k reduction further amplifies this imbalance, because the models renormalize the expert weights over the surviving experts, inflating the contribution of the MoE output relative to the residual. The resulting norm imbalance between the MoE block output and the residual stream drives both the overall performance degradation under reduction and the disproportionate performance collapse for lower-resource languages.

Building on this diagnosis, we propose two complementary inference-time interventions that rescale the MoE block's output before it is added to the residual stream: (i) a grid-searched *global* α, as a simple yet strong baseline that finds an optimal scalar for uniformly rescaling the MoE-layer output for all tokens, and (ii) a parameter-free, per-token *token* α, a token-specific renormalization strategy that stems directly from the routing-level diagnosis. Both are training-free, add no latency, and are orthogonal to the choice of the reduction method. As shown in Figure 1, they substantially improve overall retention (from 47.0% to 90.2% at two retained experts) and approximately halve the multilingual gap for every k (e.g., from 10.0pp to 4.4pp at $k = 3$ retained experts). We further observe that even non-renormalizing models exhibit a small but consistent norm sensitivity at aggressive expert reduction, but in the opposite direction (*undershoot*), which the same scalar correction recovers.

Contributions. ① We present the first systematic multilingual evaluation of MoE models under top-k reduction, covering 100 languages, three massively multilingual MoE models, and five tasks. ② We identify the cause of the resulting multilingual gap: a pre-existing imbalance between the MoE and residual-stream norms is amplified by the renormalization over the surviving experts, with the amplification disproportionately affecting low-resource languages. ③ We propose two training-free and latency-free interventions as a remedy: (i) a grid-searched global (i.e., model-level) scalar—tuned on a small amount of general-purpose text data—for rescaling the MoE outputs and (ii) a parameter-free per-token correction derived from the routing weights. Both interventions substantially close the multilingual gap and improve overall performance retention under top-k reduction. Furthermore, we provide evidence that the proposed interventions are useful beyond MoE architectures evaluated in the main body of the paper.

2 Related Work

Test-time top-k reduction. Due to increasing training and inference costs of large-scale MoEs, a range of inexpensive test-time methods evolved for reducing the number of expert activations in pretrained MoEs *without gradient updates*. These approaches operate at the token level, the layer level, or both. At the token level, NAEE (Lu et al., 2024) performs dynamic expert skipping in Mixtral. The second expert (in top-2 routing) is skipped whenever its gating weight falls below a threshold, calibrated for each layer on a small held-out set. AdapMoE (Zhong et al., 2024) replaces the fixed top-k with a sensitivity-aware gate whose offline-calibrated threshold accounts for the heterogeneous robustness of early versus late layers. At the layer level, LExI (Chitty-Venkata et al., 2025) observes that expert redundancy varies significantly with depth and casts per-layer top-k allocation as a data-free evolutionary search over a Frobenius-norm perturbation proxy, assigning more experts to high-sensitivity layers under a fixed global ac-

tivation budget. Bridging the two levels, Alloc-MoE (Liu et al., 2026) formalizes the number of activations as an expert activation budget and coordinates layer-level and token-level allocation: a dynamic-programming pass over sensitivity profiles (Alloc-L) determines per-layer budgets, and a within-layer redistribution stage (Alloc-T) reallocates activations toward tokens with less concentrated routing distributions. All four approaches represent pure test-time interventions (i.e., they do not require any gradient updates) and require only lightweight calibration, making them directly applicable to any publicly released MoE checkpoint. However, the effectiveness of all four approaches is only evaluated on English benchmarks.

Our work is orthogonal on two axes: we study *multilingual* MoEs and the effect of expert reduction on performance across languages, irrespective of the expert reduction method. We thus resort to a *uniform* test-time top-k reduction and characterize the impact of weight renormalization under expert reduction as a distinct cause of degradation.

Norm Analysis in Language Models. Activation magnitudes in dense language models have been analyzed along several axes, including outlier dimensions within hidden states (Timkey and van Schijndel, 2021; Kovaleva et al., 2021; Dettmers et al., 2022) and outlier tokens that produce massive activations (Sun et al., 2024; Yu et al., 2024). A norm ratio between the layer output and the residual has long been used as an indicator of pretraining instability; widely adopted approaches like batch or layer normalization have been introduced to control the growth of the residual norm with depth (Ioffe and Szegedy, 2015; Wang et al., 2024; Vaswani et al., 2017; Kim et al., 2025). In the MoE landscape, Lo et al. (2025) measure per-expert output norms and relate them to routing scores. They, however, (i) do not consider the aggregated MoE block's contribution to the residual, (ii) carry out the analysis only for English, and (iii) do not propose any inference-time interventions to remedy the inflated norms. No prior work measures the magnitude of the MoE block outputs *relative* to the residual stream, nor establishes connections between a norm imbalance and multilingual performance under expert reduction pressure.

3 Preliminaries

MoE Block. We consider MoE models that replace the dense feed-forward block of a transformer with a sparse mixture of size k over m experts $\{E_i\}_{i=1}^{m}$. At any layer, a token representation $x \in \mathbb{R}^d$ is fed through a learned router $g : \mathbb{R}^d \to \mathbb{R}^m$ followed by a softmax over all m experts that assigns a gate weight $w_i(x)$ to each expert: $w_i(x) = \mathit{softmax}\,(g(x))_i$. Next, for each token, a set of top-k experts with the largest gate weights is selected, $\mathcal{S}_k(x) = \mathrm{top}_k\big(\{w_i(x)\}_{i=1}^{m}\big)$, followed by ℓ_1 renormalization so that the selected gate weights constitute a probability distribution:

$$\tilde{w}_i(x) = \frac{w_i(x)}{\sum_{j\in\mathcal{S}_k(x)} w_j(x)}, \;\; i \in \mathcal{S}_k(x). \tag{1}$$

The output of the MoE block is finally the gate-weighted sum of the outputs of selected experts:

$$\mathrm{MoE}(x) = \sum_{i\in\mathcal{S}_k(x)} \tilde{w}_i(x) \cdot E_i(x). \tag{2}$$

Throughout the paper, k denotes the number of experts the model was trained to activate per token.

We observe that recent MoE releases such as Qwen3 (Yang et al., 2025), Qwen3.5 (Qwen Team, 2026), Gemma4 (Gemma Team, 2026), and GPT-OSS (Agarwal et al., 2025) consistently apply some form of *gate-weight renormalization* over the selected experts, either as an explicit ℓ_1 normalization over the top-k subset as above, or, equivalently, as a softmax applied to the top-k largest gate weights (e.g., in GPT-OSS). We suspect that this design choice is motivated by training stability: with renormalization, the expert weights do not need to absorb scale variation produced by the raw gate scores, since the gate distribution always sums to one over the active set. Other variants exist (e.g., no renormalization, as in Olmoe (Muennighoff et al., 2025) or DeepSeek-v2 (Liu et al., 2024); see Appendix B for more details).

Top-k Reduction. *Top-k reduction* is a parameter-free inference-time intervention that activates only the top-r experts per token, with $r<k$. We replace the top-k set $\mathcal{S}_k(x)$ with $\mathcal{S}_r(x) = \mathrm{top}_r\big(\{w_i(x)\}_{i=1}^{m}\big)$, the r experts with the highest gate weights. Under *standard* top-k reduction, the surviving experts flow through the very same gate-normalization pipeline, with ℓ_1 renormalization now applied over the reduced set, i.e., with r and $\mathcal{S}_r$ instead of k and $\mathcal{S}_k$, respectively, in Eqs. (1) and (2). In what follows, we show that this renormalization over surviving r experts—rather than the reduction in the number of experts itself—is responsible for the negative effects from Figure 1.

4 Analysis

4.1 Experimental Details

We analyze and evaluate three massively multilingual MoE models: Qwen3 (30B-A3B) (Yang et al., 2025), Qwen3.5 (35B-A3B) (Qwen Team, 2026), and Gemma4 (26B-A4B) (Gemma Team, 2026), using the `base` version of each. For all three models, k=8. We evaluate them on a total of 100 languages, across five tasks: Belebele (Bandarkar et al., 2024), GlobalMMLU (Singh et al., 2025), XCOPA (Ponti et al., 2020), XStoryCloze (Lin et al., 2022), and x2en translation on Flores+ (NLLB Team et al., 2024).[1] We perform zero-shot evaluation on all tasks by gradually reducing the number of active experts from the default of $r = k = 8$ down to $r = 2$. Unless otherwise specified, we report average performance across all tasks and languages. We report two primary measures: (1) the absolute drop in average performance w.r.t r=k=8 and the relative retained average performance, i.e., the ratio of performance for a given r divided by the performance with r=k=8.

To analyze trends across resource levels, we cluster the evaluation languages into five tiers, following Joshi et al. (2020), with T5 containing the highest-resource languages and T1 the lowest-resource ones. For cross-tier comparisons, we limit the evaluation to Belebele and Flores+, the only tasks that cover all evaluation languages; we do this to avoid confounding in the tier averages from per-task language availability. We restrict the evaluation of Qwen3 to its 73 supported languages (out of the 100 covered by our evaluation). Qwen3.5 supports all our evaluation languages; Gemma4 is declared to support 200 languages, but we could not find their list: we assume that it supports all of our evaluation languages. We provide full details on the evaluation protocol in Appendix A.

4.2 Multilingual Evaluation

Retention Behavior. Table 1 summarizes the results of our multilingual evaluation. Qwen3.5 shows the best average performance, followed by Qwen3 and Gemma4. The performance ranking reflects the total number of parameters of the evaluated models. Note that Table 1 does not allow for cross-tier comparisons, but only for cross-model and cross-k comparisons. This is because different tasks cover different language sets. For all three models, we observe robust performance across tasks and languages down to r=6 (and even r=5 for Qwen3.5 and Gemma4), retaining more than 95% of the original performance (i.e., performance with r=k=8). As r=6 corresponds to dropping roughly 25% of the sparsely activated experts, this result indicates that MoEs tolerate a moderate reduction of their sparse compute at inference time without major degradation, presumably because the remaining dense computation—attention, residual MLP, and any shared expert—carries the bulk of the signal. However, retaining r=6 does not yield the same active-*parameter* reduction across models, and consequently not the same inference speed-up. The rate of parameter reduction depends not only on the chosen r, but also on the size of shared components that are always active and thus unaffected by the top-k reduction.

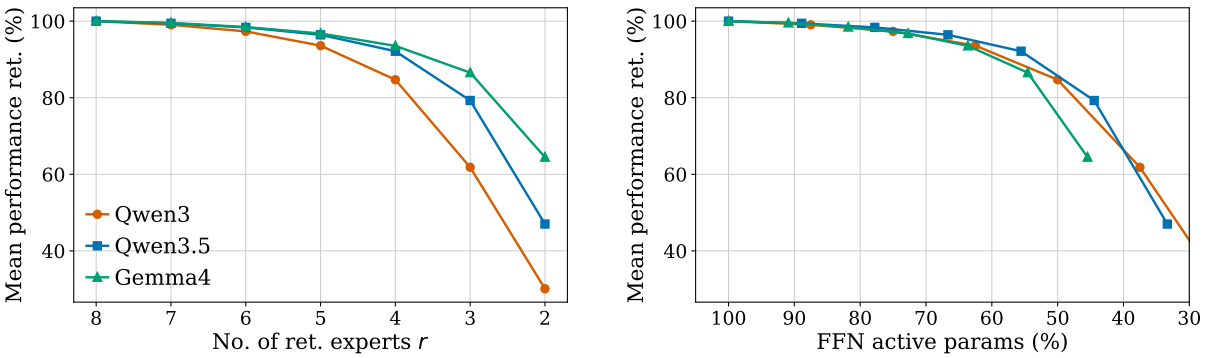
We display the retained performance as (a) a function of k and (b) fraction of the active FFN parameters in Figures 2a and 2b, respectively. All models degrade sharply once fewer than 5 experts are activated. Qwen3.5 and Gemma4 retain moderate performance even at low r, whereas Qwen3 degrades more sharply. We hypothesize that this is because Qwen3 has no shared expert and therefore relies more heavily on the sparsely activated experts. Furthermore, Gemma4 degrades most gracefully as a function of r (Figure 2a), but Qwen3.5 degrades most gracefully as a function of active-parameter fraction (Figure 2b). This is because of Gemma4's large shared expert, which is three times the size of its sparsely activated experts and keeps Gemma4's active-parameter floor high regardless of r. Regarding the inference speedup—the practically relevant axis—Qwen3.5 is the most reduction-friendly architecture, owing to its *small* shared expert: a large shared expert (Gemma4) limits the achievable FLOPs savings because it dominates the active-parameter budget, whereas the absence of a shared expert (Qwen3) collapses the performance for a reduced sparse path (r<k).

Multilingual Retention Gap. In addition to the steep degradation at low k, we observe a widening performance retention gap between high- and low-resource languages as r decreases, which we refer to as the *multilingual retention gap*. We show the multilingual retention gap between the high-resource (T5) and the low-resource languages (T1)

[1] We restrict the translation evaluation to x2en to avoid performance flooring effects associated with the generation of low-resource languages, i.e., that models cannot degrade further. Flooring performance dilutes the retention analysis.

Tier	Qwen3-30B-A3B					Gemma4-26B-A4B					Qwen3.5-35B-A3B				
	r=8	r=6	Ret$_6$	r=5	Ret$_5$	r=8	r=6	Ret$_6$	r=5	Ret$_5$	r=8	r=6	Ret$_6$	r=5	Ret$_5$
T5	78.5	77.4	98.6	75.5	96.1	76.2	75.3	98.9	74.1	97.2	82.9	82.1	99.1	81.0	97.7
T4	78.6	76.7	97.6	73.5	93.5	76.1	74.9	98.4	73.5	96.5	83.8	82.9	99.0	82.0	97.9
T3	78.2	75.2	96.2	71.9	92.0	75.1	73.7	98.1	72.0	95.8	84.1	83.0	98.7	81.6	97.0
T2	66.8	64.5	96.8	60.0	90.1	59.9	58.2	97.3	56.5	94.4	70.8	69.2	97.7	67.6	95.2
T1	72.6	69.3	95.4	64.5	88.8	61.3	59.3	96.9	57.7	94.1	71.6	70.0	97.6	68.6	95.7
Avg	75.8	73.3	96.6	69.7	91.8	68.0	66.4	97.6	64.8	95.2	77.3	76.0	98.2	74.7	96.5

Table 1: Mean task performance at r=k=8 (full), r=6, and r=5, with retentions Ret$_6$, Ret$_5$ (%). Averaged across Belebele, GlobalMMLU, XCOPA, XStoryCloze. Per-language mean across available tasks, then mean across languages within tier. Qwen3 is restricted to its 73 supported languages.



(a) (b)

Figure 2: Mean performance retention across the three MoE models as a function of (a) the number r of retained top-k experts and (b) the fraction of active parameters of the FFN layer (including any shared experts).

as a function of r in Figure 3.[2]. The results closely mirror the overall retention trends from Figure 2: Gemma4 widens the gap the least as a function of r, Qwen3.5 widens the gap the least as a function of active-parameter fraction, and Qwen3—which lacks a shared expert—exhibits the largest multilingual gap on both axes. We omit $r = 2$ for Qwen3 because the overall model performance completely collapses. At near-random performance levels for all languages, none of them has "anything left to lose", so the gap is close to 0. This prevents a meaningful cross-language analysis. We observe a (much) softer version of the same effect for Qwen3.5, where at $r = 2$ the gap somewhat deviates from the overall trajectory observed for $k > 2$ as some languages hit floor performance.

Source of Degradation. To understand the source of the performance degradation, we analyze the retained direction of the MoE block outputs when reducing from r=k=8 to r=4 experts. We compute the average per-token cosine similarity across language tiers between the full top-k and reduced top-r outputs across all layers, and report the results in the top half of Table 2. While Qwen3.5 and Gemma4 exhibit a largely consistent decrease in similarity as we move from higher- to lower-resource language tiers, the same does not hold for Qwen3; yet Qwen3 is the model with the largest multilingual retention gap (see Figure 3). This implies that the multilingual retention gap cannot be solely attributed to a divergence in the direction between the full top-k and reduced top-r outputs.

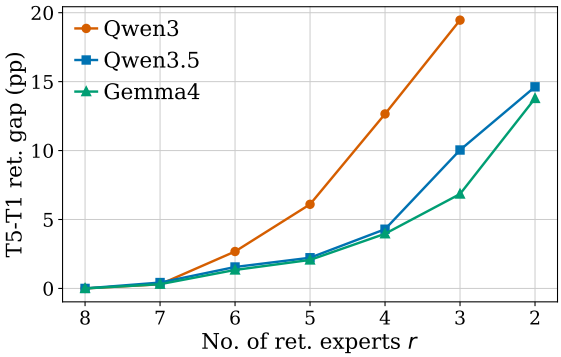


Figure 3: High- to low-resource (T5–T1) multilingual retention gap (pp) across the three MoE models as a function of the number r of retained top-k experts.

We therefore turn to the magnitudes of the signals involved (i.e., L_2 norms of MoE outputs) and report, in the bottom half of Table 2, the

[2]We show the complementary view of multilingual retention gap against the percentage of retained active parameters of the FFN in Appendix D.

Model	T5	T4	T3	T2	T1
Cosine Similarity (r=k=8 vs. r=4)					
Qwen3	0.770	0.759	0.756	0.771	0.772
Qwen3.5	0.809	0.796	0.785	0.767	0.772
Gemma4	0.911	0.907	0.903	0.902	0.904
MoE/Residual Norm Ratio (r=k=8)					
Qwen3	0.131	0.150	0.158	0.159	0.163
Qwen3.5	0.183	0.201	0.207	0.216	0.214
Gemma4	0.064	0.064	0.064	0.068	0.066

Table 2: Evidence of a language-tier-dependent source of degradation. *Top:* mean cosine similarity between full (r=k=8) and reduced (r=4) MoE block outputs. *Bottom:* MoE-to-residual norm ratio at r=k=8.

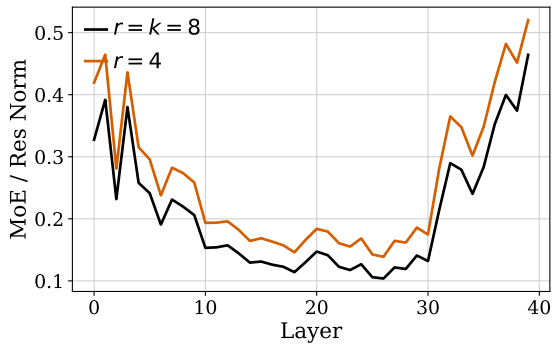


Figure 4: Per-layer MoE-to-residual norm ratio for Qwen3.5, comparing the reference r=k=8 to the renormalized output after reducing to r=4, averaged across 100 languages. Across all layers, top-k reduction overshoots the reference ratio, amplifying the relative MoE contribution to the residual stream.

per-tier MoE-to-residual L_2-norm ratio at k=8—a quantity that, to the best of our knowledge, has not yet been examined in prior magnitude analyses of MoE models (Sun et al., 2024; Lo et al., 2025). The MoE-to-residual norm ratio displays a language-tier-dependent trend: it largely consistently grows from high- to low-resource tiers for all three MoE models: the cause for this trend is the generally smaller residual-stream norms of lower-resource languages. As a consequence, the relative MoE contribution is substantially larger for low(er)-resource languages: from T5 to T1, it increases by roughly 24% for Qwen3, 17% for Qwen3.5, and only 3% for Gemma4—mirroring the order of models by size of their multilingual gaps. We interpret this as follows: low-resource languages have a smaller residual buffer to absorb the directional shift introduced by expert reduction and therefore disproportionately suffer from it.

Finally, we examine how this magnitude imbalance interacts with top-k reduction itself. Figure 4 compares the per-layer MoE-to-residual L_2−norm ratio at the reference r=k=8 to the same ratio after renormalization at r=4 reduction. We observe that, for all three models and the vast majority of layers, the reduced (r=4) MoE-to-residual ratio *overshoots* the reference one (r=k=8). This means that top-k reduction systematically amplifies the MoE component's relative contribution w.r.t the residual stream (we provide the corresponding plots for Qwen3 and Gemma4 in Appendix E). This overshoot is a direct consequence of the gate-weight renormalization in the MoE module (Section 3), which redistributes the full probability mass (after softmax) among the surviving experts, thereby inflating their effective contributions. Prior work on analyzing the design dimensions of MoE models (Lu et al., 2024; Chitty-Venkata et al., 2025) did not deliberately examine this renormalization as an important design dimension. The top-r MoE output is already a noisy approximation of the top-k output (see the top half of Table 2); the norm overshoot (relative to the residual) thus inflates this already noisier signal. The norm overshoot is the connection between the magnitude change and directional damage: it amplifies the directional perturbation introduced by top-k reduction. The amplified noisy signal then propagates through the residual stream. Representations of low-resource languages, which inherently exhibit lower residual norms, are most affected by this amplification; consequently, these languages have the lowest performance retention under inference-time top-k reduction.

5 Intervention

Our analysis suggests that controlling the norms of MoE block outputs is crucial for mitigating performance loss under aggressive test-time top-k reduction. Accordingly, we introduce two interventions that rescale the top-r MoE contribution so as to restore the cosine similarity between the top-k and top-r MoE outputs: (1) *global* α, a naive baseline that grid-searches for a model-level optimal scalar for rescaling, and *token* α, a principled, per-token correction, directly informed by our analysis.

Global α. As a baseline, we rescale all top-r MoE contributions (i.e., model-level, for all tokens and across all layers) with a single global scalar α, found via grid search. The point at which α is applied depends on the exact MoE architecture. In models that add the ℓ_1-weighted expert sum directly to the residual stream or shared expert output (Qwen3 and Qwen3.5), α multiplies the output of

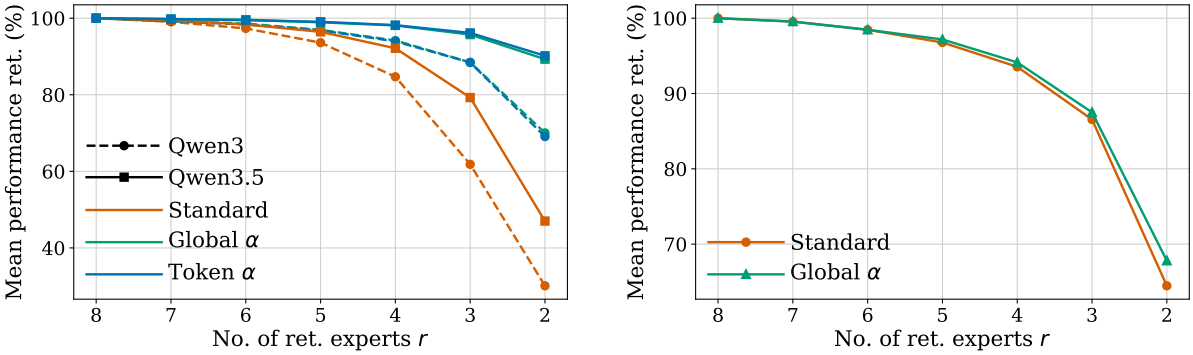


Figure 5: Mean retention as a function of r under standard top-k reduction, the global α, and token α, grouped by post-ℓ_1 normalization scheme: (a) ℓ_1-normalized (Qwen3, Qwen3.5); (b) ℓ_1-normalized + RMSNorm (Gemma4).

the MoE block. In models that apply an additional normalization between the ℓ_1 step and the residual (shared expert) addition—such as Gemma4, which inserts an RMSNorm layer (Zhang and Sennrich, 2019)—the scaling must be applied *after* that normalization, since otherwise the correction is rescaled and washed away. The two variants can be unified as follows:

$$\mathrm{MoE}_r^{\alpha}(x) \;=\; \alpha \cdot \mathrm{Norm}(\mathrm{MoE}_r(x))\,, \tag{3}$$

where $\mathrm{Norm}(\cdot)$ denotes the architecture-specific post-ℓ_1 normalization, i.e., RMSNorm for Gemma4 and an identity function for Qwen3 and Qwen3.5.

We choose α that maximizes the median cosine similarity between the top-k and top-r MoE outputs, after their addition to the residual stream, pooled across all tokens, layers, and validation languages. For the validation set, we randomly take sentences from Flores+, 10 for each of our 100 evaluation languages. We grid-search α with the step 0.05 over $\alpha \in [0.25, 1.0]$.

Token α. Under standard top-k reduction, the gate-weights of surviving experts are divided by the smaller mass $\sum_{j\in\mathcal{S}_r(x)} w_j(x)$ rather than by the training-time mass $\sum_{j\in\mathcal{S}_k(x)} w_j(x)$. This suggests a principled per-token correction: rescale $\mathrm{MoE}_r(x)$ by the ratio of these two masses, directly undoing the token-specific inflation. We set

$$\alpha^{\text{token}}(x) \;=\; \frac{\sum_{j\in\mathcal{S}_r(x)} w_j(x)}{\sum_{j\in\mathcal{S}_k(x)} w_j(x)}, \tag{4}$$

and apply it at the same point as global α. Note that this is equivalent to retaining the training-time ℓ_1 renormalization (over the top-k set) but zeroing out the gate-weight contributions of experts not in $\mathcal{S}_r(x)$. Token α has three practical advantages. First, the top-k sum in the denominator can be computed at virtually *no additional cost* during routing. Second, the resulting rescaling is token-, layer-, and language-adaptive by design: the top-k sum is computed per input token, each layer has its own router (and its own gate-weight sum), and the activated experts vary with the input language. Third, it eliminates the hyperparameter α and the validation step needed to tune it.

	Qwen3		Qwen3.5		Gemma4	
r	global α	token α	global α	token α	global α	token α
7	0.95	0.94	0.95	0.92	1.00	0.93
6	0.85	0.86	0.85	0.84	1.00	0.86
5	0.80	0.78	0.80	0.75	0.95	0.77
4	0.70	0.69	0.70	0.65	0.95	0.67
3	0.60	0.57	0.60	0.54	0.95	0.55
2	0.50	0.44	0.50	0.40	0.90	0.41

Table 3: Global vs. token α for different r. For token α, we report the median of $\alpha^{\text{token}}(x)$.

Overall Retention Results. Figure 5 shows the overall mean retention with our interventions. For Qwen3 (Qwen3.5) both interventions substantially close the retention gap opened by standard top-k reduction, and the gains grow sharply as r decreases. At r=4, standard top-k reduction drops to 84.7% (92.2%), while the interventions recover to 94.2% (98.1%) with global α and 94.0% (98.2%) with token α. At r=2 standard top-k reduction collapses to 30.1% on Qwen3 (47.0% on Qwen3.5); our interventions more than double these numbers: global α recovers to 70.0% (89.3%), and token α to 69.1% (90.2%). Token α tracks global α closely at every r, showing that the more principled per-token correction yields the same benefits, without the additional cost of tuning the rescaling scalar. We provide further evidence for the practical equivalence of token α and global α in Table 3: the median per-token $\alpha^{\text{token}}(x)$ tracks the grid-searched global α closely, and the median cosine similar-

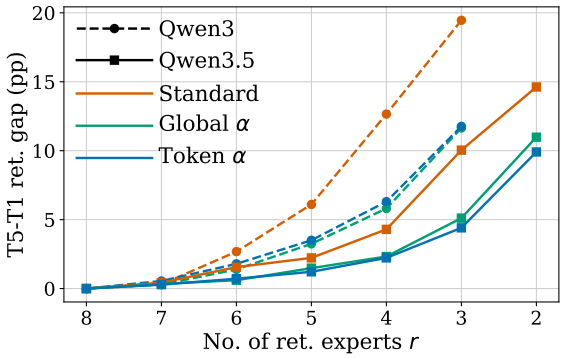


Figure 6: High- to low-resource (T5–T1) multilingual retention gap (pp) under standard top-k reduction, global α, and token α, for Qwen3 and Qwen3.5.

ity between the full top-k and reduced top-r MoE block outputs is nearly identical for both interventions. Surprisingly, despite its flexibility, token α does not yield a performance advantage over the single scalar global α. This questions the suggestions from prior work (Li et al., 2023; Liu et al., 2026) that MoE routing offers an exploitable token-level gate-weight spread under top-k reduction.

For Gemma4, global α follows the same trend as for Qwen3 and Qwen3.5—growing retention improvements over standard top-k reduction with the decrease in r—but these are markedly smaller in magnitude: at $r{=}2$, retention improves from 64.5% to 67.8% ($+3.3$ pp); at $r{=}4$, from 93.5% to 94.1% ($+0.6$ pp). The selected α ranges from 0.90 at $r{=}2$ to 1.00 at $r{\geq}6$. It is because this MoE—due to the integrated RMSNorm—was trained to rely almost exclusively on direction; the small but consistent improvements from our interventions at aggressive reductions confirm the value of magnitude correction. Token α, however, is destructive for Gemma4, because the correction is too aggressive given the architecture's pervasive RMSNorm, which trains the model to operate on near-unit-norm vectors.

Multilingual Gap Results. Beyond overall retention, both interventions substantially reduce the multilingual gap, as shown in Figure 6 for Qwen3 and Qwen3.5; we provide the corresponding plot for Gemma4 in Appendix G. The pattern mirrors that of overall retention: both interventions approximately halve the gap at every r, with the growth of absolute reduction as r decreases. For Qwen3 (Qwen3.5), the gap shrinks from 19.5pp (10.0pp) under standard top-k to 11.7pp (5.1pp) under global α and 11.8pp (4.4pp) under token α at $r{=}3$. For Gemma4, the behavior mirrors that for overall retention: the gain under global α is consistent, but comparably small in magnitude.

Other Multilingual MoEs. We analyze three contemporary MoE families (Qwen3, Qwen3.5, Gemma4) that share an explicit gate-weight renormalization step, the same number of top-k experts ($k{=}8$) and a large total number of experts (128 for Gemma4 and Qwen3, 256 for Qwen3.5), but the diagnosis and the proposed interventions apply to a broader set of MoE models. In Appendix I, we provide evidence that our intervention improves top-k reduction for other ℓ_1-renormalizing architectures (e.g., Mixtral (Jiang et al., 2024) with top-2 routing and 8 experts) as well as for alternative renormalization schemes such as GPT-OSS's softmax-over-top-k (Agarwal et al., 2025). Surprisingly, we show that even non-renormalizing architectures show a norm sensitivity at aggressive top-k reduction. For Olmoe (Muennighoff et al., 2025), we observe the opposite phenomenon for small r: the MoE contribution to the residual stream *undershoots* the training-time magnitude rather than overshooting it, and global α with $\alpha{>}1$ yields consistent improvements. The size of these gains is comparable to Gemma4's: small but consistent across r, confirming that the underlying mechanism—a mismatch between training-time and test-time MoE magnitudes—generalizes beyond the renormalizing MoE architectures.

6 Conclusion

This work offers the first systematic study of test-time top-k reduction in multilingual MoE language models, evaluating three MoEs across 100 languages and 5 tasks. We first reveal that top-k reduction induces a structural *multilingual gap*: low-resource languages collapse earlier than high-resource ones. We then trace this gap to a renormalization-induced norm overshoot inside the MoE block: the increased norms of surviving experts inflate the MoE contribution to the residual stream; this disproportionately affects low-resource languages with inherently smaller residual norms. Building on this diagnosis, we introduce two training- and latency-free interventions: a grid-searched global scalar (global α) and a per-token correction (token α) derived directly from the routing weights. Both substantially recover performance and reduce the multilingual gap. Finally, we provide further evidence for the effectiveness of the proposed interventions, demonstrating their utility for a broad range of MoE architectures.

7 Limitations

Most of our analysis relies on the relative retained downstream performance under top-k reduction compared to the performance at the full top-k, evaluated across a broad range of languages with varying resource levels. Such an analysis inherently requires models and tasks that produce a meaningful performance gradient across languages, particularly for low-resource ones: a model that reaches near-random performance on a language cannot degrade further, which dilutes the analysis. Both the models and the tasks that meet this requirement are limited.

On the model side, the MoE variants we evaluate already represent the strongest multilingual capabilities available to date, with Qwen3—a model with a stronger multilingual focus than comparable contenders such as GPT-OSS or Phi3.5—being just multilingual enough to expose the required gradient.

On the task side, the multilingual gap analysis requires that the tasks support all languages equally, or that the language set is reduced to the intersection of languages covered by the evaluated tasks. This is a prerequisite for cross-tier comparisons: otherwise, a task with high retained performance that covers only a subset of the language tiers confounds the analysis. We observed exactly this behavior when including XCOPA and XStoryCloze, which both retain gracefully and predominantly cover high-resource languages. Including them in the multilingual gap analysis substantially lifted the high-resource (T5) retention, suggesting both a larger multilingual gap and a stronger intervention effect (i.e., there was less to correct for T5, while T1 remained unchanged). We therefore deliberately omit these two tasks from the gap analysis for the soundness of our conclusions. Appendix H further reports how the gap behaves when additionally including GlobalMMLU and restricting the languages to the intersection of the three tasks resulting in only 39 languages (Belebele, GlobalMMLU, and translation) skewed towards higher resource levels.

References

Sandhini Agarwal, Lama Ahmad, Jason Ai, Sam Altman, Andy Applebaum, Edwin Arbus, Rahul K Arora, Yu Bai, Bowen Baker, Haiming Bao, and 1 others. 2025. gpt-oss-120b & gpt-oss-20b model card. *arXiv preprint arXiv:2508.10925*.

Lucas Bandarkar, Davis Liang, Benjamin Muller, Mikel Artetxe, Satya Narayan Shukla, Donald Husa, Naman Goyal, Abhinandan Krishnan, Luke Zettlemoyer, and Madian Khabsa. 2024. The belebele benchmark: a parallel reading comprehension dataset in 122 language variants. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 749–775, Bangkok, Thailand. Association for Computational Linguistics.

Krishna Teja Chitty-Venkata, Sandeep Madireddy, Murali Emani, and Venkatram Vishwanath. 2025. Lexi: Layer-adaptive active experts for efficient moe model inference. *arXiv preprint arXiv:2509.02753*.

Tim Dettmers, Mike Lewis, Younes Belkada, and Luke Zettlemoyer. 2022. Gpt3. int8 (): 8-bit matrix multiplication for transformers at scale. *Advances in neural information processing systems*, 35:30318–30332.

William Fedus, Barret Zoph, and Noam Shazeer. 2022. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39.

Leo Gao, Jonathan Tow, Baber Abbasi, Stella Biderman, Sid Black, Anthony DiPofi, Charles Foster, Laurence Golding, Jeffrey Hsu, Alain Le Noac'h, Haonan Li, Kyle McDonell, Niklas Muennighoff, Chris Ociepa, Jason Phang, Laria Reynolds, Hailey Schoelkopf, Aviya Skowron, Lintang Sutawika, and 5 others. 2024. The language model evaluation harness.

Gemma Team. 2026. Gemma 4: Open models built from gemini 3 research and technology. Accessed: 2026-05-13.

Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr.

Albert Q Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, and 1 others. 2024. Mixtral of experts. *arXiv preprint arXiv:2401.04088*.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.

Jeonghoon Kim, Byeongchan Lee, Cheonbok Park, Yeontaek Oh, Beomjun Kim, Taehwan Yoo, Seongjin Shin, Dongyoon Han, Jinwoo Shin, and Kang Min Yoo. 2025. Peri-ln: Revisiting normalization layer in the transformer architecture. *arXiv preprint arXiv:2502.02732*.

Olga Kovaleva, Saurabh Kulshreshtha, Anna Rogers, and Anna Rumshisky. 2021. BERT busters: Outlier dimensions that disrupt transformers. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3392–3405, Online. Association for Computational Linguistics.

Dmitry Lepikhin, HyoukJoong Lee, Yuanzhong Xu, Dehao Chen, Orhan Firat, Yanping Huang, Maxim Krikun, Noam Shazeer, and Zhifeng Chen. 2020. Gshard: Scaling giant models with conditional computation and automatic sharding. *arXiv preprint arXiv:2006.16668*.

Jiamin Li, Qiang Su, Yitao Yang, Yimin Jiang, Cong Wang, and Hong Xu. 2023. Adaptive gating in mixture-of-experts based language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 3577–3587, Singapore. Association for Computational Linguistics.

Xi Victoria Lin, Todor Mihaylov, Mikel Artetxe, Tianlu Wang, Shuohui Chen, Daniel Simig, Myle Ott, Naman Goyal, Shruti Bhosale, Jingfei Du, Ramakanth Pasunuru, Sam Shleifer, Punit Singh Koura, Vishrav Chaudhary, Brian O'Horo, Jeff Wang, Luke Zettlemoyer, Zornitsa Kozareva, Mona Diab, and 2 others. 2022. Few-shot learning with multilingual generative language models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9019–9052, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Aixin Liu, Bei Feng, Bin Wang, Bingxuan Wang, Bo Liu, Chenggang Zhao, Chengqi Dengr, Chong Ruan, Damai Dai, Daya Guo, and 1 others. 2024. Deepseek-v2: A strong, economical, and efficient mixture-of-experts language model. *arXiv preprint arXiv:2405.04434*.

Baihui Liu, Kaiyuan Tian, Wei Wang, Zhaoning Zhang, Linbo Qiao, and Dongsheng Li. 2026. Alloc-moe: Budget-aware expert activation allocation for efficient mixture-of-experts inference. *arXiv preprint arXiv:2604.08133*.

Ka Man Lo, Zeyu Huang, Zihan Qiu, Zili Wang, and Jie Fu. 2025. A closer look into mixture-of-experts in large language models. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 4427–4447, Albuquerque, New Mexico. Association for Computational Linguistics.

Xudong Lu, Qi Liu, Yuhui Xu, Aojun Zhou, Siyuan Huang, Bo Zhang, Junchi Yan, and Hongsheng Li. 2024. Not all experts are equal: Efficient expert pruning and skipping for mixture-of-experts large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6159–6172.

Niklas Muennighoff, Luca Soldaini, Dirk Groeneveld, Kyle Lo, Jacob Morrison, Sewon Min, Weijia Shi, Pete Walsh, Oyvind Tafjord, Nathan Lambert, and 1 others. 2025. Olmoe: Open mixture-of-experts language models. In *International Conference on Learning Representations*, volume 2025, pages 62061–62121.

NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, and 20 others. 2024. Scaling neural machine translation to 200 languages. *Nature*, 630(8018):841–846.

Edoardo Maria Ponti, Goran Glavaš, Olga Majewska, Qianchu Liu, Ivan Vulić, and Anna Korhonen. 2020. XCOPA: A multilingual dataset for causal commonsense reasoning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2362–2376, Online. Association for Computational Linguistics.

Qwen Team. 2026. Qwen3.5: Towards native multimodal agents.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. 2017. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*.

Shivalika Singh, Angelika Romanou, Clémentine Fourrier, David Ifeoluwa Adelani, Jian Gang Ngui, Daniel Vila-Suero, Peerat Limkonchotiwat, Kelly Marchisio, Wei Qi Leong, Yosephine Susanto, Raymond Ng, Shayne Longpre, Sebastian Ruder, Wei-Yin Ko, Antoine Bosselut, Alice Oh, Andre Martins, Leshem Choshen, Daphne Ippolito, and 4 others. 2025. Global MMLU: Understanding and addressing cultural and linguistic biases in multilingual evaluation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 18761–18799, Vienna, Austria. Association for Computational Linguistics.

Mingjie Sun, Xinlei Chen, J Zico Kolter, and Zhuang Liu. 2024. Massive activations in large language models. *arXiv preprint arXiv:2402.17762*.

William Timkey and Marten van Schijndel. 2021. All bark and no bite: Rogue dimensions in transformer language models obscure representational quality. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4527–4546, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Hongyu Wang, Shuming Ma, Li Dong, Shaohan Huang, Dongdong Zhang, and Furu Wei. 2024. Deepnet:

Scaling transformers to 1,000 layers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(10):6761–6774.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and 1 others. 2019. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Mengxia Yu, De Wang, Qi Shan, Colorado J Reed, and Alvin Wan. 2024. The super weight in large language models. *arXiv preprint arXiv:2411.07191*.

Biao Zhang and Rico Sennrich. 2019. Root mean square layer normalization. *Advances in neural information processing systems*, 32.

Shuzhang Zhong, Ling Liang, Yuan Wang, Runsheng Wang, Ru Huang, and Meng Li. 2024. Adapmoe: Adaptive sensitivity-based expert gating and management for efficient moe inference. In *Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design*, pages 1–9.

A Experimental Details

Models. We evaluate the `base` version of each model: Qwen3 (Apache 2.0) (Yang et al., 2025), Qwen3.5 (Apache 2.0) (Qwen Team, 2026), and Gemma4 (Apache 2.0) (Gemma Team, 2026). All models and datasets are loaded through Hugging Face Transformers (Wolf et al., 2019) (Apache 2.0). Non-translation tasks are evaluated through lm-evaluation-harness (Gao et al., 2024) (MIT).

Tasks. We use five multilingual benchmarks under zero-shot prompting. Belebele (Bandarkar et al., 2024) (CC BY-SA 4.0) provides 900 multiple-choice reading-comprehension questions per language. GlobalMMLU (Singh et al., 2025) (Apache 2.0) provides 14 000 multi-subject multiple-choice questions per language. XCOPA (Ponti et al., 2020) (CC BY 4.0) provides 500 causal-commonsense items per language. XStoryCloze (Lin et al., 2022) (CC BY-SA 4.0) provides 1 511 items per language. Flores+ (NLLB Team et al., 2024) (CC BY-SA 4.0) provides 1 024 parallel sentences per language in its devtest split; we evaluate x2en (translation into English) to avoid floor effects on low-resource targets. We report accuracy on all tasks except translation, where we report chrF.

Translation sampling. To reduce GPU cost, we evaluate Flores+ on a random sample of 200 sentence pairs per language rather than the full 1 024-sentence devtest. We verified that the 200-sentence sample yields chrF scores approximately matching those on the full set; for the purpose of analyzing per-tier gradients, the difference is negligible in favor of the substantial reduction in GPU compute.

Reduction sweep and tiering. We sweep the number of activated experts $r \in \{2, 3, 4, 5, 6, 8\}$ for all models. Multilingual evaluation covers up to 100 languages depending on per-task availability; Qwen3 is restricted to its 73 officially supported languages. For the cross-tier (T5–T1) analyses, we limit to Belebele and translation to avoid confounding from per-task language availability. Tier assignments follow Joshi et al. (2020).

Compute. All evaluations are run on NVIDIA H100 GPUs. We estimate the total compute for the full experimental sweep at ∼4,000 GPU hours.

B MoE Architectural Variants

We summarize how the architectures referenced in Section 3 differ along the gate-weight renormalization axis. Table 4 reports the architectural properties of the three main-body MoE models.

ℓ_1 renormalization over the top-k subset. Mixtral (Jiang et al., 2024) (8 experts, top-2 routing), Qwen3 (Yang et al., 2025), Qwen3.5 (Qwen Team, 2026), and Gemma4 (Gemma Team, 2026) all renormalize the gate weights over the active set. Gemma4 additionally applies an RMSNorm after the ℓ_1 step, making the block output near-magnitude-invariant by construction.

Softmax restricted to the top-k subset. GPT-OSS (Agarwal et al., 2025) computes the softmax only over the top-k logits, which is functionally equivalent to ℓ_1 renormalization of a softmax over all experts.

No renormalization.. Olmoe (Muennighoff et al., 2025) and DeepSeek-v2 (Liu et al., 2024) apply the raw (non-renormalized) gate weights to the selected experts; the gate distribution does not sum to one over the active set. We include DeepSeek-v2 as an example; we do not evaluate it.

C α Grid Search

The selection criterion follows Section 5: for each candidate α, we pool the per-token cosine similari-

	Qwen3	Qwen3.5	Gemma4	Mixtral	DS-V2-Lite	GPT-OSS	Olmoe
Total params	30B	35B	26B	47B	16B	21B	7B
Active params	3B	3B	4B	13B	2.4B	3.6B	1.3B
Vocab size	151 936	248 320	262 144	32 000	102 400	201 088	50 304
Layers	48	40	30	32	27	24	16
# total experts	128	256	128	8	64	32	64
# active experts	8	8	8	2	6	4	8
Shared expert ($\times$ active)	—	1$\times$	3$\times$	—	2$\times$	—	—
Renormalization	ℓ_1	ℓ_1	ℓ_1+RMSNorm	ℓ_1	none	softmax-k	none

Table 4: Architectural properties of the MoE models considered. The shared-expert row reports the total intermediate size of shared experts as a multiple of a single routed expert's intermediate size (— indicates no shared expert). "softmax-k" denotes a softmax restricted to the top-k subset.

ties between the full top-k and the rescaled top-r hidden state *after* the MoE output has been added to the residual stream, across all tokens, all MoE layers, and all 100 validation languages, and take the median of that pool. The selected α is the value that maximizes this pooled median, and is applied uniformly across all layers, tokens, and languages at inference. We adopt the after-residual measurement point because it is the only point that is architecture-agnostic across the ℓ_1-only (Qwen3, Qwen3.5) and ℓ_1+RMSNorm (Gemma4) variants; we verified that, for the ℓ_1-only models, measuring at the MoE block output, after the shared-expert sum, or after the residual yields essentially the same global α. The selected values are reported in Table 3.

D Multilingual Gap (Active-Parameters)

Figure 7 reports the T5–T1 multilingual retention gap as a function of the FFN active-parameter fraction considering shared experts, complementing the gap-vs-k view in Figure 3. Under this axis, Qwen3.5's small shared expert achieves the smallest gap per active parameter.

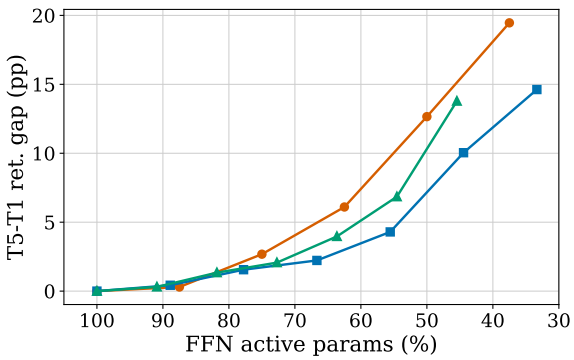


Figure 7: High- to low-resource (T5–T1) multilingual retention gap (pp) as a function of the FFN active-parameter fraction considering shared experts.

E Norm Ratio for Qwen3 and Gemma4

Figures 8 and 9 show the per-layer MoE-to-residual norm ratio for Qwen3 and Gemma4, respectively, complementing the main-body view for Qwen3.5 in Figure 4. Both models exhibit the same overshoot pattern under top-k reduction. We point out that for Gemma4 the overshoot is erased by the RMSNorm that follows the renormalization, explaining the model's weaker sensitivity to renormalization.

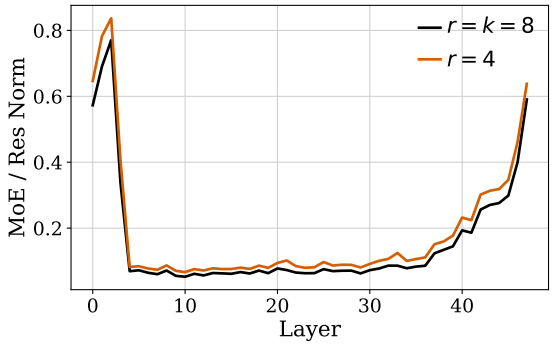


Figure 8: Per-layer MoE-to-residual norm ratio for Qwen3, comparing the reference $r{=}k{=}8$ to the renormalized output after reducing to $r{=}4$, averaged across 100 languages.

F Source of Degradation: Per-Layer Details

We describe the metric aggregation used for Table 2 and Figures 4, 8, and 9. All quantities are computed on the full Flores+ dev set across all 100 languages. For both the cosine similarity (top half of Table 2) and the MoE-to-residual norm ratio (bottom half), the aggregation proceeds in three stages: (i) for each (language, layer) pair, we take the mean over all tokens; (ii) for each language, we take the mean over all MoE layers; (iii) for each resource tier, we take the mean over the languages in that tier. The cosine similarities compare the full

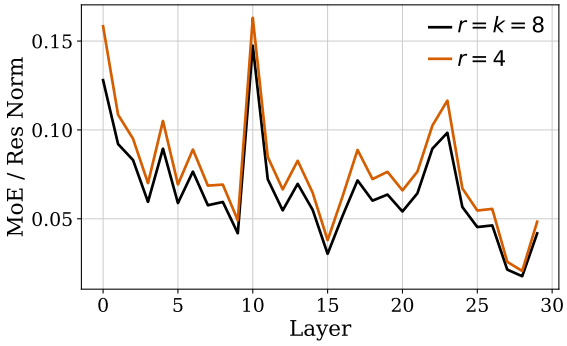


Figure 9: Per-layer MoE-to-residual norm ratio for Gemma4, comparing the reference $r{=}k{=}8$ to the renormalized output after reducing to $r{=}4$, averaged across 100 languages.

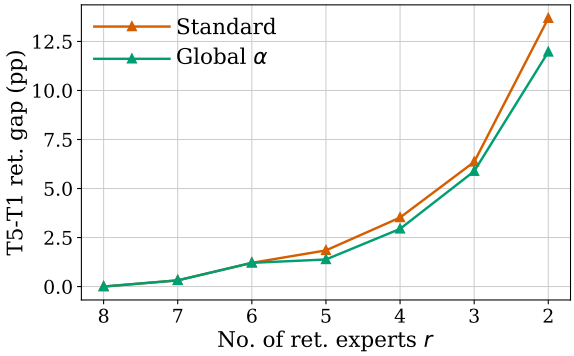


Figure 10: High- to low-resource (T5–T1) multilingual retention gap (pp) as a function of r for Gemma4 under top-k reduction and global α.

top-$k{=}8$ MoE block output against the renormalized top-$k{=}4$ output. The MoE/residual norm ratio is computed at $k{=}8$. The per-layer MoE/residual overshoot figures (Figures 4, 8, and 9) follow the same per-(language, layer) token mean as (i), but in step (iii) we average across all 100 languages rather than per tier.

	Qwen3		Qwen3.5		Gemma4	
r	Ratio	Over.	Ratio	Over.	Ratio	Over.
8 (ref)	0.157	1.000	0.208	1.000	0.065	1.000
7	0.162	1.031	0.217	1.044	0.067	1.027
6	0.167	1.068	0.228	1.096	0.069	1.063
5	0.175	1.116	0.241	1.160	0.073	1.112
4	0.185	1.178	0.258	1.242	0.077	1.183
3	0.198	1.265	0.281	1.352	0.084	1.291
2	0.219	1.397	0.314	1.511	0.097	1.479

Table 5: Per-r MoE-to-residual norm ratio and overshoot factor by model. The *Ratio* column is the mean MoE-to-residual norm ratio under top-k reduction, averaged per (language, layer), then over layers within each language, then over all 100 languages. The *Overshoot* column reports the same quantity normalized by the reference at $r{=}k{=}8$.

G Multilingual Gap for Gemma4

Figure 10 shows the T5–T1 multilingual retention gap on Gemma4 under top-k reduction and the global α intervention, complementing the main-body view for Qwen3 and Qwen3.5 in Figure 6. The gap reduction is smaller than for the ℓ_1-only architectures, consistent with the smaller retention gains observed for Gemma4 (Section 5).

H Robustness of Multilingual Gap

The multilingual gap analysis in the main body uses Belebele and translation because those are the only two tasks that cover all 100 evaluated languages. To check that the gap and its reduction under our interventions are not artifacts of that particular task selection, we additionally include GlobalMMLU—restricting the language pool to the 39 languages supported by all three tasks. We point out that these 39 languages are generally higher resourced. Table 6 reports the T5−T1 retention gap on this 39-language pool. The gap reduction under global α and token α is smaller than on the full 100-language pool, as expected given the higher-resource skew, but the qualitative pattern persists: standard top-k reduction widens the gap as k decreases, and both interventions reduce it at lower k on every model.

	Qwen3			Qwen3.5			Gemma4	
r	Std	Glo α	Tok α	Std	Glo α	Tok α	Std	Glo α
7	2.0	2.0	2.1	0.1	0.3	0.2	0.1	0.1
6	4.1	3.4	3.7	1.0	0.7	0.5	1.2	1.2
5	7.7	4.7	4.9	2.1	1.1	1.0	1.9	0.8
4	13.3	7.5	8.7	4.3	2.1	2.0	3.0	3.0
3	17.6	13.1	13.4	10.4	4.8	4.1	5.5	5.3
2	–	–	–	15.8	10.2	8.9	10.2	9.2

Table 6: High- to low-resource (T5–T1) retention gap (pp) on Belebele, Translation and GlobalMMLU (39 languages). Qwen3 at $r{=}2$ is omitted because top-k reduction collapses for that model.

I Results for Mixtral, GPT-OSS, Olmoe

We provide additional evidence that the diagnosis and the proposed correction generalize beyond the three models analyzed in the main body. We eval-

uate three additional models with different MoE properties (Table 4): Mixtral (Jiang et al., 2024) (explicit ℓ_1 renormalization over the top-2 subset), GPT-OSS (Agarwal et al., 2025) (softmax restricted to the top-4 subset), and Olmoe (Muennighoff et al., 2025) (no renormalization). We limit the evaluation to Belebele English.

Renormalizing Architectures Mixtral and GPT-OSS exhibit the same renormalization-induced norm overshoot as Qwen3.5, Qwen3, and Gemma4 (Tables 7 and 8): the MoE-to-residual norm ratio rises monotonically as k decreases, reaching $1.08\times$ at $k{=}1$ for Mixtral and $1.20\times$ for GPT-OSS. Applying token α retains accuracy better than standard top-k reduction, and the improvement grows with the overshoot: on Mixtral (GPT-OSS),token α lifts retention from 78.9% to 86.4% at $k{=}1$ (from 14.8% to 53.4%).

	Standard		**Token α**		**MoE/Res**	**Over.**
r	Abs	Ret	Abs	Ret		
2	74.9	100.0	-	-	0.167	1.000
1	59.1	78.9	64.7	86.4	0.181	1.079

Table 7: Mixtral on Belebele English across different values of r.

	Standard		**Token α**		**MoE/Res**	**Over.**
r	Abs	Ret	Abs	Ret		
4	95.1	100.0	-	-	0.305	1.000
3	94.4	99.3	94.1	98.9	0.310	1.017
2	89.1	93.7	89.7	94.3	0.320	1.049
1	14.1	14.8	50.8	53.4	0.366	1.200

Table 8: GPT-OSS on Belebele English across different values of r.

	Standard		**Global α**			**MoE/Res**	**Over.**
r	Abs	Ret	α	Abs	Ret		
8	44.7	100.0	-	-	-	0.319	1.000
7	44.8	100.2	1.00	44.8	100.2	0.318	0.997
6	44.2	99.0	1.00	44.2	99.0	0.317	0.992
5	41.9	93.8	1.05	41.6	93.0	0.315	0.985
4	40.2	90.0	1.05	40.3	90.3	0.312	0.978
3	34.9	78.1	1.05	35.6	79.6	0.307	0.960
2	28.3	63.4	1.20	29.1	65.2	0.292	0.915

Table 9: Olmoe on Belebele English across different values of r.

Non-Renormalizing Architecture. Olmoe shows the opposite pattern (Table 9): the MoE-to-residual norm ratio *undershoots* reference, monotonically decreasing as k decreases (down to $0.92\times$ at $r{=}2$). Token α is not applicable here—without renormalization there is no inflation to undo and the ratio of routed masses is by construction ≤ 1, which would compound rather than counteract the undershoot. We therefore evaluate global α with the grid restricted to $\alpha \geq 1$. The gains are small but consistent across k, comparable in magnitude to those observed for Gemma4, consistent with our diagnosis that the renormalization-induced overshoot is the dominant lever and that non-renormalizing architectures are correspondingly less sensitive to magnitude correction.

J Language Coverage

Tier	Language	Qwen3	Qwen3.5	Belebele	Translation	GlobalMMLU	XCOPA	XStory
T5	Arabic	x	x	x	x	x		x
T5	Chinese (S.)	x	x	x	x	x	x	x
T5	English	x	x	x		x		x
T5	French	x	x	x	x	x		
T5	German	x	x	x	x	x		
T5	Japanese	x	x	x	x	x		
T5	Spanish	x	x	x	x			x
T4	Basque	x	x	x	x			x
T4	Catalan	x	x	x	x			
T4	Croatian	x	x	x	x			
T4	Czech	x	x	x	x	x		
T4	Dutch	x	x	x	x	x		
T4	Finnish	x	x	x	x			
T4	Hindi	x	x	x	x	x		x
T4	Hungarian	x	x	x	x			
T4	Italian	x	x	x	x		x	
T4	Korean	x	x	x	x	x		
T4	Persian	x	x	x	x	x		
T4	Polish	x	x	x	x	x		
T4	Portuguese	x	x	x	x	x		
T4	Russian	x	x	x	x	x		x
T4	Serbian	x	x	x	x	x		
T4	Swedish	x	x	x	x	x		
T4	Turkish	x	x	x	x		x	
T4	Vietnamese	x	x	x	x	x	x	
T3	Afrikaans	x	x	x	x			
T3	Bengali	x	x	x	x	x		
T3	Bulgarian	x	x	x	x			
T3	Cebuano	x	x	x	x			
T3	Danish	x	x	x	x			
T3	Estonian	x	x	x	x		x	
T3	Filipino	x	x	x	x	x		
T3	Georgian	x	x	x	x			
T3	Greek	x	x	x	x	x		
T3	Hebrew	x	x	x	x	x		
T3	Indonesian	x	x	x	x	x	x	x
T3	Kazakh	x	x	x	x			
T3	Latvian	x	x	x	x			
T3	Lithuanian	x	x	x	x	x		
T3	Malay	x	x	x	x	x		
T3	Romanian	x	x	x	x	x		
T3	Slovak	x	x	x	x			
T3	Slovenian	x	x	x	x			
T3	Tamil	x	x	x	x		x	
T3	Thai	x	x	x	x		x	
T3	Ukrainian	x	x	x	x	x		
T3	Urdu	x	x	x	x			
T3	Uzbek	x	x	x	x			

Table 10: Per-language coverage of the five evaluation tasks. The “Qwen3” and “Qwen3.5” columns mark each model’s officially supported languages (73 for Qwen3, all 100 for Qwen3.5); the remaining columns mark per-task language availability (Belebele, FLORES Translation, GlobalMMLU, XCOPA, XStoryCloze). Part 1: tiers T5–T3.

Tier	Language	Qwen3	Qwen3.5	Belebele	Translation	GlobalMMLU	XCOPA	XStory
T2	Amharic		x	x	x	x		
T2	Haitian Creole	x	x	x	x		x	
T2	Hausa		x	x	x	x		
T2	Icelandic	x	x	x	x			
T2	Lao	x	x	x	x			
T2	Maltese	x	x	x	x			
T2	Marathi	x	x	x	x			
T2	Punjabi	x	x	x	x			
T2	Swahili	x	x	x	x	x	x	x
T2	Tigrinya		x	x	x			
T2	Tswana		x	x	x			
T2	Wolof		x	x	x			
T2	Xhosa		x	x	x			
T2	Yoruba		x	x	x	x		
T2	Zulu		x	x	x			
T1	Albanian	x	x	x	x			
T1	Armenian	x	x	x	x			
T1	Assamese	x	x	x	x			
T1	Azerbaijani	x	x	x	x			
T1	Bambara		x	x	x			
T1	Burmese	x	x	x	x			x
T1	Chichewa		x	x	x	x		
T1	Guarani		x	x	x		x	
T1	Gujarati	x	x	x	x			
T1	Igbo		x	x	x	x		
T1	Javanese	x	x	x	x			
T1	Kannada	x	x	x	x			
T1	Khmer	x	x	x	x			
T1	Kinyarwanda		x	x	x			
T1	Kurdish		x	x	x			
T1	Kyrgyz		x	x	x	x		
T1	Lingala		x	x	x			
T1	Luganda		x	x	x			
T1	Macedonian	x	x	x	x			
T1	Malagasy		x	x	x	x		
T1	Malayalam	x	x	x	x			
T1	Maori		x	x	x			
T1	Mongolian		x	x	x			
T1	Nepali	x	x	x	x	x		
T1	Norwegian	x	x	x	x			
T1	Oriya	x	x	x	x			
T1	Oromo		x	x	x			
T1	Pashto		x	x	x			
T1	Shona		x	x	x	x		
T1	Sinhala	x	x	x	x	x		
T1	Somali		x	x	x	x		
T1	Southern Sotho		x	x	x			
T1	Sundanese	x	x	x	x			
T1	Swati		x	x	x			
T1	Tajik	x	x	x	x			
T1	Telugu	x	x	x	x	x		x
T1	Tibetan		x	x	x			

Table 11: Per-language coverage of the five evaluation tasks. The “Qwen3” and “Qwen3.5” columns mark each model’s officially supported languages (73 for Qwen3, all 100 for Qwen3.5); the remaining columns mark per-task language availability (Belebele, FLORES Translation, GlobalMMLU, XCOPA, XStoryCloze). Part 2: tiers T2–T1.

K Detailed Results: Retention

	Qwen3																	
	Standard						Global α						Token α					
Tier/Language	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2
T5	99.6	98.9	96.8	91.6	74.5	36.9	99.7	99.5	98.6	97.0	94.1	83.9	99.7	99.5	98.6	97.1	94.1	84.2
English	99.7	99.0	97.0	92.9	78.8	47.0	99.9	99.7	98.9	96.7	94.7	87.8	99.8	99.5	98.8	97.5	94.6	88.9
French	99.5	99.0	97.1	90.4	72.8	29.4	99.5	99.8	98.6	97.9	95.5	83.5	99.3	99.6	98.7	98.1	95.0	84.9
German	99.7	99.4	97.8	92.2	73.8	27.3	100.0	99.9	98.8	97.4	95.1	83.5	100.1	100.1	99.1	97.3	94.4	83.8
Chinese (S.)	99.5	98.1	95.9	92.0	79.0	46.6	99.4	99.0	98.1	96.9	93.9	85.7	99.3	99.3	98.2	96.4	94.2	85.0
Spanish	99.5	99.3	97.8	93.5	78.4	39.3	99.8	99.8	98.8	97.7	94.7	86.5	100.1	99.6	99.0	97.4	94.8	86.5
Japanese	99.4	98.5	96.0	89.6	68.6	28.0	99.7	99.0	98.3	96.5	92.4	80.7	99.7	99.1	98.0	96.7	92.8	80.0
Arabic	99.5	99.0	96.2	90.8	69.9	40.9	99.6	99.2	98.5	95.7	92.7	79.3	99.9	99.2	98.2	96.6	92.7	80.5
T4	99.0	98.1	95.0	87.7	67.1	31.2	99.2	98.9	97.8	95.8	91.6	77.0	99.3	98.8	97.8	95.6	91.9	77.7
Catalan	99.0	98.4	97.1	90.6	69.2	22.3	99.5	99.7	99.3	97.1	94.7	84.4	99.9	99.5	98.7	97.5	94.1	85.1
Portuguese	99.5	98.6	96.2	89.1	69.7	28.6	99.4	99.1	98.3	97.2	94.2	83.7	99.4	99.0	98.1	96.6	94.8	82.7
Italian	99.3	98.4	97.5	94.4	80.7	38.6	99.4	99.4	98.0	97.5	94.8	85.8	99.3	99.4	99.2	96.7	94.4	87.8
Swedish	97.9	97.1	95.0	88.5	66.7	28.4	98.2	97.7	96.7	95.5	91.7	77.3	98.1	97.4	96.3	95.1	91.4	76.3
Croatian	98.7	98.0	94.9	88.3	63.8	20.7	99.3	99.0	97.9	95.6	91.9	77.7	99.4	98.6	98.5	96.1	91.7	79.3
Vietnamese	98.9	97.8	95.5	89.1	70.9	38.1	99.0	98.1	96.7	94.5	91.3	79.3	98.9	97.9	96.7	95.0	91.2	80.2
Czech	98.1	97.1	93.6	87.6	64.9	27.4	98.4	98.1	97.1	94.7	91.0	74.6	98.3	97.9	96.7	94.4	91.7	76.5
Hungarian	99.9	98.8	94.0	84.3	61.3	24.1	100.0	99.8	98.6	96.4	92.5	73.9	99.8	99.4	98.8	96.5	92.4	75.8
Russian	98.3	97.9	96.0	91.7	73.6	38.0	98.3	98.4	97.6	96.1	92.6	81.2	98.5	98.3	97.7	96.4	93.1	82.4
Serbian	99.4	98.4	95.1	85.5	60.9	28.0	99.4	99.1	98.1	95.6	90.2	72.1	99.5	98.7	97.8	95.4	90.6	73.9
Finnish	99.6	99.0	95.9	84.6	61.3	23.3	100.1	99.7	98.5	96.1	91.0	73.4	100.2	100.1	98.8	96.0	91.7	74.1
Dutch	99.9	99.5	96.6	89.3	67.7	28.3	100.2	99.7	98.8	96.6	93.4	80.0	100.2	99.8	98.4	96.8	93.7	81.2
Turkish	99.2	98.2	94.1	88.3	66.7	41.5	99.6	98.5	97.8	96.2	91.7	76.3	99.8	99.1	98.6	95.6	91.3	77.4
Korean	99.2	98.0	95.8	88.1	68.7	30.2	99.5	99.0	98.2	96.2	91.9	78.0	99.4	99.3	98.2	96.3	92.4	78.5
Polish	99.2	97.9	95.7	88.2	67.2	28.3	99.2	99.1	97.8	95.9	90.9	76.0	99.8	98.9	97.7	95.5	92.2	76.6
Persian	97.3	95.8	89.6	76.2	56.2	28.4	97.5	97.1	95.5	93.3	86.3	67.0	97.8	96.9	95.5	92.5	86.8	66.3
Hindi	99.5	98.0	94.7	88.5	69.0	41.7	99.6	99.2	98.3	96.2	90.3	74.0	99.4	98.9	98.3	96.0	90.8	70.9
Basque	99.1	98.4	92.2	86.7	69.1	46.3	99.3	98.8	96.6	93.2	88.5	71.3	99.0	98.5	96.9	92.8	89.1	74.1
T3	98.9	97.2	94.0	85.2	61.5	28.6	99.2	98.7	97.1	94.6	88.9	70.4	99.2	98.4	97.3	94.6	89.1	69.0
Afrikaans	99.1	97.8	95.1	87.6	63.5	23.5	99.6	99.6	97.7	94.9	91.0	75.4	99.5	98.6	97.5	95.7	91.4	73.4
Danish	99.4	98.3	95.3	89.1	65.6	22.8	99.7	98.8	98.1	96.4	93.8	80.0	99.4	98.7	97.8	96.1	94.3	81.0
Bulgarian	99.4	98.0	95.2	89.2	66.6	22.0	99.6	99.9	98.5	97.3	94.1	78.2	99.7	99.7	98.5	96.7	93.7	81.1
Slovak	99.5	97.6	95.7	87.7	64.5	22.9	99.6	99.8	99.3	96.7	93.2	76.9	99.4	99.3	98.6	97.1	94.0	79.0
Romanian	98.0	96.5	94.0	87.6	65.8	26.2	98.1	97.4	96.5	94.5	91.1	75.6	97.9	97.6	96.5	94.6	91.0	75.5
Latvian	99.9	98.6	96.0	83.3	59.8	23.4	100.3	99.7	98.6	95.4	90.3	69.6	99.9	99.6	98.5	96.1	90.2	71.5
Slovenian	98.9	97.8	94.0	86.2	59.6	23.8	99.2	98.5	97.7	95.1	90.1	71.5	99.4	98.5	97.7	95.2	91.0	74.4
Indonesian	99.4	99.0	96.3	89.7	75.9	44.7	99.6	99.2	98.8	96.4	92.6	81.6	99.7	99.3	98.5	96.5	92.5	80.3
Ukrainian	97.6	96.6	93.7	87.0	63.0	26.3	98.0	97.3	96.1	94.4	89.6	74.2	97.5	97.0	95.9	94.4	90.0	62.7
Malay	99.0	98.1	96.3	86.3	66.6	27.7	99.4	99.2	98.5	96.8	90.6	73.6	99.4	98.9	98.1	96.5	91.0	74.2
Filipino	99.8	98.1	95.2	85.8	62.2	28.3	99.9	99.1	98.0	95.5	89.5	69.3	99.9	99.4	97.5	95.4	89.2	68.5
Greek	97.7	95.8	93.1	86.7	62.2	26.0	97.7	96.7	96.1	94.0	89.5	71.3	97.4	96.9	96.0	94.3	90.1	68.9
Cebuano	99.1	95.3	90.3	81.6	58.7	24.4	99.5	97.6	96.6	93.7	84.2	62.3	99.4	98.1	96.6	94.0	85.2	67.2
Hebrew	97.8	95.6	91.1	81.4	58.4	29.6	98.0	97.2	96.0	94.0	87.9	71.5	98.3	96.8	95.9	93.9	88.4	67.1
Thai	99.3	98.0	97.4	90.7	69.2	42.7	100.2	99.2	97.6	96.9	93.5	78.4	100.0	99.1	97.8	97.4	92.7	80.9
Urdu	98.8	96.1	92.3	83.7	54.7	22.7	99.5	99.1	96.4	93.9	85.8	62.6	99.5	98.3	97.5	94.5	86.7	63.1
Lithuanian	98.8	97.4	93.2	82.0	59.4	28.8	99.0	98.4	97.2	94.3	87.9	70.2	99.0	98.1	97.2	93.6	88.0	69.1
Estonian	99.9	98.2	97.5	88.9	68.9	45.0	99.9	99.2	98.2	93.6	88.0	74.9	99.6	98.9	98.3	93.4	89.1	75.9
Uzbek	99.0	96.5	92.3	80.0	52.9	26.2	99.8	99.2	94.1	91.7	85.4	61.0	100.2	98.5	96.7	91.6	84.4	42.5
Bengali	99.1	96.7	92.3	80.8	56.9	29.0	98.8	98.2	96.7	93.7	85.4	62.5	99.3	98.1	96.2	92.4	85.0	59.4
Tamil	99.1	96.2	91.0	82.0	62.6	44.0	98.7	97.3	95.6	93.3	86.9	67.7	99.6	98.2	96.5	93.5	87.0	61.8
Kazakh	97.4	96.4	91.4	80.8	51.1	23.3	98.4	98.9	94.9	90.7	83.1	60.5	98.6	98.4	95.9	90.2	85.4	59.4
Georgian	99.7	97.0	92.8	81.8	47.3	24.3	99.7	99.6	96.1	93.4	81.9	50.2	100.0	98.1	96.9	93.4	79.4	49.3

Table 12: Retention (%) at r=7..2 for Qwen3 under top-k reduction, global α, and token α. Averaged across Belebele, GlobalMMLU, XCOPA, XStoryCloze, Translation. Per-language mean across available tasks, then mean across languages within tier. Qwen3 is restricted to its 73 supported languages. Part 1: tiers T5–T3.

	Qwen3																	
	Standard						Global α						Token α					
Tier/Language	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2
T2	99.5	97.2	92.0	81.0	57.0	34.4	99.1	98.2	96.0	91.3	83.7	62.8	99.3	97.8	95.0	90.4	82.9	62.1
Marathi	98.8	95.8	91.5	80.8	52.8	25.7	99.7	98.0	96.5	91.9	84.9	61.5	99.1	97.4	95.2	91.1	83.8	60.8
Punjabi	99.1	96.9	90.5	75.6	51.2	23.9	99.3	98.2	95.1	92.1	85.9	60.2	99.8	97.7	94.0	89.9	83.6	55.4
Icelandic	99.2	97.0	92.9	82.0	54.4	27.1	97.8	98.7	96.4	91.7	82.9	61.8	98.0	98.1	94.9	90.5	81.4	65.6
Maltese	100.1	96.7	91.1	79.4	54.2	27.3	99.5	97.9	95.3	90.4	80.0	59.3	100.4	97.7	94.2	88.4	80.4	55.4
Lao	99.6	97.9	90.4	78.6	45.1	21.3	99.1	99.3	95.1	89.4	81.8	54.1	99.3	98.2	95.2	89.7	82.6	55.3
Haitian Creole	100.2	97.3	93.0	83.5	68.1	52.0	99.3	97.3	97.4	91.2	83.0	70.5	99.2	97.2	96.2	91.0	82.4	70.4
Swahili	99.2	98.7	94.8	86.9	73.3	63.5	98.8	98.0	96.5	92.3	87.2	72.3	99.4	98.1	95.7	92.3	86.2	71.8
T1	98.9	96.2	91.1	79.8	53.9	26.6	99.1	98.1	95.7	92.0	84.4	60.1	98.9	97.6	95.3	91.6	84.1	57.5
Norwegian	99.5	97.9	96.0	88.6	62.2	22.8	99.3	98.6	98.7	96.8	93.8	80.6	99.1	98.9	98.6	97.5	94.0	82.6
Macedonian	99.8	97.7	93.4	84.8	61.2	22.4	99.7	99.3	97.2	96.2	91.0	74.2	100.0	99.0	97.5	95.8	91.6	74.8
Albanian	100.0	98.3	93.9	84.3	58.8	23.8	100.5	99.4	97.0	94.0	88.3	72.2	99.5	98.6	97.4	94.0	88.5	61.0
Armenian	99.0	97.2	91.7	78.5	51.4	18.9	98.2	98.4	98.0	94.0	84.8	59.3	98.7	98.4	97.1	92.7	85.1	55.5
Malayalam	98.1	96.0	89.8	77.2	51.6	22.5	99.0	98.7	95.0	90.1	84.0	56.3	98.8	98.5	94.9	90.9	85.5	54.9
Gujarati	98.3	95.7	91.9	79.7	50.5	21.0	98.8	97.8	94.7	92.0	87.1	58.4	98.7	97.5	94.7	91.0	86.7	47.9
Javanese	100.0	97.8	92.6	78.8	57.2	26.0	100.1	98.3	97.2	94.1	86.3	64.5	100.3	98.0	96.4	94.1	84.5	60.6
Kannada	99.9	96.1	89.5	76.9	51.4	21.2	100.0	98.3	95.7	91.0	85.4	51.7	100.0	97.8	94.2	89.4	84.7	54.5
Nepali	96.8	93.9	89.4	80.2	56.2	31.1	97.3	96.3	94.2	90.6	81.8	58.7	97.4	95.7	94.2	89.5	81.3	51.9
Oriya	99.5	95.2	88.9	74.6	47.5	22.3	99.2	97.8	93.9	88.1	79.3	50.9	99.0	97.1	91.8	86.5	80.0	52.3
Azerbaijani	98.8	97.1	92.4	79.9	56.5	24.7	99.5	99.2	95.9	93.8	86.2	60.8	99.8	98.4	96.4	93.4	85.8	57.7
Sundanese	99.5	96.2	92.1	79.8	57.3	25.6	99.7	97.8	95.4	90.8	82.5	58.5	98.7	97.7	95.3	91.0	80.4	59.0
Telugu	98.3	96.8	91.8	82.0	61.4	42.8	98.6	97.6	96.5	92.7	86.0	57.7	98.2	97.1	95.9	91.0	86.0	56.5
Assamese	98.9	96.6	89.4	75.9	50.7	26.1	99.9	98.1	96.2	89.5	80.3	52.0	99.0	97.6	94.6	90.2	78.9	49.5
Khmer	98.3	92.5	87.5	79.8	39.4	20.3	98.5	99.6	96.7	92.6	79.3	53.9	98.9	97.6	96.6	94.2	83.8	53.5
Tajik	98.3	96.9	91.4	77.1	46.1	27.6	98.4	96.3	91.9	87.0	77.5	51.1	97.8	95.4	91.1	85.2	74.3	44.7
Sinhala	97.6	94.0	87.4	75.2	52.7	35.6	97.4	95.2	92.2	87.2	78.5	53.5	97.6	95.0	92.1	87.3	77.3	51.8
Burmese	99.4	95.7	91.6	82.6	58.3	44.1	100.6	98.8	96.1	95.5	87.8	67.1	99.0	98.3	96.4	94.1	85.6	65.9
Avg	99.1	97.3	93.6	84.7	61.8	30.1	99.2	98.6	97.0	94.2	88.5	70.0	99.2	98.3	96.8	93.9	88.4	69.1

Table 13: Retention (%) at r=7..2 for Qwen3 under top-k reduction, global α, and token α. Averaged across Belebele, GlobalMMLU, XCOPA, XStoryCloze, Translation. Per-language mean across available tasks, then mean across languages within tier. Qwen3 is restricted to its 73 supported languages. Part 2: tiers T2–T1 and overall mean.

	Qwen3.5																	
	Standard						Global α						Token α					
Tier/Language	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2
T5	99.7	99.3	97.7	94.8	86.6	62.4	100.0	99.9	99.7	99.1	97.9	94.5	99.9	99.9	99.6	99.1	98.1	94.6
English	99.7	99.2	98.3	96.9	91.7	75.9	99.8	99.8	99.7	99.4	98.6	96.5	100.1	99.9	99.6	99.3	98.6	96.6
German	99.7	99.1	97.6	94.2	84.7	58.2	99.9	99.8	99.6	99.0	97.5	94.3	99.7	99.8	99.4	99.1	97.8	94.7
French	99.4	99.2	98.3	96.3	87.3	59.6	99.6	99.7	99.7	99.4	98.7	95.8	99.9	99.7	99.9	99.5	98.7	95.6
Chinese (S.)	99.7	99.7	98.3	94.6	87.4	68.3	100.2	99.8	99.6	98.8	97.5	94.8	99.9	99.8	99.6	98.9	98.4	95.5
Spanish	99.9	99.5	98.4	95.6	87.7	62.2	100.0	100.1	99.9	99.2	98.2	95.5	99.8	100.2	100.0	98.8	97.9	95.2
Arabic	99.5	98.7	96.5	93.2	83.8	56.6	100.0	99.5	99.1	98.3	97.2	92.5	99.7	99.1	98.6	98.1	97.2	92.0
Japanese	99.7	99.7	96.7	92.9	83.5	56.2	100.4	100.5	100.4	99.5	97.9	92.1	100.1	100.8	100.2	99.6	98.5	92.8
T4	99.5	98.8	97.4	94.5	84.2	54.1	99.8	99.8	99.5	99.0	97.7	93.7	99.9	99.7	99.4	98.9	97.7	93.8
Catalan	99.5	100.0	98.7	96.7	86.4	51.9	99.7	100.1	99.4	99.7	97.9	94.2	99.9	99.7	99.7	99.3	98.8	96.2
Portuguese	99.4	98.2	96.3	92.4	80.7	54.9	99.8	99.8	99.2	98.2	97.9	94.8	99.9	99.9	99.6	99.0	98.4	95.1
Swedish	99.5	98.5	97.6	94.5	87.6	54.1	99.6	99.6	99.3	98.8	97.6	92.7	99.5	99.3	99.1	98.8	97.0	92.5
Italian	99.6	99.7	99.5	96.6	89.0	65.0	100.1	100.5	100.2	99.8	98.1	94.6	99.7	100.2	100.0	99.8	97.8	94.7
Serbian	99.7	98.8	97.9	95.7	81.6	48.1	99.6	99.9	99.7	98.5	97.0	92.0	99.7	99.7	99.3	98.0	97.0	91.3
Czech	99.3	98.1	97.3	94.6	85.2	54.1	99.6	99.5	99.1	98.7	97.7	93.4	100.1	99.2	99.3	98.8	97.6	92.5
Vietnamese	99.9	99.4	98.5	95.2	84.1	61.7	100.2	99.9	99.2	99.0	97.2	94.3	100.2	100.0	99.4	98.7	97.7	93.9
Russian	99.9	99.4	97.8	95.6	85.6	61.0	100.1	100.1	99.7	98.5	97.3	92.7	100.2	99.8	99.5	98.7	96.6	93.1
Hungarian	99.9	98.7	94.7	91.8	80.5	47.0	99.9	99.8	99.8	99.4	98.9	95.1	100.0	99.8	100.1	100.2	98.7	95.6
Croatian	99.2	97.1	96.9	94.2	86.2	48.2	100.2	101.3	99.8	101.0	101.1	98.8	100.7	100.9	100.0	100.7	101.1	98.7
Finnish	98.9	96.6	94.3	92.0	82.8	46.0	99.5	99.0	100.2	100.3	98.5	93.8	99.1	98.6	98.9	98.8	99.4	94.9
Dutch	100.2	99.2	98.2	94.7	83.8	56.7	100.1	99.7	99.6	99.2	97.5	94.5	99.9	99.6	99.7	99.2	97.8	94.5
Turkish	99.3	98.5	98.4	94.3	83.9	58.8	99.7	99.6	99.1	98.6	98.0	92.5	99.6	99.4	99.4	97.9	96.1	92.0
Persian	99.0	98.8	97.3	94.3	84.4	45.7	99.5	99.4	98.4	97.8	96.2	92.8	99.5	99.0	98.8	98.0	96.6	92.5
Korean	99.9	99.5	97.7	94.0	82.2	55.5	100.1	99.9	99.6	99.3	97.6	93.2	100.0	99.6	99.6	99.0	97.8	93.3
Polish	99.7	98.8	97.5	95.7	84.0	49.4	99.8	99.7	99.2	98.7	97.0	92.9	99.7	99.5	99.1	98.6	97.0	92.4
Hindi	99.4	98.8	97.4	94.3	83.8	56.9	99.6	99.2	99.0	98.3	96.3	92.1	99.9	99.4	99.0	98.4	96.4	92.3
Basque	99.7	99.3	97.2	94.6	83.4	58.5	100.2	100.1	100.1	98.2	96.7	91.4	100.3	100.5	99.4	98.7	97.4	92.6
T3	99.5	98.6	96.9	93.6	82.5	48.2	99.8	99.6	99.2	99.0	97.5	92.8	99.8	99.7	99.4	98.8	97.5	93.1
Afrikaans	99.9	99.0	97.7	95.3	86.4	43.2	100.0	100.1	99.4	99.5	99.1	95.1	99.9	100.0	99.4	99.4	98.8	95.7
Danish	99.2	98.3	96.0	93.1	84.1	55.3	99.4	99.3	99.1	99.7	98.5	96.1	99.6	99.3	98.8	98.8	98.0	96.4
Romanian	99.8	98.2	96.0	92.4	84.7	52.2	99.7	99.5	99.1	99.2	97.4	93.4	99.9	99.6	99.5	98.7	97.5	92.8
Bulgarian	100.1	99.4	98.0	96.6	86.8	52.5	99.9	100.1	100.6	100.7	99.6	97.5	99.9	100.3	100.3	100.7	100.0	96.3
Malay	99.7	99.2	98.2	96.1	86.8	54.4	100.0	99.5	99.4	98.8	96.8	92.8	99.9	99.6	99.5	98.8	97.1	93.4
Slovak	99.6	97.3	97.2	94.3	83.0	47.6	99.9	99.0	98.9	99.3	98.9	95.0	99.9	99.6	98.9	98.2	98.6	95.6
Indonesian	99.6	99.0	98.3	96.6	89.2	68.1	100.0	99.5	99.1	98.8	97.2	94.6	99.8	99.7	99.1	98.7	97.2	93.7
Latvian	99.1	98.3	97.2	91.6	80.6	41.6	99.7	99.7	99.2	99.0	98.4	93.4	99.5	99.6	99.2	98.7	97.9	93.7
Ukrainian	99.4	98.1	97.2	92.1	83.8	51.9	100.0	99.6	99.3	98.6	96.9	91.8	99.7	99.4	100.1	99.0	97.3	92.2
Filipino	99.7	98.8	97.7	93.1	81.7	46.6	99.7	99.4	99.1	98.0	95.6	90.5	99.9	99.5	98.9	97.8	96.3	91.4
Slovenian	98.9	97.7	96.5	94.6	81.9	42.1	99.1	98.7	98.5	99.2	98.2	94.2	99.6	99.4	98.7	99.8	98.0	93.5
Greek	99.1	98.2	95.1	92.3	82.8	47.1	99.7	100.0	99.2	99.0	96.9	92.4	99.5	99.4	99.4	99.0	97.4	93.0
Hebrew	100.3	98.9	97.5	94.0	83.6	49.3	100.2	100.2	100.0	99.2	97.6	92.8	100.2	100.0	100.0	99.1	97.2	92.0
Estonian	99.6	98.9	97.1	93.0	83.4	48.8	99.6	99.7	99.3	99.0	98.9	91.5	99.9	99.1	99.9	98.8	97.9	91.9
Lithuanian	99.1	98.7	98.6	95.1	83.2	46.3	99.8	100.4	100.1	99.4	97.4	92.1	99.8	100.2	99.8	98.9	96.8	90.8
Uzbek	100.0	99.1	98.2	92.6	75.9	42.8	100.2	100.2	99.6	99.2	98.4	91.3	100.0	99.9	99.4	99.3	96.9	91.5
Cebuano	100.4	98.7	97.4	93.4	81.1	45.3	100.9	99.5	99.2	98.7	95.8	90.6	100.4	99.5	99.6	98.2	96.5	91.8
Urdu	99.2	98.1	96.0	91.6	72.8	36.8	99.8	99.3	98.6	98.0	97.1	91.8	99.8	99.6	99.5	99.1	97.5	91.3
Bengali	99.5	98.6	96.7	93.0	80.3	46.3	99.7	99.4	98.5	97.8	95.8	89.4	100.2	99.3	98.7	98.0	95.9	90.9
Georgian	99.2	98.8	95.9	91.8	77.9	35.2	99.7	99.1	98.6	99.1	97.1	90.4	99.7	99.8	99.6	99.2	97.1	92.9
Kazakh	99.0	97.9	95.6	93.0	76.4	38.3	99.9	99.3	98.9	98.7	96.9	91.3	99.8	99.9	99.3	98.7	97.2	92.6
Thai	100.1	98.9	93.2	93.0	88.5	62.5	99.4	99.5	99.0	99.0	97.8	95.0	100.0	100.3	99.6	98.2	97.7	95.7
Tamil	99.0	98.9	96.9	93.1	82.9	55.0	99.1	99.6	99.3	98.8	96.8	90.6	99.3	99.3	100.0	98.6	96.9	91.6

Table 14: Retention (%) at r=7..2 for Qwen3.5 under top-k reduction, global α, and token α. Averaged across Belebele, GlobalMMLU, XCOPA, XStoryCloze, Translation. Per-language mean across available tasks, then mean across languages within tier. Part 1: tiers T5–T3.

	Qwen3.5																	
	Standard						Global α						Token α					
Tier/Language	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2
T2	99.6	98.3	95.9	90.5	75.8	42.1	99.7	99.5	98.6	97.4	94.2	85.6	99.7	99.3	98.8	97.7	95.0	87.4
Maltese	100.0	98.7	98.0	94.8	81.0	39.8	99.6	100.5	99.6	98.9	97.4	90.6	99.7	99.9	99.7	99.0	97.5	92.8
Icelandic	100.3	98.8	98.8	94.2	82.3	40.1	101.1	100.5	100.3	99.9	97.7	92.3	100.5	100.3	100.9	100.0	98.5	94.4
Marathi	99.7	98.7	97.8	94.2	81.4	41.8	100.6	100.0	100.3	99.5	97.1	92.1	100.3	99.9	100.0	99.0	96.7	92.3
Punjabi	100.7	100.0	98.4	94.7	83.6	37.0	100.4	100.4	99.9	99.9	98.4	92.5	100.6	99.8	100.1	99.9	98.4	93.1
Haitian Creole	99.1	97.8	95.9	93.3	81.2	54.3	99.0	100.8	100.0	99.0	97.0	89.4	99.6	99.8	99.4	98.5	96.6	90.2
Swahili	99.4	98.5	96.5	93.4	84.6	59.8	99.9	98.5	98.1	96.8	94.6	88.0	99.7	98.8	98.3	96.3	94.8	89.3
Lao	98.7	98.5	96.5	89.3	68.9	33.0	98.5	99.6	99.4	98.5	97.2	88.8	99.4	99.4	99.4	98.7	96.7	89.6
Xhosa	99.0	98.0	94.8	86.9	67.8	34.0	100.1	99.1	99.8	96.9	93.0	83.9	99.8	99.0	99.1	97.0	94.0	86.3
Zulu	99.5	98.5	94.1	91.2	76.4	42.5	99.8	99.1	98.1	97.6	95.7	87.0	98.8	99.8	98.5	98.3	95.1	88.8
Amharic	99.6	98.2	93.6	86.6	70.9	37.6	100.0	99.1	97.9	97.4	92.1	81.1	99.9	98.9	98.1	97.4	93.6	83.9
Hausa	99.7	98.4	95.9	88.2	73.0	45.4	99.7	98.8	98.2	96.5	90.7	80.5	99.6	98.8	98.3	95.9	93.5	84.4
Tigrinya	99.5	97.2	94.5	88.3	69.0	35.1	99.5	98.1	95.8	95.5	92.1	79.6	99.4	98.6	97.0	97.3	92.8	81.9
Tswana	99.9	98.4	93.7	86.3	68.1	38.2	99.1	99.4	97.3	96.9	90.3	81.8	99.3	98.3	98.2	97.8	94.1	84.7
Yoruba	99.5	97.5	94.6	88.1	71.2	47.1	100.5	100.5	97.3	95.0	92.1	80.6	99.8	99.4	98.6	95.5	92.3	82.1
Wolof	99.9	96.7	95.0	88.2	78.0	45.6	97.7	98.2	97.1	92.2	88.1	76.5	99.3	99.2	96.2	94.3	90.8	77.2
T1	99.2	97.9	95.7	90.3	74.9	41.9	99.7	99.4	98.4	97.3	93.8	85.5	99.7	99.3	98.6	97.5	94.7	86.9
Norwegian	99.1	98.4	96.8	93.1	83.5	48.7	99.8	99.7	99.8	99.7	99.1	95.7	99.7	99.7	99.7	99.2	98.6	96.4
Macedonian	99.6	98.3	95.4	95.0	85.1	45.9	99.8	99.8	100.1	99.7	98.5	95.1	100.2	99.6	99.3	99.0	99.0	95.9
Albanian	99.4	97.2	96.9	92.7	80.2	42.2	99.8	100.3	99.6	99.1	98.2	93.4	99.9	99.2	99.4	99.3	97.9	94.0
Armenian	99.3	98.3	97.1	94.1	78.8	42.8	100.3	100.1	99.8	99.2	98.1	92.9	100.6	100.7	100.4	99.7	98.5	94.0
Malayalam	99.4	98.9	97.4	93.5	79.8	34.1	100.1	99.7	99.9	99.9	97.7	91.6	99.7	100.2	99.6	99.9	97.8	93.8
Javanese	99.7	98.7	98.0	92.7	75.0	40.5	99.6	100.2	99.1	98.3	96.8	89.5	99.5	100.1	99.3	99.4	97.1	91.7
Gujarati	99.5	98.7	97.1	92.4	79.6	35.1	100.6	100.1	99.7	98.8	96.0	89.0	100.1	99.2	99.2	98.4	96.3	89.6
Oriya	100.0	99.5	98.3	94.3	75.9	38.9	100.6	101.1	99.7	99.4	97.0	90.0	99.9	100.4	100.1	99.7	97.3	91.0
Sundanese	99.3	97.9	96.0	89.8	76.5	36.4	99.0	99.2	98.9	97.1	95.2	90.5	99.3	99.4	99.0	97.9	95.6	90.9
Kannada	99.6	98.9	98.1	92.2	65.8	31.6	99.7	99.8	99.9	98.6	97.1	90.3	100.1	100.1	99.9	98.8	97.1	92.0
Nepali	100.1	99.5	97.1	92.9	78.2	43.6	100.2	100.2	100.1	98.9	96.1	89.5	100.1	100.6	100.1	99.1	96.4	88.8
Tajik	101.1	99.0	96.7	93.3	80.2	38.4	100.7	100.7	100.2	98.6	96.4	88.4	100.5	100.3	100.0	99.0	96.4	90.8
Telugu	99.2	98.5	97.3	92.4	79.5	51.5	99.8	99.4	98.8	97.5	95.1	89.0	99.5	99.4	99.1	97.8	95.0	90.1
Sinhala	99.0	98.2	95.2	90.4	77.1	39.1	99.0	99.0	98.3	97.4	94.7	87.8	99.7	99.2	99.1	98.4	95.0	89.7
Kurdish	99.0	98.6	94.9	89.4	67.7	32.4	99.0	99.5	99.1	97.3	92.6	85.0	100.2	99.1	98.3	97.1	94.9	84.6
Assamese	99.1	99.2	96.4	93.3	76.2	37.6	100.0	99.4	98.9	97.0	95.1	87.5	99.6	99.4	99.4	97.7	95.7	88.8
Khmer	99.3	97.9	96.5	93.4	70.7	35.6	100.1	100.1	99.6	98.9	97.4	90.4	100.6	100.3	99.9	99.5	98.9	93.0
Azerbaijani	98.8	96.9	96.4	94.4	80.5	42.0	99.2	99.7	98.6	98.4	95.3	88.2	99.0	99.5	98.8	98.0	96.5	89.0
Kyrgyz	99.7	98.3	95.0	90.0	78.0	41.8	99.5	99.2	98.1	96.6	94.7	88.3	99.3	99.0	98.3	96.5	95.1	87.2
Pashto	99.7	98.0	96.1	89.5	72.6	36.5	99.9	99.7	99.0	98.0	94.8	86.1	100.3	99.4	99.8	98.9	94.4	86.3
Mongolian	99.2	97.9	96.1	89.5	65.3	32.3	99.9	100.4	98.6	99.7	96.2	87.9	99.6	99.8	100.0	98.8	96.0	89.7
Burmese	99.4	96.7	96.3	92.5	81.8	53.2	99.9	100.1	99.1	98.7	96.4	90.5	100.3	99.2	99.2	98.5	97.1	91.4
Southern Sotho	98.9	99.2	95.5	88.5	71.8	37.8	99.6	99.2	97.0	97.4	90.7	80.1	99.6	99.0	98.9	98.0	93.5	83.1
Malagasy	98.8	97.7	94.8	89.6	76.0	42.4	99.7	98.7	97.7	96.4	91.8	82.3	99.4	98.9	98.0	96.9	93.9	86.0
Kinyarwanda	99.1	97.0	97.2	92.2	75.6	38.0	100.8	100.0	99.3	98.2	95.8	85.6	100.0	99.2	98.2	97.7	97.2	87.7
Maori	98.5	96.0	91.9	84.3	70.8	40.6	99.8	99.9	97.1	97.2	91.9	82.0	99.5	100.0	98.1	97.5	93.9	84.4
Somali	100.1	98.8	95.1	88.3	70.1	42.8	100.4	100.6	99.7	97.8	92.8	81.6	100.3	100.0	99.2	97.3	94.2	81.7
Shona	98.8	97.2	95.1	89.0	72.9	42.8	99.3	98.5	98.6	96.0	92.1	80.6	99.4	98.4	97.9	95.6	92.9	83.5
Igbo	99.8	97.6	95.3	88.2	74.0	43.9	100.0	98.4	98.1	95.9	91.1	80.3	99.9	99.2	98.1	96.5	92.0	80.9
Chichewa	100.1	98.3	95.1	90.0	70.8	47.3	99.5	98.7	98.0	96.8	90.1	80.3	100.0	99.5	97.0	96.8	92.7	84.2
Swati	97.3	97.8	91.7	82.2	70.1	40.4	98.3	97.8	94.7	92.6	87.7	73.1	98.8	97.7	94.5	93.3	88.3	77.4
Guarani	99.2	97.6	94.2	87.2	74.0	59.7	100.3	98.5	95.7	97.2	91.5	84.1	99.6	97.9	97.2	96.4	93.0	83.4
Lingala	98.5	96.7	94.9	87.4	73.8	45.7	99.2	99.2	96.5	93.5	88.2	78.3	98.7	99.4	97.9	94.1	91.6	79.6
Oromo	99.2	95.5	94.1	86.9	66.6	40.8	98.9	98.9	95.3	91.4	85.6	70.2	98.3	97.4	96.2	93.9	89.1	75.2
Luganda	97.8	98.0	94.6	88.5	74.9	45.9	99.5	97.6	96.8	94.6	87.5	76.0	98.6	98.8	97.8	96.7	90.6	79.9
Tibetan	97.1	94.2	89.4	82.1	67.4	44.2	97.6	99.1	96.3	93.3	87.8	75.2	99.6	96.2	97.0	94.9	89.0	73.9
Bambara	99.0	98.6	95.9	85.4	74.2	57.4	100.8	96.9	97.0	93.8	87.9	77.0	99.0	97.7	95.9	89.8	85.1	76.2
Avg	99.4	98.4	96.4	92.1	79.3	47.0	99.8	99.6	98.9	98.1	95.7	89.3	99.8	99.5	99.1	98.2	96.2	90.2

Table 15: Retention (%) at r=7..2 for Qwen3.5 under top-k reduction, global α, and token α. Averaged across Belebele, GlobalMMLU, XCOPA, XStoryCloze, Translation. Per-language mean across available tasks, then mean across languages within tier. Part 2: tiers T2–T1 and overall mean.

	Gemma4											
	Standard						Global α					
Tier/Language	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2
T5	99.7	99.2	97.9	95.7	90.4	73.2	99.7	99.2	97.9	95.9	90.9	75.2
English	99.8	99.2	98.2	96.2	91.9	78.0	99.8	99.2	98.2	96.6	92.1	79.8
French	99.5	99.1	97.0	95.2	90.9	72.3	99.5	99.1	97.6	95.5	91.0	74.3
Spanish	100.0	99.4	99.1	97.3	91.5	78.6	100.0	99.4	98.9	96.4	92.4	79.6
German	99.3	98.4	97.3	95.2	90.2	70.7	99.3	98.4	97.3	95.7	91.0	72.6
Chinese (S.)	99.7	99.5	98.2	96.0	90.8	74.5	99.7	99.5	98.4	96.3	91.1	77.0
Arabic	99.8	99.1	97.0	95.1	87.7	68.8	99.8	99.1	96.9	95.0	88.8	70.9
Japanese	100.1	99.6	98.2	95.1	89.8	69.5	100.1	99.6	98.0	95.4	90.2	72.1
T4	99.7	98.9	97.6	95.3	89.7	70.7	99.7	98.9	97.9	95.5	90.5	73.1
Catalan	99.8	99.2	98.1	96.5	92.2	74.2	99.8	99.2	98.6	95.6	92.6	75.9
Italian	99.6	98.8	97.1	96.1	92.5	79.0	99.6	98.8	97.6	95.7	92.5	78.7
Croatian	99.7	99.4	98.4	95.0	89.3	69.2	99.7	99.4	97.7	95.7	90.0	72.8
Turkish	99.3	98.6	99.0	97.1	92.1	74.4	99.3	98.6	97.7	96.7	92.8	78.1
Portuguese	100.1	99.5	97.7	96.1	89.4	70.9	100.1	99.5	98.3	96.8	90.7	74.7
Swedish	99.7	98.9	97.9	95.1	89.6	69.9	99.7	98.9	97.8	95.3	89.6	72.8
Vietnamese	99.4	99.0	97.8	94.8	90.3	72.3	99.4	99.0	97.7	95.6	90.5	75.3
Russian	100.0	99.0	98.2	96.3	91.0	74.8	100.0	99.0	98.6	96.2	91.1	75.7
Hungarian	100.0	98.7	97.7	95.6	90.9	69.8	100.0	98.7	98.8	96.6	92.6	71.7
Czech	99.7	99.0	97.3	94.9	88.6	70.1	99.7	99.0	97.5	95.2	89.0	72.4
Serbian	99.6	98.4	96.4	93.3	87.0	65.4	99.6	98.4	97.0	93.9	88.1	68.5
Finnish	99.9	99.3	97.9	96.0	89.4	69.7	99.9	99.3	98.4	96.5	92.5	72.6
Dutch	99.5	98.4	97.4	95.1	90.2	71.2	99.5	98.4	97.7	94.5	89.5	74.4
Polish	99.4	99.1	97.7	94.7	89.3	68.9	99.4	99.1	97.7	95.7	90.1	71.1
Basque	99.8	98.8	98.1	95.2	88.5	70.1	99.8	98.8	97.7	95.7	90.7	72.0
Korean	99.4	98.1	97.0	94.6	88.4	67.3	99.4	98.1	97.2	94.2	88.7	69.0
Hindi	99.4	98.8	96.9	94.8	88.6	70.7	99.4	98.8	98.2	96.0	90.0	74.1
Persian	99.8	98.9	96.7	93.7	86.9	63.8	99.8	98.9	97.4	93.6	88.2	65.9
T3	99.6	98.8	97.4	94.8	88.8	67.2	99.6	98.8	97.7	95.3	89.6	70.7
Afrikaans	99.7	99.3	98.9	95.8	89.6	68.7	99.7	99.3	98.3	95.5	89.9	71.6
Danish	100.0	99.0	98.3	96.5	92.5	75.4	100.0	99.0	98.1	97.4	92.6	77.8
Bulgarian	100.0	99.2	97.6	95.5	90.8	71.8	100.0	99.2	98.3	96.5	91.3	75.5
Indonesian	99.6	99.0	98.1	95.7	90.4	76.4	99.6	99.0	98.2	96.1	91.5	78.4
Slovak	99.6	99.6	98.9	96.4	91.3	74.9	99.6	99.6	98.8	96.9	91.7	77.5
Estonian	99.1	99.2	97.9	95.3	91.0	73.7	99.1	99.2	98.7	96.1	91.1	75.3
Latvian	99.2	98.7	97.5	94.7	89.0	66.1	99.2	98.7	97.5	95.1	90.5	69.6
Romanian	99.8	99.3	97.6	95.1	89.6	69.4	99.8	99.3	97.0	94.4	90.1	71.9
Slovenian	99.4	98.2	96.8	95.9	91.0	69.7	99.4	98.2	97.0	95.7	91.8	73.1
Filipino	99.9	98.0	96.7	93.4	88.5	67.8	99.9	98.0	97.2	93.9	89.1	69.7
Cebuano	98.7	98.5	95.6	92.8	86.8	66.3	98.7	98.5	96.8	92.8	86.9	69.6
Malay	99.3	98.5	97.0	94.4	87.4	67.2	99.3	98.5	97.3	94.3	88.1	70.1
Ukrainian	99.6	98.8	97.6	94.1	87.8	67.0	99.6	98.8	97.5	95.2	88.4	70.5
Hebrew	99.8	98.6	96.9	94.2	87.4	62.0	99.8	98.6	97.0	93.8	87.3	66.7
Greek	100.3	98.6	97.8	96.1	89.5	66.7	100.3	98.6	98.1	95.7	90.2	72.8
Lithuanian	99.8	98.7	96.8	93.1	85.0	64.3	99.8	98.7	97.6	93.6	86.5	66.2
Thai	99.8	99.7	97.9	96.4	91.6	74.5	99.8	99.7	98.1	97.4	92.4	75.3
Uzbek	99.7	98.6	96.3	92.5	84.8	55.2	99.7	98.6	97.3	94.5	88.6	64.2
Georgian	99.8	98.6	96.5	94.8	87.5	61.3	99.8	98.6	96.9	94.9	88.2	64.5
Tamil	99.6	99.6	97.9	96.9	91.3	69.4	99.6	99.6	99.1	96.8	90.8	73.8
Urdu	100.0	98.8	97.2	93.1	86.4	58.7	100.0	98.8	97.7	93.7	86.1	60.4
Kazakh	100.1	99.3	96.7	93.4	86.5	55.7	100.1	99.3	97.8	95.6	89.0	65.3
Bengali	99.0	97.4	96.8	94.3	86.6	64.2	99.0	97.4	97.4	95.1	88.2	67.0

Table 16: Retention (%) at r=7..2 for Gemma4 under top-k reduction and global α. Averaged across Belebele, GlobalMMLU, XCOPA, XStoryCloze, Translation. Per-language mean across available tasks, then mean across languages within tier. Part 1: tiers T5–T3.

	Gemma4											
	Standard						Global α					
Tier/Language	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2
T2	99.5	98.3	96.2	91.9	83.8	61.0	99.5	98.3	96.8	92.9	85.1	64.7
Maltese	99.2	99.0	97.0	93.3	85.5	62.7	99.2	99.0	96.2	93.5	87.7	66.2
Icelandic	98.8	97.1	96.8	94.3	89.3	68.8	98.8	97.1	96.3	94.2	88.8	71.2
Swahili	99.2	98.3	96.7	94.2	88.2	71.8	99.2	98.3	97.1	94.9	88.1	74.0
Marathi	100.3	99.3	98.5	95.6	87.8	58.1	100.3	99.3	99.4	96.4	88.2	66.8
Punjabi	99.2	98.7	97.5	94.0	85.7	57.7	99.2	98.7	96.8	95.3	88.4	61.0
Haitian Creole	100.8	99.4	97.5	94.4	87.6	68.6	100.8	99.4	98.5	95.0	89.4	72.7
Lao	98.7	99.2	96.4	91.9	86.1	60.3	98.7	99.2	96.1	93.5	85.5	62.7
Amharic	99.9	97.0	95.1	89.9	82.1	55.6	99.9	97.0	96.2	92.0	81.8	61.8
Xhosa	99.7	98.4	94.5	89.3	79.6	57.7	99.7	98.4	97.2	91.6	83.2	61.6
Zulu	100.3	99.7	96.5	92.1	85.8	61.8	100.3	99.7	97.4	90.9	82.3	64.2
Hausa	98.9	95.8	93.5	87.7	79.9	56.8	98.9	95.8	95.1	90.9	81.5	61.3
Tswana	98.7	96.9	95.6	91.3	79.0	58.7	98.7	96.9	95.3	89.2	78.1	61.6
Tigrinya	99.0	96.7	93.8	88.1	75.8	46.3	99.0	96.7	94.5	90.2	81.0	49.3
Yoruba	99.2	98.4	95.8	90.1	80.7	60.8	99.2	98.4	96.0	90.9	84.3	64.3
Wolof	100.7	100.0	98.4	91.9	83.3	68.8	100.7	100.0	99.5	95.5	87.8	72.4
T1	99.4	98.0	96.0	92.2	84.0	59.5	99.4	98.0	96.5	92.9	85.1	63.2
Norwegian	99.5	98.2	97.7	95.8	91.8	73.1	99.5	98.2	98.5	96.3	93.2	78.3
Macedonian	99.4	99.2	98.2	95.9	90.4	69.3	99.4	99.2	98.3	94.9	91.7	74.1
Albanian	100.0	98.7	97.8	95.4	89.2	66.9	100.0	98.7	96.9	94.9	89.9	69.7
Armenian	99.1	99.4	97.7	94.7	86.6	62.0	99.1	99.4	97.0	94.5	86.6	65.2
Malayalam	99.3	98.3	96.2	91.0	85.6	57.6	99.3	98.3	96.5	93.1	85.1	64.4
Gujarati	99.8	97.1	96.4	92.3	86.4	55.7	99.8	97.1	96.1	92.7	88.2	61.5
Javanese	100.0	97.9	96.2	90.8	82.6	60.2	100.0	97.9	96.2	91.6	82.5	62.7
Tajik	99.0	97.7	96.1	91.5	80.7	54.8	99.0	97.7	96.3	92.2	84.6	61.4
Kannada	100.8	97.7	95.7	94.6	87.6	57.3	100.8	97.7	96.8	94.3	87.0	65.1
Telugu	99.5	98.4	95.7	93.9	87.4	63.8	99.5	98.4	97.6	94.3	88.7	66.6
Nepali	99.3	97.8	96.0	93.5	84.8	61.2	99.3	97.8	97.2	93.3	86.0	64.7
Khmer	99.2	99.1	97.7	93.0	86.8	60.7	99.2	99.1	96.5	95.1	87.1	65.2
Kurdish	99.2	95.9	94.3	91.1	80.2	53.3	99.2	95.9	94.7	91.8	82.0	56.4
Sundanese	98.3	97.9	95.5	92.6	81.7	53.4	98.3	97.9	96.1	91.9	83.7	59.5
Oriya	99.4	97.8	96.1	93.0	82.9	49.8	99.4	97.8	97.8	95.4	84.7	55.6
Azerbaijani	98.7	97.8	96.9	94.5	85.0	61.4	98.7	97.8	97.6	95.6	86.4	60.7
Mongolian	98.9	98.7	96.5	90.6	82.4	60.5	98.9	98.7	95.9	92.5	86.1	61.8
Sinhala	100.1	98.6	96.2	93.0	83.5	58.2	100.1	98.6	97.1	92.8	85.8	59.8
Burmese	99.3	98.5	96.4	94.3	86.5	63.8	99.3	98.5	98.1	94.8	88.0	66.7
Assamese	98.7	97.1	96.8	93.9	84.6	50.7	98.7	97.1	98.1	94.8	86.8	61.1
Pashto	100.8	98.3	94.7	90.7	81.8	56.5	100.8	98.3	95.6	90.8	81.5	56.1
Kyrgyz	99.6	98.7	97.3	93.3	85.4	59.0	99.6	98.7	98.6	94.3	85.4	61.1
Malagasy	99.1	97.7	95.4	90.2	82.6	58.5	99.1	97.7	96.2	91.9	84.6	61.3
Southern Sotho	98.7	95.9	93.8	90.4	80.3	54.8	98.7	95.9	92.5	90.0	82.8	59.6
Kinyarwanda	98.7	98.8	95.9	92.5	81.0	50.0	98.7	98.8	96.7	91.2	82.8	57.2
Somali	100.0	97.7	96.7	90.6	82.6	59.3	100.0	97.7	96.9	91.4	85.0	63.4
Shona	99.1	97.6	94.1	91.6	83.1	59.0	99.1	97.6	95.5	91.8	83.4	62.5
Chichewa	99.4	97.1	94.7	91.4	84.3	62.0	99.4	97.1	94.7	91.0	82.0	65.4
Igbo	99.5	96.9	93.6	90.2	80.5	59.7	99.5	96.9	96.4	90.6	79.0	65.1
Maori	98.3	96.9	93.7	89.9	84.5	61.8	98.3	96.9	95.0	90.7	79.9	64.0
Guarani	100.1	99.1	95.5	92.7	87.3	72.1	100.1	99.1	96.0	92.4	88.4	70.9
Swati	100.7	99.0	97.6	90.6	78.0	53.3	100.7	99.0	99.0	91.8	81.6	56.7
Oromo	99.5	100.4	97.4	92.8	87.3	58.8	99.5	100.4	97.5	95.4	84.8	60.6
Lingala	99.0	99.6	98.1	93.7	84.7	63.2	99.0	99.6	97.1	92.5	86.1	69.9
Luganda	101.4	97.6	96.9	91.4	80.5	62.5	101.4	97.6	97.8	95.0	84.0	62.3
Tibetan	97.3	95.4	93.1	87.2	77.2	53.2	97.3	95.4	94.2	90.3	77.4	54.8
Bambara	99.6	96.3	93.6	86.7	81.4	64.5	99.6	96.3	92.4	90.4	84.5	67.8
Avg	99.5	98.5	96.8	93.5	86.6	64.5	99.5	98.5	97.2	94.1	87.5	67.8

Table 17: Retention (%) at r=7..2 for Gemma4 under top-k reduction and global α. Averaged across Belebele, GlobalMMLU, XCOPA, XStoryCloze, Translation. Per-language mean across available tasks, then mean across languages within tier. Part 2: tiers T2–T1 and overall mean.

L Detailed Results: Multilingual Gap

	Qwen3																	
	Standard						Global α						Token α					
Tier/Language	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2
T5	99.5	98.9	97.0	91.7	71.4	24.6	99.8	99.7	99.0	97.8	95.9	86.3	99.8	99.6	98.9	97.9	95.7	88.1
English	99.3	99.5	97.7	93.6	73.1	28.6	100.0	100.4	100.1	97.6	97.5	90.8	100.0	100.0	99.5	98.4	97.5	93.7
French	99.3	99.0	97.6	91.5	75.6	26.1	99.3	100.0	98.9	98.9	97.4	87.8	99.2	99.8	99.2	99.3	96.8	90.3
German	99.5	99.3	97.6	92.6	74.3	22.9	100.0	100.0	98.6	97.9	96.8	86.4	100.2	100.3	99.1	97.4	95.7	87.7
Arabic	99.4	98.4	95.7	89.6	61.0	23.3	99.4	99.1	98.9	96.4	94.7	80.5	99.7	99.1	98.6	97.4	94.5	83.7
Spanish	99.5	99.1	97.6	93.1	74.8	23.3	100.1	99.9	99.0	98.7	95.2	87.4	100.0	99.5	99.0	97.9	95.8	88.0
Chinese (S.)	100.0	98.3	96.8	91.9	74.1	24.9	99.7	99.3	98.9	98.1	95.9	87.8	99.7	99.0	98.3	97.0	95.0	88.5
Japanese	99.4	99.0	96.3	89.6	66.9	23.1	99.8	99.2	98.8	97.2	93.7	83.7	99.7	99.5	98.4	97.4	94.5	84.7
T4	99.4	98.5	95.2	87.6	65.6	23.6	99.7	99.5	98.4	96.5	92.9	78.5	99.8	99.4	98.4	96.4	93.5	80.3
Portuguese	99.4	99.0	96.9	89.6	71.5	25.7	99.5	99.6	99.1	98.3	96.2	88.3	99.5	99.6	98.9	97.8	97.1	89.0
Catalan	99.0	98.4	97.1	90.6	69.2	22.3	99.5	99.7	99.3	97.1	94.7	84.4	99.9	99.5	98.7	97.5	94.1	85.1
Swedish	99.5	98.8	96.8	91.0	68.9	24.8	99.7	99.4	98.7	98.0	94.8	82.5	99.7	99.1	98.2	97.9	95.0	81.8
Serbian	99.5	98.0	95.1	85.7	60.8	22.0	99.4	99.5	98.4	96.3	92.5	75.5	99.7	99.2	98.4	96.5	93.1	80.1
Croatian	98.7	98.0	94.9	88.3	63.8	20.7	99.3	99.0	97.9	95.6	91.9	77.7	99.4	98.6	98.5	96.1	91.7	79.3
Czech	100.0	99.5	95.8	90.5	66.9	23.6	100.4	100.3	99.4	97.3	94.4	79.1	100.1	100.0	99.0	96.8	95.5	82.7
Italian	100.2	99.4	98.3	95.2	78.7	23.9	100.3	99.8	99.2	98.1	96.5	87.9	100.4	100.0	99.6	97.5	96.2	90.6
Russian	99.1	98.8	96.1	92.1	70.8	22.7	99.4	100.6	99.3	97.9	94.7	84.2	99.9	100.2	99.4	98.2	95.7	87.2
Vietnamese	99.9	98.8	96.9	89.5	68.4	22.9	99.9	99.0	98.4	96.5	94.6	84.0	100.1	99.2	98.4	97.0	94.9	86.7
Hungarian	99.9	98.8	94.0	84.3	61.3	24.1	100.0	99.8	98.6	96.4	92.5	73.9	99.8	99.4	98.8	96.5	92.4	75.8
Dutch	99.7	99.6	96.6	90.1	68.0	23.9	100.1	99.8	99.3	97.5	94.9	84.1	100.2	100.0	98.8	97.8	96.1	87.2
Turkish	98.8	97.2	92.4	83.2	58.3	23.0	99.4	99.2	97.8	96.3	92.7	72.4	99.5	99.3	98.4	96.4	92.0	75.8
Korean	98.8	97.9	96.4	88.1	68.7	26.2	99.4	99.1	98.9	96.9	93.1	80.4	99.4	99.4	98.8	97.2	94.2	82.2
Polish	99.2	98.2	96.1	88.5	68.8	23.2	99.4	99.3	98.0	96.5	91.7	79.7	100.0	99.3	97.9	95.9	93.9	81.7
Persian	99.2	98.0	92.6	80.5	59.2	23.6	99.5	99.4	97.9	96.0	90.5	71.4	99.9	99.4	98.2	95.5	91.5	71.6
Finnish	99.6	99.0	95.9	84.6	61.3	23.3	100.1	99.7	98.5	96.1	91.0	73.4	100.2	100.1	98.8	96.0	91.7	74.1
Hindi	100.0	97.7	93.2	85.3	61.5	24.4	100.3	99.2	97.9	95.8	91.0	74.3	99.5	99.1	98.3	95.9	92.3	70.1
Basque	99.0	97.2	88.7	79.7	55.3	25.1	98.9	98.0	94.8	90.4	84.4	60.1	98.7	97.5	95.1	89.4	85.2	64.0
T3	99.2	97.4	94.0	85.0	60.3	22.9	99.5	99.1	97.5	95.2	89.7	71.1	99.6	98.8	97.6	95.2	90.0	69.6
Afrikaans	99.1	97.8	95.1	87.6	63.5	23.5	99.6	99.6	97.7	94.9	91.0	75.4	99.5	98.6	97.5	95.7	91.4	73.4
Danish	99.4	98.3	95.3	89.1	65.6	22.8	99.7	98.8	98.1	96.4	93.8	80.0	99.4	98.7	97.8	96.1	94.3	81.0
Romanian	99.8	98.1	96.2	89.8	68.1	22.0	100.0	99.3	98.8	96.9	94.6	81.1	99.7	99.6	98.9	97.3	94.5	82.7
Indonesian	99.5	98.3	95.4	87.7	69.4	24.1	99.7	99.3	98.8	97.7	94.5	85.6	99.8	99.1	98.8	97.2	94.5	84.8
Malay	98.8	98.0	96.6	87.3	69.7	22.0	99.1	99.2	98.7	97.3	93.0	79.3	99.4	99.1	98.7	97.0	94.0	81.1
Bulgarian	99.4	98.0	95.2	89.2	66.6	22.0	99.6	99.9	98.5	97.3	94.1	78.2	99.7	99.7	98.5	96.7	93.7	81.1
Slovak	99.5	97.6	95.7	87.7	64.5	22.9	99.6	99.8	99.3	96.7	93.2	76.9	99.4	99.3	98.6	97.1	94.0	79.0
Filipino	99.8	98.3	95.2	85.7	62.4	23.0	99.9	99.4	98.3	96.2	91.0	70.1	99.9	99.6	97.8	96.4	90.4	71.5
Hebrew	99.4	97.1	93.1	83.6	59.6	23.9	99.8	98.9	97.8	96.0	90.4	75.8	100.4	98.5	98.0	96.0	90.9	68.8
Ukrainian	99.5	98.4	95.5	89.3	65.1	21.6	99.9	99.4	98.2	96.8	93.2	79.1	99.4	99.1	98.1	97.1	94.1	63.9
Latvian	99.9	98.6	96.0	83.3	59.8	23.4	100.3	99.7	98.6	95.4	90.3	69.6	99.9	99.6	98.5	96.1	90.2	71.5
Greek	99.4	97.7	95.0	88.9	63.4	20.7	99.5	98.5	98.2	96.7	92.7	75.4	99.2	98.6	98.0	97.0	93.7	71.9
Slovenian	98.9	97.8	94.0	86.2	59.6	23.8	99.2	98.5	97.7	95.1	90.1	71.5	99.4	98.5	97.7	95.2	91.0	74.4
Lithuanian	98.6	97.3	93.2	82.1	59.9	22.6	98.6	98.4	97.0	94.9	89.2	72.8	98.6	98.2	97.3	94.4	89.3	71.3
Cebuano	99.1	95.3	90.3	81.6	58.7	24.4	99.5	97.6	96.6	93.7	84.2	62.3	99.4	98.1	96.6	94.0	85.2	67.2
Thai	98.8	97.6	95.9	86.2	60.1	20.1	100.0	99.6	98.3	97.1	92.0	73.7	99.7	99.3	98.2	97.0	92.3	76.8
Estonian	98.8	96.9	94.7	84.6	58.8	23.6	99.5	98.9	96.4	93.2	87.9	70.0	99.2	98.2	96.5	93.0	87.8	72.0
Bengali	99.5	96.9	92.5	80.9	55.4	22.3	99.1	98.4	97.0	94.4	87.1	65.4	99.6	98.4	96.5	93.4	87.4	62.5
Urdu	98.8	96.1	92.3	83.7	54.7	22.7	99.5	99.1	96.4	93.9	85.8	62.6	99.5	98.3	97.5	94.5	86.7	63.1
Uzbek	99.0	96.5	92.3	80.0	52.9	26.2	99.8	99.2	94.1	91.7	85.4	61.0	100.2	98.5	96.7	91.6	84.4	42.5
Tamil	99.1	96.6	89.2	78.5	49.6	22.1	99.1	98.6	95.8	92.8	85.4	58.6	100.3	98.9	96.2	92.9	85.1	50.5
Kazakh	97.4	96.4	91.4	80.8	51.1	23.3	98.4	98.9	94.9	90.7	83.1	60.5	98.6	98.4	95.9	90.2	85.4	59.4
Georgian	99.7	97.0	92.8	81.8	47.3	24.3	99.7	99.6	96.1	93.4	81.9	50.2	100.0	98.1	96.9	93.4	79.4	49.3

Table 18: Retention (%) at r=7..2 for Qwen3 under top-k reduction, global α, and token α. Averaged across Belebele and Translation. Per-language mean across available tasks, then mean across languages within tier. Qwen3 is restricted to its 73 supported languages. Part 1: tiers T5–T3.

	Qwen3																	
	Standard						Global α						Token α					
Tier/Language	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2
T2	99.3	96.6	90.7	78.6	52.2	27.2	98.9	98.0	95.3	90.0	81.7	58.0	99.2	97.4	94.4	89.1	80.8	57.8
Marathi	98.8	95.8	91.5	80.8	52.8	25.7	99.7	98.0	96.5	91.9	84.9	61.5	99.1	97.4	95.2	91.1	83.8	60.8
Punjabi	99.1	96.9	90.5	75.6	51.2	23.9	99.3	98.2	95.1	92.1	85.9	60.2	99.8	97.7	94.0	89.9	83.6	55.4
Icelandic	99.2	97.0	92.9	82.0	54.4	27.1	97.8	98.7	96.4	91.7	82.9	61.8	98.0	98.1	94.9	90.5	81.4	65.6
Maltese	100.1	96.7	91.1	79.4	54.2	27.3	99.5	97.9	95.3	90.4	80.0	59.3	100.4	97.7	94.2	88.4	80.4	55.4
Lao	99.6	97.9	90.4	78.6	45.1	21.3	99.1	99.3	95.1	89.4	81.8	54.1	99.3	98.2	95.2	89.7	82.6	55.3
Haitian Creole	99.3	96.6	89.5	78.1	55.1	30.3	99.0	97.2	96.3	87.9	78.4	59.7	98.2	97.2	95.0	87.4	77.3	61.5
Swahili	98.8	95.1	89.1	75.5	52.6	34.5	98.1	96.4	92.1	86.6	78.1	49.4	99.3	95.9	92.3	86.4	76.8	50.6
T1	99.2	96.3	90.9	79.0	51.9	23.5	99.5	98.3	95.8	92.0	84.2	58.4	99.2	97.8	95.4	91.6	83.9	55.7
Norwegian	99.5	97.9	96.0	88.6	62.2	22.8	99.3	98.6	98.7	96.8	93.8	80.6	99.1	98.9	98.6	97.5	94.0	82.6
Macedonian	99.8	97.7	93.4	84.8	61.2	22.4	99.7	99.3	97.2	96.2	91.0	74.2	100.0	99.0	97.5	95.8	91.6	74.8
Albanian	100.0	98.3	93.9	84.3	58.8	23.8	100.5	99.4	97.0	94.0	88.3	72.2	99.5	98.6	97.4	94.0	88.5	61.0
Armenian	99.0	97.2	91.7	78.5	51.4	18.9	98.2	98.4	98.0	94.0	84.8	59.3	98.7	98.4	97.1	92.7	85.1	55.5
Malayalam	98.1	96.0	89.8	77.2	51.6	22.5	99.0	98.7	95.0	90.1	84.0	56.3	98.8	98.5	94.9	90.9	85.5	54.9
Gujarati	98.3	95.7	91.9	79.7	50.5	21.0	98.8	97.8	94.7	92.0	87.1	58.4	98.7	97.5	94.7	91.0	86.7	47.9
Javanese	100.0	97.8	92.6	78.8	57.2	26.0	100.1	98.3	97.2	94.1	86.3	64.5	100.3	98.0	96.4	94.1	84.5	60.6
Nepali	98.7	95.1	90.2	81.0	54.9	25.6	99.2	98.3	95.7	92.3	83.2	59.2	99.5	98.0	96.0	91.1	83.0	50.1
Kannada	99.9	96.1	89.5	76.9	51.4	21.2	100.0	98.3	95.7	91.0	85.4	51.7	100.0	97.8	94.2	89.4	84.7	54.5
Telugu	99.9	97.2	90.1	75.4	49.1	22.8	100.0	97.9	95.8	91.2	84.2	43.8	99.6	97.5	95.2	88.6	85.1	43.7
Oriya	99.5	95.2	88.9	74.6	47.5	22.3	99.2	97.8	93.9	88.1	79.3	50.9	99.0	97.1	91.8	86.5	80.0	52.3
Azerbaijani	98.8	97.1	92.4	79.9	56.5	24.7	99.5	99.2	95.9	93.8	86.2	60.8	99.8	98.4	96.4	93.4	85.8	57.7
Sundanese	99.5	96.2	92.1	79.8	57.3	25.6	99.7	97.8	95.4	90.8	82.5	58.5	98.7	97.7	95.3	91.0	80.4	59.0
Assamese	98.9	96.6	89.4	75.9	50.7	26.1	99.9	98.1	96.2	89.5	80.3	52.0	99.0	97.6	94.6	90.2	78.9	49.5
Khmer	98.3	92.5	87.5	79.8	39.4	20.3	98.5	99.6	96.7	92.6	79.3	53.9	98.9	97.6	96.6	94.2	83.8	53.5
Sinhala	99.2	95.1	87.3	73.5	47.5	26.8	99.0	96.4	92.8	86.4	77.0	47.7	99.7	96.2	93.2	87.6	76.3	46.7
Tajik	98.3	96.9	91.4	77.1	46.1	27.6	98.4	96.3	91.9	87.0	77.5	51.1	97.8	95.4	91.1	85.2	74.3	44.7
Burmese	99.6	94.2	88.8	77.1	41.5	23.2	101.4	99.2	96.3	96.3	85.7	55.1	99.3	98.7	96.0	94.6	82.5	54.4
Avg	99.3	97.5	93.5	84.2	59.8	23.8	99.5	98.9	97.2	94.5	89.0	70.0	99.5	98.6	97.1	94.3	89.0	69.4

Table 19: Retention (%) at r=7..2 for Qwen3 under top-k reduction, global α, and token α. Averaged across Belebele and Translation. Per-language mean across available tasks, then mean across languages within tier. Qwen3 is restricted to its 73 supported languages. Part 2: tiers T2–T1 and overall mean.

	Qwen3.5																	
	Standard						Global α						Token α					
Tier/Language	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2
T5	99.7	99.4	97.8	94.3	84.0	54.2	100.1	100.1	100.1	99.8	99.1	96.6	100.0	100.1	100.0	99.9	99.3	97.1
English	99.7	99.0	98.8	97.4	90.0	65.4	99.9	99.4	100.0	99.9	99.7	99.1	100.1	99.7	100.0	99.9	99.3	98.6
German	99.9	99.5	98.4	94.6	83.5	53.1	100.2	100.2	100.4	99.8	98.8	96.8	99.8	100.1	99.9	100.1	99.3	97.6
French	99.5	99.7	98.8	96.8	87.1	59.1	99.5	100.0	100.0	100.0	99.7	97.8	100.0	99.9	100.4	100.2	99.6	97.8
Arabic	99.2	98.2	94.7	91.0	77.9	45.9	100.1	99.8	99.3	99.6	99.4	95.7	99.8	99.3	99.2	99.1	99.4	94.8
Spanish	100.0	99.6	98.8	96.2	87.6	55.1	100.0	100.2	99.9	99.7	99.7	98.3	99.9	100.2	100.1	100.0	99.1	98.0
Chinese (S.)	99.7	100.2	98.8	92.2	80.2	49.6	100.2	99.9	99.6	98.9	97.0	95.0	100.1	100.1	99.7	99.1	98.6	97.5
Japanese	99.7	99.9	96.2	91.9	81.9	51.5	100.7	101.2	101.4	100.4	99.2	93.4	100.3	101.5	100.9	100.5	100.0	95.1
T4	99.5	98.7	97.4	94.6	82.8	49.5	99.8	100.0	99.8	99.6	98.7	95.5	99.9	99.8	99.7	99.4	98.8	96.0
Portuguese	99.6	98.7	97.8	95.1	81.1	53.0	99.7	99.9	99.6	99.3	99.1	96.8	99.8	99.8	99.7	99.3	99.0	97.1
Catalan	99.5	100.0	98.7	96.7	86.4	51.9	99.7	100.1	99.4	99.7	97.9	94.2	99.9	99.7	99.7	99.3	98.8	96.2
Swedish	99.4	98.4	97.4	94.4	88.8	53.1	99.6	99.7	99.5	99.5	99.4	96.5	99.5	99.3	99.3	99.5	98.5	96.7
Serbian	99.6	99.0	98.4	96.6	80.0	45.8	99.4	100.3	100.3	99.3	98.5	94.9	99.5	100.0	99.7	98.5	98.1	94.8
Czech	99.1	97.8	97.7	95.2	86.2	52.3	99.6	99.6	99.5	99.6	98.9	96.5	100.1	99.3	99.8	99.3	98.6	96.2
Russian	99.6	99.5	96.8	95.8	82.7	53.0	100.1	100.0	100.0	99.1	98.9	96.2	100.4	100.0	99.6	99.4	98.3	96.9
Persian	98.8	98.9	97.5	94.8	85.1	42.7	99.4	99.6	98.9	98.4	97.3	95.6	99.6	99.3	99.4	98.9	97.8	95.5
Hungarian	99.9	98.7	94.7	91.8	80.5	47.0	99.9	99.8	99.8	99.4	98.9	95.1	100.0	99.8	100.1	100.2	98.7	95.6
Italian	99.3	99.4	99.0	96.3	87.4	58.0	99.7	99.8	99.8	99.8	99.3	97.0	99.5	99.5	99.8	99.8	99.2	97.2
Vietnamese	99.9	99.8	98.8	94.2	77.3	51.8	100.5	100.5	100.4	100.1	99.2	96.9	100.2	100.6	100.2	100.0	99.8	97.3
Turkish	99.5	98.1	97.0	92.4	80.4	46.8	100.3	100.2	99.8	99.7	99.2	95.3	100.1	100.0	100.1	99.0	98.6	95.2
Croatian	99.2	97.1	96.9	94.2	86.2	48.2	100.2	101.3	99.8	101.0	101.1	98.8	100.7	100.9	100.0	100.7	101.1	98.7
Finnish	98.9	96.6	94.3	92.0	82.8	46.0	99.5	99.0	100.2	100.3	98.5	93.8	99.1	98.6	98.9	98.8	99.4	94.9
Korean	99.9	99.7	97.5	93.6	79.6	52.7	100.2	100.1	100.0	100.3	98.6	94.5	99.9	99.7	100.1	99.5	98.6	95.3
Dutch	100.3	99.4	98.8	96.0	83.6	54.0	100.2	99.9	100.0	100.0	98.7	96.9	100.0	99.7	100.1	99.7	99.0	96.8
Polish	99.8	98.7	97.3	96.2	82.5	42.3	99.7	99.8	99.4	99.0	98.0	95.6	99.7	99.4	99.3	99.0	98.2	96.0
Hindi	99.7	98.6	97.5	94.5	82.3	48.9	99.9	99.3	99.3	99.1	97.5	93.9	99.9	99.7	99.3	99.2	97.6	93.7
Basque	99.0	98.8	96.4	93.0	77.8	44.0	99.6	100.4	100.1	98.5	96.7	90.9	99.9	100.6	100.0	99.2	98.4	93.3
T3	99.6	98.5	96.8	93.3	82.0	44.9	99.9	99.7	99.5	99.3	98.1	93.9	99.9	99.7	99.7	99.3	98.1	94.4
Afrikaans	99.9	99.0	97.7	95.3	86.4	43.2	100.0	100.1	99.4	99.5	99.1	95.1	99.9	100.0	99.4	99.4	98.8	95.7
Danish	99.2	98.3	96.0	93.1	84.1	55.3	99.4	99.3	99.1	99.7	98.5	96.1	99.6	99.3	98.8	98.8	98.0	96.4
Malay	99.6	99.0	98.6	96.7	87.1	52.4	100.1	99.6	99.8	99.7	98.2	95.7	100.0	99.6	99.9	99.5	98.4	96.3
Romanian	100.0	98.0	95.5	91.1	83.3	46.5	99.7	99.8	99.6	99.8	98.3	95.8	100.1	99.9	100.1	99.2	98.6	95.4
Indonesian	99.8	99.2	98.7	96.0	85.9	54.3	100.1	100.1	99.9	99.7	98.7	97.5	99.9	100.3	99.7	99.9	98.8	97.4
Bulgarian	100.1	99.4	98.0	96.6	86.8	52.5	99.9	100.1	100.6	100.7	99.6	97.5	99.9	100.3	100.3	100.7	100.0	96.3
Filipino	99.7	98.6	97.9	92.8	80.5	40.3	99.7	99.5	99.4	98.5	96.4	92.2	99.9	99.8	99.3	98.3	97.3	92.9
Slovak	99.6	97.3	97.2	94.3	83.0	47.6	99.9	99.0	98.9	99.3	98.9	95.0	99.9	99.6	98.9	98.2	98.6	95.6
Ukrainian	99.4	97.9	97.4	92.0	84.0	51.0	100.0	99.9	99.9	99.3	98.4	94.7	99.9	99.5	100.9	99.8	98.9	95.3
Hebrew	100.3	98.5	97.5	94.0	84.0	47.5	100.1	100.3	100.5	100.0	98.9	96.1	100.0	100.1	100.6	99.9	99.0	96.4
Latvian	99.1	98.3	97.2	91.6	80.6	41.6	99.7	99.7	99.2	99.0	98.4	93.4	99.5	99.6	99.2	98.7	97.9	93.7
Slovenian	98.9	97.7	96.5	94.6	81.9	42.1	99.1	98.7	98.5	99.2	98.2	94.2	99.6	99.4	98.7	99.8	98.0	93.5
Estonian	99.8	98.4	96.7	90.6	80.5	35.7	99.2	99.8	99.5	99.3	99.2	92.0	100.1	99.0	100.1	99.5	98.3	93.4
Greek	99.0	98.3	95.0	93.3	84.7	46.0	99.9	100.3	99.9	100.2	98.2	95.1	99.4	99.5	100.1	99.9	98.8	95.8
Lithuanian	98.8	98.5	99.1	95.8	82.6	42.8	99.9	100.8	101.0	100.5	99.2	95.0	99.7	100.3	100.4	99.8	98.2	94.2
Uzbek	100.0	99.1	98.2	92.6	75.9	42.8	100.2	100.2	99.6	99.2	98.4	91.3	100.0	99.9	99.4	99.3	96.9	91.5
Cebuano	100.4	98.7	97.4	93.4	81.1	45.3	100.9	99.5	99.2	98.7	95.8	90.6	100.4	99.5	99.6	98.2	96.5	91.8
Bengali	99.7	98.7	97.2	94.2	81.4	45.1	99.7	99.3	98.6	98.3	96.9	91.0	100.4	99.4	98.9	98.5	96.7	93.5
Urdu	99.2	98.1	96.0	91.6	72.8	36.8	99.8	99.3	98.6	98.0	97.1	91.8	99.8	99.6	99.5	99.1	97.5	91.3
Thai	100.1	97.7	89.0	90.2	85.8	51.2	100.0	99.9	99.6	99.7	98.7	95.7	100.1	100.1	100.0	99.6	98.5	96.3
Georgian	99.2	98.8	95.9	91.8	77.9	35.2	99.7	99.1	98.6	99.1	97.1	90.4	99.7	99.8	99.6	99.2	97.1	92.9
Kazakh	99.0	97.9	95.6	93.0	76.4	38.3	99.9	99.3	98.9	98.7	96.9	91.3	99.8	99.9	99.3	98.7	97.2	92.6
Tamil	99.2	99.4	97.0	92.3	79.3	39.3	99.6	99.5	99.9	98.9	97.7	91.0	100.0	99.4	100.1	99.2	98.3	92.7

Table 20: Retention (%) at r=7..2 for Qwen3.5 under top-k reduction, global α, and token α. Averaged across Belebele and Translation. Per-language mean across available tasks, then mean across languages within tier. Part 1: tiers T5–T3.

	Qwen3.5																	
	Standard						Global α						Token α					
Tier/Language	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2
T2	99.6	98.3	95.7	90.3	74.7	39.1	99.8	99.7	98.7	97.6	94.6	85.7	99.8	99.4	98.9	97.9	95.4	87.6
Maltese	100.0	98.7	98.0	94.8	81.0	39.8	99.6	100.5	99.6	98.9	97.4	90.6	99.7	99.9	99.7	99.0	97.5	92.8
Swahili	99.1	98.4	95.8	91.5	79.2	41.6	100.4	99.5	99.2	98.1	96.6	89.3	100.2	99.3	99.1	97.6	96.7	90.9
Icelandic	100.3	98.8	98.8	94.2	82.3	40.1	101.1	100.5	100.3	99.9	97.7	92.3	100.5	100.3	100.9	100.0	98.5	94.4
Marathi	99.7	98.7	97.8	94.2	81.4	41.8	100.6	100.0	100.3	99.5	97.1	92.1	100.3	99.9	100.0	99.0	96.7	92.3
Punjabi	100.7	100.0	98.4	94.7	83.6	37.0	100.4	100.4	99.9	99.9	98.4	92.5	100.6	99.8	100.1	99.9	98.4	93.1
Haitian Creole	98.9	97.3	95.4	93.3	76.5	40.7	99.2	101.1	100.2	99.4	97.4	87.9	99.7	100.7	100.0	99.3	97.4	89.1
Lao	98.7	98.5	96.5	89.3	68.9	33.0	98.5	99.6	99.4	98.5	97.2	88.8	99.4	99.4	99.4	98.7	96.7	89.6
Amharic	99.6	98.3	92.7	86.2	68.8	31.3	100.3	99.5	98.4	98.7	93.1	81.5	100.1	99.1	98.6	98.6	95.0	85.3
Xhosa	99.0	98.0	94.8	86.9	67.8	34.0	100.1	99.1	99.8	96.9	93.0	83.9	99.8	99.0	99.1	97.0	94.0	86.3
Zulu	99.5	98.5	94.1	91.2	76.4	42.5	99.8	99.1	98.1	97.6	95.7	87.0	98.8	99.8	98.5	98.3	95.1	88.8
Hausa	99.9	98.9	96.9	88.5	72.3	42.1	99.8	98.9	98.6	97.6	91.9	81.2	99.8	98.9	98.5	96.6	95.0	86.8
Tigrinya	99.5	97.2	94.5	88.3	69.0	35.1	99.5	98.1	95.8	95.5	92.1	79.6	99.4	98.6	97.0	97.3	92.8	81.9
Tswana	99.9	98.4	93.7	86.3	68.1	38.2	99.1	99.4	97.3	96.9	90.3	81.8	99.3	98.3	98.2	97.8	94.1	84.7
Yoruba	99.2	97.5	93.9	87.1	68.0	42.8	100.7	100.9	97.0	94.5	92.4	79.9	99.6	99.2	98.8	95.3	91.9	81.3
Wolof	99.9	96.7	95.0	88.2	78.0	45.6	97.7	98.2	97.1	92.2	88.1	76.5	99.3	99.2	96.2	94.3	90.8	77.2
T1	99.2	97.9	95.6	90.0	74.0	39.6	99.8	99.5	98.6	97.4	94.0	85.6	99.7	99.4	98.8	97.6	94.9	87.1
Norwegian	99.1	98.4	96.8	93.1	83.5	48.7	99.8	99.7	99.8	99.7	99.1	95.7	99.7	99.7	99.7	99.2	98.6	96.4
Macedonian	99.6	98.3	95.4	95.0	85.1	45.9	99.8	99.8	100.1	99.7	98.5	95.1	100.2	99.6	99.3	99.0	99.0	95.9
Albanian	99.4	97.2	96.9	92.7	80.2	42.2	99.8	100.3	99.6	99.1	98.2	93.4	99.9	99.2	99.4	99.3	97.9	94.0
Armenian	99.3	98.3	97.1	94.1	78.8	42.8	100.3	100.1	99.8	99.2	98.1	92.9	100.6	100.7	100.4	99.7	98.5	94.0
Malayalam	99.4	98.9	97.4	93.5	79.8	34.1	100.1	99.7	99.9	99.9	97.7	91.6	99.7	100.2	99.6	99.9	97.8	93.8
Javanese	99.7	98.7	98.0	92.7	75.0	40.5	99.6	100.2	99.1	98.3	96.8	89.5	99.5	100.1	99.3	99.4	97.1	91.7
Gujarati	99.5	98.7	97.1	92.4	79.6	35.1	100.6	100.1	99.7	98.8	96.0	89.0	100.1	99.2	99.2	98.4	96.3	89.6
Oriya	100.0	99.5	98.3	94.3	75.9	38.9	100.6	101.1	99.7	99.4	97.0	90.0	99.9	100.4	100.1	99.7	97.3	91.0
Nepali	100.5	100.2	98.1	94.8	78.5	40.8	100.7	101.1	101.6	100.4	98.2	93.1	100.4	101.9	101.6	100.7	98.5	91.8
Telugu	99.0	98.1	97.0	90.3	73.3	36.0	99.7	99.4	98.9	97.4	95.2	88.9	99.2	99.5	99.7	98.0	95.5	91.8
Sundanese	99.3	97.9	96.0	89.8	76.5	36.4	99.0	99.2	98.9	97.1	95.2	90.5	99.3	99.4	99.0	97.9	95.6	90.9
Sinhala	98.8	98.4	95.2	90.1	77.5	34.2	98.8	99.0	99.0	98.1	96.1	89.7	99.9	99.5	100.2	99.7	95.9	92.1
Kannada	99.6	98.9	98.1	92.2	65.8	31.6	99.7	99.8	99.9	98.6	97.1	90.3	100.1	100.1	99.9	98.8	97.1	92.0
Tajik	101.1	99.0	96.7	93.3	80.2	38.4	100.7	100.7	100.2	98.6	96.4	88.4	100.5	100.3	100.0	99.0	96.4	90.8
Kyrgyz	99.8	98.1	94.8	89.8	78.3	38.1	99.5	99.4	98.2	96.9	95.6	90.5	99.0	99.2	98.6	96.6	96.5	90.3
Kurdish	99.0	98.6	94.9	89.4	67.7	32.4	99.0	99.5	99.1	97.3	92.6	85.0	100.2	99.1	98.3	97.1	94.9	84.6
Assamese	99.1	99.2	96.4	93.3	76.2	37.6	100.0	99.4	98.9	97.0	95.1	87.5	99.6	99.4	99.4	97.7	95.7	88.8
Khmer	99.3	97.9	96.5	93.4	70.7	35.6	100.1	100.1	99.6	98.9	97.4	90.4	100.6	100.3	99.9	99.5	98.9	93.0
Azerbaijani	98.8	96.9	96.4	94.4	80.5	42.0	99.2	99.7	98.6	98.4	95.3	88.2	99.0	99.5	98.8	98.0	96.5	89.0
Pashto	99.7	98.0	96.1	89.5	72.6	36.5	99.9	99.7	99.0	98.0	94.8	86.1	100.3	99.4	99.8	98.9	94.4	86.3
Burmese	99.1	95.8	95.5	90.5	76.2	37.1	99.8	99.8	99.2	99.2	96.4	89.5	100.4	99.3	99.4	98.9	97.9	91.4
Mongolian	99.2	97.9	96.1	89.5	65.3	32.3	99.9	100.4	98.6	99.7	96.2	87.9	99.6	99.8	100.0	98.8	96.0	89.7
Malagasy	98.3	97.4	94.2	88.9	74.3	36.8	99.6	98.9	97.9	96.9	91.8	82.7	99.3	98.8	98.1	97.7	94.6	87.5
Southern Sotho	98.9	99.2	95.5	88.5	71.8	37.8	99.6	99.2	97.0	97.4	90.7	80.1	99.6	99.0	98.9	98.0	93.5	83.1
Kinyarwanda	99.1	97.0	97.2	92.2	75.6	38.0	100.8	100.0	99.3	98.2	95.8	85.6	100.0	99.2	98.2	97.7	97.2	87.7
Somali	100.4	98.9	94.8	87.8	67.4	38.0	100.8	101.5	101.0	99.5	94.2	82.9	100.8	100.9	100.4	98.6	96.0	82.6
Maori	98.5	96.0	91.9	84.3	70.8	40.6	99.8	99.9	97.1	97.2	91.9	82.0	99.5	100.0	98.1	97.5	93.9	84.4
Shona	98.8	97.1	95.7	89.1	71.2	37.1	99.3	98.6	99.3	97.1	93.7	81.2	99.7	98.4	98.3	96.1	94.0	84.3
Igbo	100.1	97.4	95.0	86.7	71.4	39.9	100.1	98.2	98.1	96.1	90.9	80.5	100.0	99.4	98.1	96.8	91.3	80.3
Chichewa	100.1	97.9	94.8	89.6	66.3	43.1	99.1	98.4	97.9	97.4	90.0	80.0	100.3	99.6	96.5	97.3	93.9	85.5
Swati	97.3	97.8	91.7	82.2	70.1	40.4	98.3	97.8	94.7	92.6	87.7	73.1	98.8	97.7	94.5	93.3	88.3	77.4
Guarani	100.1	97.2	91.2	82.8	64.3	43.3	101.0	99.3	96.6	96.7	90.0	79.0	99.9	99.0	98.9	95.5	92.3	77.7
Lingala	98.5	96.7	94.9	87.4	73.8	45.7	99.2	99.2	96.5	93.5	88.2	78.3	98.7	99.4	97.9	94.1	91.6	79.6
Oromo	99.2	95.5	94.1	86.9	66.6	40.8	98.9	98.9	95.3	91.4	85.6	70.2	98.3	97.4	96.2	93.9	89.1	75.2
Luganda	97.8	98.0	94.6	88.5	74.9	45.9	99.5	97.6	96.8	94.6	87.5	76.0	98.6	98.8	97.8	96.7	90.6	79.9
Tibetan	97.1	94.2	89.4	82.1	67.4	44.2	97.6	99.1	96.3	93.3	87.8	75.2	99.6	96.2	97.0	94.9	89.0	73.9
Bambara	99.0	98.6	95.9	85.4	74.2	57.4	100.8	96.9	97.0	93.8	87.9	77.0	99.0	97.7	95.9	89.8	85.1	76.2
Avg	99.4	98.3	96.3	91.9	78.2	43.6	99.8	99.7	99.1	98.4	96.2	90.1	99.8	99.6	99.3	98.5	96.7	91.2

Table 21: Retention (%) at r=7..2 for Qwen3.5 under top-k reduction, global α, and token α. Averaged across Belebele and Translation. Per-language mean across available tasks, then mean across languages within tier. Part 2: tiers T2–T1 and overall mean.

	Gemma4											
	Standard						Global α					
Tier/Language	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2
T5	99.7	99.2	98.0	96.2	91.0	72.4	99.7	99.2	98.0	96.2	91.6	74.5
English	99.9	98.8	97.6	95.6	91.3	72.1	99.9	98.8	97.6	96.3	91.9	76.5
French	99.4	99.3	97.2	96.1	93.7	76.4	99.4	99.3	97.8	96.2	93.3	77.7
German	99.3	98.8	97.9	96.5	92.5	74.3	99.3	98.8	98.1	97.0	93.3	75.8
Arabic	99.7	99.3	97.0	95.8	87.6	65.8	99.7	99.3	96.8	95.4	89.1	68.2
Chinese (S.)	99.3	99.1	98.0	95.6	89.0	69.5	99.3	99.1	98.4	96.0	89.2	71.5
Spanish	100.3	99.6	99.7	97.8	90.9	75.8	100.3	99.6	99.0	96.4	92.0	77.0
Japanese	100.3	99.8	98.8	95.7	91.7	72.9	100.3	99.8	98.5	96.4	92.0	74.6
T4	99.6	99.0	97.8	95.7	90.6	71.1	99.6	99.0	98.1	96.0	91.4	73.8
Catalan	99.8	99.2	98.1	96.5	92.2	74.2	99.8	99.2	98.6	95.6	92.6	75.9
Portuguese	100.2	99.7	98.2	97.3	92.3	75.2	100.2	99.7	99.1	98.2	93.7	79.4
Swedish	99.6	99.1	98.2	96.2	92.5	74.7	99.6	99.1	98.3	96.4	92.2	77.2
Serbian	99.5	98.7	96.7	94.3	89.5	69.0	99.5	98.7	97.6	94.7	90.3	72.0
Czech	99.7	99.0	97.6	96.0	90.7	74.7	99.7	99.0	97.9	96.4	91.1	76.2
Croatian	99.7	99.4	98.4	95.0	89.3	69.2	99.7	99.4	97.7	95.7	90.0	72.8
Italian	99.1	98.5	96.8	95.4	91.8	76.4	99.1	98.5	96.8	95.2	92.2	77.1
Vietnamese	98.7	98.4	98.0	95.1	91.0	69.6	98.7	98.4	97.4	95.3	91.0	74.2
Turkish	99.3	98.2	98.9	96.1	90.3	67.4	99.3	98.2	97.2	95.9	90.6	72.9
Russian	100.5	99.3	98.4	96.9	92.9	75.8	100.5	99.3	99.0	96.8	91.8	76.0
Hungarian	100.0	98.7	97.7	95.6	90.9	69.8	100.0	98.7	98.8	96.6	92.6	71.7
Dutch	99.5	98.3	97.4	95.5	92.2	74.8	99.5	98.3	98.0	94.8	90.7	78.6
Finnish	99.9	99.3	97.9	96.0	89.4	69.7	99.9	99.3	98.4	96.5	92.5	72.6
Persian	100.1	99.2	97.1	94.8	89.6	67.0	100.1	99.2	98.5	94.7	90.9	68.7
Polish	99.4	99.6	98.2	95.4	91.3	72.3	99.4	99.6	98.0	96.8	92.5	74.2
Hindi	99.3	99.6	97.1	95.8	88.5	67.9	99.3	99.6	98.5	96.8	90.7	72.0
Korean	99.4	98.2	97.4	95.4	90.0	68.8	99.4	98.2	97.5	95.0	89.8	70.4
Basque	99.8	99.2	98.0	94.5	87.0	62.9	99.8	99.2	97.9	96.0	90.5	66.4
T3	99.7	98.9	97.6	95.1	89.2	67.2	99.7	98.9	97.9	95.5	90.0	70.8
Afrikaans	99.7	99.3	98.9	95.8	89.6	68.7	99.7	99.3	98.3	95.5	89.9	71.6
Danish	100.0	99.0	98.3	96.5	92.5	75.4	100.0	99.0	98.1	97.4	92.6	77.8
Romanian	99.7	99.7	98.0	96.1	92.0	72.7	99.7	99.7	97.2	95.2	92.4	75.1
Filipino	99.8	97.6	96.6	93.7	90.0	69.9	99.8	97.6	97.4	94.0	90.6	72.3
Bulgarian	100.0	99.2	97.6	95.5	90.8	71.8	100.0	99.2	98.3	96.5	91.3	75.5
Malay	99.4	98.8	97.4	95.5	89.4	70.4	99.4	98.8	98.1	95.1	89.8	72.8
Hebrew	100.0	98.9	97.6	95.4	90.0	64.4	100.0	98.9	97.7	94.6	89.8	69.6
Indonesian	99.5	99.1	98.7	96.5	91.2	71.7	99.5	99.1	98.4	96.6	91.6	74.8
Slovak	99.6	99.6	98.9	96.4	91.3	74.9	99.6	99.6	98.8	96.9	91.7	77.5
Latvian	99.2	98.7	97.5	94.7	89.0	66.1	99.2	98.7	97.5	95.1	90.5	69.6
Ukrainian	99.8	99.2	98.2	94.9	89.4	70.2	99.8	99.2	98.1	96.5	90.0	73.8
Greek	100.7	98.8	98.7	97.7	92.1	71.1	100.7	98.8	99.1	97.0	92.5	78.2
Lithuanian	100.0	98.9	97.5	93.8	86.7	68.0	100.0	98.9	98.1	94.3	88.7	69.4
Slovenian	99.4	98.2	96.8	95.9	91.0	69.7	99.4	98.2	97.0	95.7	91.8	73.1
Estonian	98.8	98.8	97.8	94.3	89.1	68.8	98.8	98.8	97.8	95.7	90.0	71.8
Cebuano	98.7	98.5	95.6	92.8	86.8	66.3	98.7	98.5	96.8	92.8	86.9	69.6
Uzbek	99.7	98.6	96.3	92.5	84.8	55.2	99.7	98.6	97.3	94.5	88.6	64.2
Bengali	98.9	97.4	97.7	95.7	88.1	65.6	98.9	97.4	98.2	96.2	89.6	68.2
Georgian	99.8	98.6	96.5	94.8	87.5	61.3	99.8	98.6	96.9	94.9	88.2	64.5
Urdu	100.0	98.8	97.2	93.1	86.4	58.7	100.0	98.8	97.7	93.7	86.1	60.4
Thai	99.7	99.6	97.9	96.1	89.5	67.9	99.7	99.6	98.0	97.3	90.6	68.2
Kazakh	100.1	99.3	96.7	93.4	86.5	55.7	100.1	99.3	97.8	95.6	89.0	65.3
Tamil	99.7	99.8	97.2	95.4	88.6	61.2	99.7	99.8	98.6	95.4	87.6	65.5

Table 22: Retention (%) at r=7..2 for Gemma4 under top-k reduction and global α. Averaged across Belebele and Translation. Per-language mean across available tasks, then mean across languages within tier. Part 1: tiers T5–T3.

	Gemma4											
	Standard						Global α					
Tier/Language	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2	Ret_7	Ret_6	Ret_5	Ret_4	Ret_3	Ret_2
T2	99.5	98.2	96.2	91.6	83.6	59.8	99.5	98.2	96.8	92.8	84.8	63.8
Maltese	99.2	99.0	97.0	93.3	85.5	62.7	99.2	99.0	96.2	93.5	87.7	66.2
Swahili	99.5	98.2	96.5	93.3	86.7	65.0	99.5	98.2	97.0	94.2	85.7	68.5
Icelandic	98.8	97.1	96.8	94.3	89.3	68.8	98.8	97.1	96.3	94.2	88.8	71.2
Marathi	100.3	99.3	98.5	95.6	87.8	58.1	100.3	99.3	99.4	96.4	88.2	66.8
Punjabi	99.2	98.7	97.5	94.0	85.7	57.7	99.2	98.7	96.8	95.3	88.4	61.0
Haitian Creole	100.7	98.9	96.8	92.8	84.8	62.7	100.7	98.9	97.6	93.8	86.0	66.4
Lao	98.7	99.2	96.4	91.9	86.1	60.3	98.7	99.2	96.1	93.5	85.5	62.7
Amharic	100.1	96.5	95.3	89.4	83.1	54.7	100.1	96.5	96.5	92.4	81.7	61.9
Hausa	98.9	95.1	93.1	86.7	80.3	55.5	98.9	95.1	95.4	90.8	82.4	61.1
Xhosa	99.7	98.4	94.5	89.3	79.6	57.7	99.7	98.4	97.2	91.6	83.2	61.6
Zulu	100.3	99.7	96.5	92.1	85.8	61.8	100.3	99.7	97.4	90.9	82.3	64.2
Tswana	98.7	96.9	95.6	91.3	79.0	58.7	98.7	96.9	95.3	89.2	78.1	61.6
Tigrinya	99.0	96.7	93.8	88.1	75.8	46.3	99.0	96.7	94.5	90.2	81.0	49.3
Yoruba	98.7	98.6	96.0	90.2	80.9	58.6	98.7	98.6	96.0	90.1	85.2	62.6
Wolof	100.7	100.0	98.4	91.9	83.3	68.8	100.7	100.0	99.5	95.5	87.8	72.4
T1	99.4	97.9	96.0	92.2	84.1	58.6	99.4	97.9	96.6	92.9	85.1	62.5
Norwegian	99.5	98.2	97.7	95.8	91.8	73.1	99.5	98.2	98.5	96.3	93.2	78.3
Macedonian	99.4	99.2	98.2	95.9	90.4	69.3	99.4	99.2	98.3	94.9	91.7	74.1
Albanian	100.0	98.7	97.8	95.4	89.2	66.9	100.0	98.7	96.9	94.9	89.9	69.7
Armenian	99.1	99.4	97.7	94.7	86.6	62.0	99.1	99.4	97.0	94.5	86.6	65.2
Malayalam	99.3	98.3	96.2	91.0	85.6	57.6	99.3	98.3	96.5	93.1	85.1	64.4
Gujarati	99.8	97.1	96.4	92.3	86.4	55.7	99.8	97.1	96.1	92.7	88.2	61.5
Nepali	99.6	97.8	96.3	94.5	86.2	62.5	99.6	97.8	98.0	94.2	87.5	66.1
Javanese	100.0	97.9	96.2	90.8	82.6	60.2	100.0	97.9	96.2	91.6	82.5	62.7
Telugu	99.9	98.7	96.1	94.9	87.8	54.9	99.9	98.7	98.4	94.4	89.4	59.3
Tajik	99.0	97.7	96.1	91.5	80.7	54.8	99.0	97.7	96.3	92.2	84.6	61.4
Sinhala	100.3	98.4	96.8	94.5	84.8	57.7	100.3	98.4	97.8	93.7	87.8	58.9
Kannada	100.8	97.7	95.7	94.6	87.6	57.3	100.8	97.7	96.8	94.3	87.0	65.1
Khmer	99.2	99.1	97.7	93.0	86.8	60.7	99.2	99.1	96.5	95.1	87.1	65.2
Kurdish	99.2	95.9	94.3	91.1	80.2	53.3	99.2	95.9	94.7	91.8	82.0	56.4
Sundanese	98.3	97.9	95.5	92.6	81.7	53.4	98.3	97.9	96.1	91.9	83.7	59.5
Oriya	99.4	97.8	96.1	93.0	82.9	49.8	99.4	97.8	97.8	95.4	84.7	55.6
Azerbaijani	98.7	97.8	96.9	94.5	85.0	61.4	98.7	97.8	97.6	95.6	86.4	60.7
Mongolian	98.9	98.7	96.5	90.6	82.4	60.5	98.9	98.7	95.9	92.5	86.1	61.8
Assamese	98.7	97.1	96.8	93.9	84.6	50.7	98.7	97.1	98.1	94.8	86.8	61.1
Kyrgyz	99.5	98.9	98.0	93.6	87.2	60.1	99.5	98.9	99.5	95.1	87.3	61.0
Pashto	100.8	98.3	94.7	90.7	81.8	56.5	100.8	98.3	95.6	90.8	81.5	56.1
Burmese	99.1	98.1	95.4	92.2	82.7	53.4	99.1	98.1	97.8	92.9	85.1	56.4
Malagasy	99.3	97.8	95.8	90.7	84.7	58.1	99.3	97.8	96.6	92.5	86.0	61.8
Southern Sotho	98.7	95.9	93.8	90.4	80.3	54.8	98.7	95.9	92.5	90.0	82.8	59.6
Kinyarwanda	98.7	98.8	95.9	92.5	81.0	50.0	98.7	98.8	96.7	91.2	82.8	57.2
Somali	100.4	97.5	96.7	90.5	84.1	59.2	100.4	97.5	97.4	91.6	86.7	64.7
Shona	98.6	96.8	93.1	92.1	84.4	58.0	98.6	96.8	95.1	92.2	84.3	62.2
Chichewa	99.4	96.4	94.2	90.9	85.2	61.5	99.4	96.4	94.1	90.8	81.0	65.9
Igbo	99.2	96.5	92.7	90.0	80.5	59.2	99.2	96.5	97.1	90.6	77.9	65.4
Maori	98.3	96.9	93.7	89.9	84.5	61.8	98.3	96.9	95.0	90.7	79.9	64.0
Swati	100.7	99.0	97.6	90.6	78.0	53.3	100.7	99.0	99.0	91.8	81.6	56.7
Guarani	100.3	98.2	94.2	90.7	83.2	59.4	100.3	98.2	94.3	90.9	83.5	58.4
Oromo	99.5	100.4	97.4	92.8	87.3	58.8	99.5	100.4	97.5	95.4	84.8	60.6
Lingala	99.0	99.6	98.1	93.7	84.7	63.2	99.0	99.6	97.1	92.5	86.1	69.9
Luganda	101.4	97.6	96.9	91.4	80.5	62.5	101.4	97.6	97.8	95.0	84.0	62.3
Tibetan	97.3	95.4	93.1	87.2	77.2	53.2	97.3	95.4	94.2	90.3	77.4	54.8
Bambara	99.6	96.3	93.6	86.7	81.4	64.5	99.6	96.3	92.4	90.4	84.5	67.8
Avg	99.6	98.5	96.8	93.7	86.9	64.0	99.6	98.5	97.3	94.3	87.8	67.5

Table 23: Retention (%) at $r{=}7..2$ for Gemma4 under top-k reduction and global α. Averaged across Belebele and Translation. Per-language mean across available tasks, then mean across languages within tier. Part 2: tiers T2–T1 and overall mean.

M Detailed Results: Belebele

Qwen3 – Belebele																						
	Standard							Global α							Token α							
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	89.1	79.4	89.1	57.0	63.8	25.3	28.4	89.1	86.7	97.3	84.9	95.2	74.4	83.4	89.1	86.8	97.4	84.7	95.1	77.3	86.7
English	91.8	85.9	93.6	67.1	73.1	26.2	28.6	91.8	89.6	97.6	89.4	97.5	83.3	90.8	91.8	90.3	98.4	89.4	97.5	86.0	93.7
French	90.6	77.8	85.9	57.6	63.6	25.1	27.7	90.6	89.1	98.4	87.1	96.2	74.8	82.6	90.6	89.6	98.9	86.4	95.5	78.6	86.7
German	90.1	80.7	89.5	59.4	66.0	24.7	27.4	90.1	87.1	96.7	86.1	95.6	74.0	82.1	90.1	86.6	96.1	84.4	93.7	77.0	85.5
Chinese (S.)	89.7	81.0	90.3	63.0	70.3	26.3	29.4	89.7	87.0	97.0	85.4	95.3	77.4	86.4	89.7	86.6	96.5	85.3	95.2	80.1	89.3
Spanish	88.3	79.9	90.4	56.2	63.6	22.0	24.9	88.3	86.9	98.4	83.6	94.6	72.8	82.4	88.3	86.4	97.9	84.2	95.3	74.9	84.8
Arabic	87.7	75.9	86.6	46.3	52.9	27.0	30.8	87.7	84.4	96.3	83.1	94.8	68.7	78.3	87.7	85.3	97.3	83.0	94.7	72.6	82.8
Japanese	85.7	74.6	87.0	49.2	57.5	25.7	30.0	85.7	82.8	96.6	79.3	92.6	69.7	81.3	85.7	82.9	96.8	80.3	93.8	71.9	83.9
T4	85.7	70.6	82.2	46.7	54.4	24.6	28.8	85.7	81.7	95.2	78.1	91.0	62.9	73.2	85.7	81.4	94.9	78.7	91.7	66.4	77.3
Portuguese	90.0	76.4	84.9	52.4	58.3	25.3	28.1	90.0	87.8	97.5	85.2	94.7	75.2	83.6	90.0	86.8	96.4	86.4	96.0	76.7	85.2
Catalan	89.1	78.1	87.7	52.9	59.4	23.4	26.3	89.1	86.1	96.6	83.7	93.9	71.7	80.4	89.1	86.2	96.8	82.3	92.4	73.7	82.7
Italian	88.7	83.1	93.7	61.6	69.4	23.0	25.9	88.7	86.2	97.2	83.9	94.6	74.6	84.1	88.7	86.0	97.0	83.7	94.4	78.9	89.0
Dutch	88.7	76.0	85.7	49.9	56.3	23.6	26.6	88.7	85.3	96.2	82.7	93.2	68.1	76.8	88.7	85.7	96.6	83.7	94.4	72.1	81.3
Croatian	88.6	74.1	83.7	46.0	51.9	22.0	24.8	88.6	82.2	92.8	78.2	88.3	61.7	69.6	88.6	82.8	93.5	78.2	88.3	66.3	74.9
Russian	88.6	77.7	87.7	52.8	59.6	25.2	28.5	88.6	86.2	97.4	82.0	92.6	69.8	78.8	88.6	86.2	97.4	83.0	93.7	76.0	85.8
Swedish	88.2	76.6	86.8	50.2	56.9	26.7	30.2	88.2	85.9	97.4	82.3	93.3	66.6	75.4	88.2	85.6	97.0	82.2	93.2	68.7	77.8
Vietnamese	88.0	75.3	85.6	50.7	57.6	26.1	29.7	88.0	84.3	95.8	82.8	94.1	71.8	81.6	88.0	84.6	96.1	83.0	94.3	74.2	84.3
Czech	86.9	74.1	85.3	47.0	54.1	25.9	29.8	86.9	84.0	96.7	80.1	92.2	62.8	72.3	86.9	82.9	95.4	81.8	94.1	65.9	75.8
Polish	86.8	69.9	80.5	48.0	55.3	23.0	26.5	86.8	82.1	94.6	76.3	88.0	61.6	70.9	86.8	81.7	94.1	79.1	91.2	64.8	74.6
Serbian	86.6	67.8	78.3	38.2	44.2	23.9	27.6	86.6	81.9	94.6	77.1	89.1	61.1	70.6	86.6	82.1	94.9	78.2	90.4	64.1	74.1
Korean	86.3	74.9	86.7	51.7	59.8	26.2	30.4	86.3	83.6	96.8	81.0	93.8	67.3	78.0	86.3	83.9	97.2	80.8	93.6	71.9	83.3
Hungarian	85.2	64.7	75.9	41.9	49.2	24.9	29.2	85.2	80.6	94.5	77.2	90.6	56.1	65.8	85.2	80.9	94.9	79.3	93.1	58.3	68.4
Finnish	84.0	65.1	77.5	41.2	49.1	24.3	29.0	84.0	80.1	95.4	74.3	88.5	55.9	66.5	84.0	79.4	94.6	75.1	89.4	60.9	72.5
Turkish	83.0	63.4	76.4	39.4	47.5	23.9	28.8	83.0	78.4	94.5	75.3	90.8	57.7	69.5	83.0	79.1	95.3	74.7	90.0	60.3	72.7
Persian	82.7	59.4	71.9	40.9	49.5	24.8	30.0	82.7	79.0	95.6	73.6	89.0	55.9	67.6	82.7	77.4	93.7	74.4	90.1	57.4	69.5
Basque	75.7	55.3	73.1	34.8	46.0	24.0	31.7	75.7	65.8	86.9	62.4	82.5	42.3	55.9	75.7	63.6	84.0	62.6	82.7	49.8	65.8
Hindi	75.3	58.9	78.2	41.2	54.7	26.2	34.8	75.3	70.3	93.4	66.9	88.8	52.4	69.6	75.3	70.7	93.8	67.3	89.4	54.9	72.9
T3	83.5	66.7	79.7	42.2	50.5	24.9	30.0	83.5	78.0	93.4	73.1	87.4	55.7	66.4	83.5	78.1	93.4	73.5	87.8	58.0	69.1
Romanian	88.4	74.4	84.2	49.1	55.5	22.3	25.3	88.4	84.1	95.1	80.8	91.3	64.7	73.1	88.4	84.7	95.7	81.0	91.6	68.2	77.1
Danish	88.2	73.7	83.5	46.9	53.1	25.3	28.7	88.2	84.0	95.2	80.1	90.8	64.4	73.0	88.2	83.6	94.7	81.1	91.9	67.2	76.2
Indonesian	88.2	72.2	81.9	49.1	55.7	25.6	29.0	88.2	85.9	97.4	81.4	92.3	71.8	81.4	88.2	85.1	96.5	82.1	93.1	70.8	80.2
Slovak	87.9	71.8	81.7	45.2	51.5	23.4	26.7	87.9	82.9	94.3	79.8	90.8	60.6	68.9	87.9	83.7	95.2	80.6	91.7	65.2	74.2
Bulgarian	87.8	73.3	83.5	48.0	54.7	26.6	30.3	87.8	83.9	95.6	80.7	91.9	64.6	73.5	87.8	83.4	95.1	79.8	90.9	67.2	76.6
Afrikaans	87.2	72.2	82.8	44.1	50.6	26.7	30.6	87.2	80.0	91.7	76.4	87.6	57.9	66.4	87.2	81.3	93.2	76.7	87.9	57.7	66.1
Slovenian	87.2	69.4	79.6	37.9	43.4	23.8	27.3	87.2	80.9	92.7	75.1	86.1	54.3	62.3	87.2	81.4	93.4	77.0	88.3	59.1	67.8
Greek	87.2	72.9	83.6	48.4	55.5	25.1	28.8	87.2	83.0	95.2	79.3	91.0	61.4	70.4	87.2	84.1	96.4	81.7	93.6	61.6	70.6
Latvian	86.7	66.1	76.3	41.3	47.7	25.1	29.0	86.7	80.2	92.6	75.8	87.4	53.7	61.9	86.7	80.9	93.3	75.7	87.3	57.8	66.7
Ukrainian	86.7	73.3	84.6	45.1	52.1	25.1	29.0	86.7	83.0	95.8	78.4	90.5	62.1	71.7	86.7	83.0	95.8	80.1	92.4	68.0	78.5
Lithuanian	86.2	62.9	72.9	39.9	46.3	24.0	27.8	86.2	79.2	91.9	74.0	85.8	57.1	66.2	86.2	79.7	92.4	74.3	86.2	58.1	67.4
Malay	86.1	70.6	81.9	51.4	59.7	23.1	26.8	86.1	83.1	96.5	79.3	92.1	65.3	75.9	86.1	82.4	95.7	80.4	93.4	67.6	78.5
Filipino	84.8	68.7	81.0	42.3	49.9	23.8	28.0	84.8	80.2	94.6	74.8	88.2	53.0	62.5	84.8	80.3	94.8	74.6	87.9	56.0	66.1
Hebrew	84.1	64.6	76.8	38.4	45.7	25.0	29.7	84.1	80.9	96.2	74.3	88.4	59.2	70.4	84.1	81.2	96.6	75.1	89.3	61.7	73.3
Thai	83.7	74.3	88.8	47.9	57.2	26.1	31.2	83.7	81.7	97.6	77.7	92.8	65.8	78.6	83.7	80.6	96.3	77.6	92.7	68.3	81.7
Estonian	82.2	64.4	78.4	40.3	49.1	24.1	29.3	82.2	74.2	90.3	68.6	83.4	50.7	61.6	82.2	73.9	89.9	68.2	83.0	53.3	64.9
Bengali	79.8	61.0	76.5	37.1	46.5	25.7	32.2	79.8	74.4	93.3	68.1	85.4	49.9	62.5	79.8	73.6	92.2	67.3	84.4	51.4	64.5
Cebuano	79.2	56.2	71.0	40.4	51.1	24.2	30.6	79.2	71.4	90.2	62.2	78.5	41.8	52.7	79.2	71.7	90.5	63.3	79.9	46.7	58.9
Uzbek	78.2	61.4	78.6	38.7	49.4	26.9	34.4	78.2	68.6	87.6	64.3	82.2	49.2	62.9	78.2	69.3	88.6	62.2	79.5	45.2	57.8
Urdu	78.0	62.2	79.8	37.3	47.9	25.8	33.0	78.0	73.0	93.6	67.0	85.9	47.6	61.0	78.0	73.4	94.2	66.2	84.9	48.9	62.7
Tamil	75.1	56.1	74.7	34.8	46.3	23.7	31.5	75.1	67.6	89.9	63.7	84.8	45.8	60.9	75.1	67.4	89.8	62.6	83.3	45.3	60.4
Georgian	74.6	56.4	75.7	33.4	44.9	25.6	34.3	74.6	69.0	92.5	59.4	79.7	37.2	49.9	74.6	68.1	91.4	59.8	80.2	41.7	55.9
Kazakh	72.6	54.9	75.7	33.9	46.7	26.6	36.6	72.6	63.9	88.1	60.4	83.3	43.3	59.7	72.6	63.1	87.0	62.3	85.9	46.4	64.0

Table 24: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Qwen3 on Belebele under top-k reduction, global α, and token α. Per-language mean within tier. Qwen3 is restricted to its 73 supported languages. Part 1: tiers T5–T3.

Qwen3 – Belebele																						
	Standard							Global α							Token α							
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T2	68.0	49.2	72.3	32.8	48.7	25.1	37.3	68.0	58.8	86.2	52.5	76.8	38.0	55.9	68.0	57.8	84.8	51.7	75.7	39.4	57.8
Icelandic	77.7	59.1	76.1	35.2	45.4	25.7	33.0	77.7	70.1	90.3	62.3	80.3	43.2	55.7	77.7	68.3	88.0	58.9	75.8	46.1	59.4
Marathi	77.2	56.2	72.8	33.9	43.9	26.8	34.7	77.2	68.9	89.2	62.6	81.0	44.2	57.3	77.2	66.8	86.5	61.8	80.0	46.6	60.3
Punjabi	73.3	49.6	67.6	33.2	45.3	24.2	33.0	73.3	64.7	88.2	61.7	84.1	41.6	56.7	73.3	62.1	84.7	57.2	78.0	43.0	58.6
Lao	65.7	48.2	73.4	28.9	44.0	23.4	35.7	65.7	55.7	84.8	49.1	74.8	34.8	53.0	65.7	55.8	84.9	51.0	77.7	35.2	53.6
Maltese	64.0	46.6	72.7	33.0	51.6	24.1	37.7	64.0	55.2	86.3	46.9	73.3	37.0	57.8	64.0	54.2	84.7	48.1	75.2	35.1	54.9
Swahili	59.1	42.6	72.0	33.2	56.2	28.1	47.6	59.1	49.0	82.9	43.6	73.7	30.1	50.9	59.1	48.6	82.1	43.9	74.2	33.3	56.4
Haitian Creole	59.0	42.2	71.6	32.3	54.8	23.1	39.2	59.0	48.3	81.9	41.7	70.6	35.4	60.1	59.0	48.6	82.3	40.7	68.9	36.3	61.6
T1	74.1	54.5	73.4	35.4	47.8	24.8	33.8	74.1	66.3	89.2	60.3	81.1	42.7	57.3	74.1	65.6	88.3	60.1	80.8	43.8	58.8
Norwegian	88.0	73.0	83.0	43.3	49.2	24.8	28.2	88.0	84.3	95.8	80.1	91.0	63.6	72.2	88.0	84.3	95.8	80.7	91.7	66.1	75.1
Albanian	83.3	64.2	77.1	43.0	51.6	22.7	27.2	83.3	77.4	92.9	71.2	85.5	53.7	64.4	83.3	76.3	91.6	71.1	85.3	54.8	65.7
Macedonian	82.9	64.0	77.2	40.0	48.3	24.2	29.2	82.9	79.0	95.3	73.2	88.3	56.0	67.6	82.9	78.6	94.8	73.9	89.1	58.2	70.2
Armenian	79.1	57.2	72.3	37.1	46.9	24.0	30.3	79.1	72.1	91.2	63.8	80.6	43.3	54.8	79.1	70.8	89.5	64.7	81.7	44.1	55.8
Javanese	77.1	54.6	70.7	37.4	48.6	25.8	33.4	77.1	70.8	91.8	63.4	82.3	44.9	58.2	77.1	72.0	93.4	62.3	80.8	42.0	54.5
Malayalam	77.1	54.4	70.6	37.1	48.1	25.6	33.1	77.1	66.8	86.6	62.1	80.5	45.3	58.8	77.1	65.9	85.4	62.7	81.3	44.4	57.6
Kannada	75.6	51.2	67.8	31.8	42.1	23.2	30.7	75.6	66.1	87.5	62.4	82.6	38.6	51.0	75.6	63.9	84.6	62.2	82.4	41.6	55.0
Azerbaijani	75.0	55.0	73.3	35.2	47.0	24.6	32.7	75.0	68.4	91.3	61.2	81.6	42.1	56.1	75.0	68.7	91.6	61.6	82.1	42.9	57.2
Gujarati	74.3	54.4	73.2	37.1	49.9	23.7	31.8	74.3	65.7	88.3	62.6	84.2	42.8	57.5	74.3	64.7	87.0	61.9	83.3	45.0	60.5
Oriya	73.3	49.7	67.7	32.1	43.8	25.3	34.5	73.3	60.8	82.9	55.0	75.0	37.6	51.2	73.3	59.1	80.6	55.0	75.0	40.6	55.3
Nepali	73.1	55.8	76.3	34.4	47.1	27.6	37.7	73.1	66.4	90.9	58.4	79.9	39.6	54.1	73.1	64.7	88.4	57.8	79.0	42.3	57.9
Khmer	71.8	54.8	76.3	33.8	47.1	23.3	32.5	71.8	64.8	90.2	56.6	78.8	40.3	56.2	71.8	64.1	89.3	57.1	79.6	41.9	58.4
Sundanese	71.3	52.0	72.9	36.6	51.2	23.8	33.3	71.3	63.3	88.8	55.6	77.9	37.9	53.1	71.3	64.3	90.2	53.9	75.5	40.6	56.9
Telugu	70.2	51.6	73.4	31.2	44.5	26.2	37.3	70.2	62.3	88.8	57.8	82.3	36.6	52.1	70.2	60.0	85.4	58.1	82.8	37.8	53.8
Assamese	69.4	49.8	71.7	33.0	47.5	26.8	38.6	69.4	60.1	86.6	54.7	78.7	37.1	53.4	69.4	60.7	87.4	51.3	73.9	37.6	54.1
Sinhala	64.9	46.0	70.9	29.7	45.7	24.0	37.0	64.9	53.3	82.2	49.2	75.9	33.3	51.4	64.9	54.4	83.9	49.3	76.0	35.4	54.6
Tajik	63.6	45.9	72.2	32.8	51.6	25.9	40.7	63.6	53.4	84.1	48.2	75.9	37.6	59.1	63.6	52.2	82.2	46.4	73.1	36.0	56.6
Burmese	63.2	46.6	73.6	31.3	49.6	25.0	39.5	63.2	57.6	91.0	50.1	79.3	38.1	60.3	63.2	55.7	88.0	51.4	81.4	37.1	58.7
Avg	80.8	64.2	78.9	42.2	51.9	24.9	31.2	80.8	75.0	92.5	70.3	86.5	54.4	66.5	80.8	74.7	92.1	70.4	86.6	56.6	69.2

Table 25: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4,3,2\}$ for Qwen3 on Belebele under top-k reduction, global α, and token α. Per-language mean within tier. Qwen3 is restricted to its 73 supported languages. Part 2: tiers T2–T1 and overall mean.

Qwen3.5 – Belebele																					
	Standard							Global α							Token α						
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	91.5	86.7	94.7	78.0	85.3	51.8	56.6	91.5	91.4	99.9	90.9	99.4	89.1	97.4	91.5	91.4	100.0	91.1	99.6	88.8	97.1
English	95.4	93.0	97.4	85.9	90.0	62.4	65.4	95.4	95.3	99.9	95.1	99.7	94.6	99.1	95.4	95.3	99.9	94.8	99.3	94.1	98.6
French	92.1	88.4	96.0	81.3	88.3	50.8	55.1	92.1	92.2	100.1	92.3	100.2	90.2	97.9	92.1	92.7	100.6	92.8	100.7	90.3	98.1
Spanish	91.9	87.0	94.7	76.1	82.8	48.9	53.2	91.9	91.2	99.3	91.4	99.5	89.6	97.5	91.9	91.8	99.9	90.3	98.3	89.0	96.9
German	91.8	86.9	94.7	77.8	84.7	50.0	54.5	91.8	91.8	100.0	91.0	99.2	89.0	97.0	91.8	92.2	100.5	91.8	100.0	89.6	97.6
Chinese (S.)	91.7	89.1	97.2	79.1	86.3	54.4	59.4	91.7	91.7	100.0	90.7	98.9	90.0	98.2	91.7	91.6	99.9	91.6	99.9	90.2	98.4
Arabic	89.8	78.6	87.5	70.1	78.1	47.4	52.8	89.8	90.1	100.4	89.4	99.6	85.6	95.3	89.8	88.6	98.6	89.6	99.8	84.7	94.3
Japanese	87.6	83.8	95.7	75.8	86.5	48.7	55.6	87.6	87.6	100.0	86.3	98.6	84.9	97.0	87.6	87.9	100.4	86.9	99.2	84.0	95.9
T4	89.7	85.1	94.9	75.8	84.5	45.6	50.8	89.7	89.2	99.4	88.3	98.4	85.3	95.1	89.7	89.1	99.3	88.6	98.7	85.5	95.3
Portuguese	92.6	87.0	94.0	78.2	84.5	47.6	51.4	92.6	91.7	99.0	92.2	99.6	89.9	97.1	92.6	91.6	98.9	92.0	99.4	89.3	96.5
Russian	92.3	88.1	95.4	80.3	87.0	50.9	55.1	92.3	90.8	98.3	90.6	98.1	88.3	95.7	92.3	91.3	98.9	90.6	98.1	89.0	96.4
Swedish	91.3	84.7	92.7	79.2	86.7	45.4	49.8	91.3	89.8	98.3	90.2	98.8	88.0	96.4	91.3	90.4	99.0	89.3	97.8	87.1	95.4
Czech	91.0	86.0	94.5	80.3	88.3	46.6	51.2	91.0	90.6	99.5	89.9	98.8	87.1	95.7	91.0	90.0	98.9	89.7	98.5	86.9	95.5
Italian	90.9	87.4	96.2	78.9	86.8	51.7	56.8	90.9	90.8	99.9	90.3	99.4	86.9	95.6	90.9	90.4	99.5	89.6	98.5	87.8	96.6
Catalan	90.7	86.9	95.8	76.3	84.2	46.2	51.0	90.7	90.7	100.0	88.4	97.5	83.4	92.0	90.7	90.2	99.5	89.8	99.0	86.7	95.6
Serbian	90.6	87.3	96.4	72.8	80.4	43.0	47.5	90.6	89.9	99.3	88.6	97.8	84.8	93.6	90.6	88.9	98.2	88.6	97.8	85.1	94.0
Hungarian	90.2	84.1	93.2	71.0	78.7	40.0	44.3	90.2	89.3	99.0	88.6	98.2	85.2	94.5	90.2	90.1	99.9	88.8	98.4	85.1	94.3
Vietnamese	90.1	85.9	95.3	76.1	84.5	52.2	58.0	90.1	89.9	99.8	89.2	99.0	87.2	96.8	90.1	89.6	99.4	90.4	100.4	87.6	97.2
Dutch	90.0	85.2	94.7	77.4	86.0	45.4	50.5	90.0	90.2	100.2	88.3	98.1	85.9	95.4	90.0	89.9	99.9	88.6	98.4	85.7	95.2
Finnish	89.9	86.0	95.7	76.2	84.8	42.2	47.0	89.9	90.6	100.7	89.2	99.3	85.1	94.7	89.9	90.1	100.2	89.9	100.0	85.0	94.6
Korean	89.7	86.8	96.8	78.2	87.2	51.6	57.5	89.7	90.3	100.7	88.9	99.1	86.6	96.5	89.7	89.7	100.0	88.7	98.9	85.8	95.7
Persian	89.1	85.3	95.8	75.7	84.9	43.8	49.1	89.1	88.3	99.1	87.2	97.9	86.4	97.0	89.1	88.6	99.4	88.1	98.9	85.8	96.3
Polish	89.0	84.4	94.9	75.9	85.3	45.7	51.3	89.0	88.2	99.1	86.9	97.6	85.6	96.1	89.0	88.4	99.4	87.4	98.3	84.7	95.1
Turkish	88.8	82.9	93.4	73.8	83.1	44.7	50.3	88.8	87.8	98.9	87.1	98.1	83.4	94.0	88.8	87.6	98.6	86.3	97.2	83.9	94.5
Croatian	87.9	85.1	96.8	77.2	87.9	45.2	51.5	87.9	89.6	101.9	89.6	101.9	87.9	100.0	87.9	89.1	101.4	90.0	102.4	88.2	100.4
Basque	87.7	81.2	92.6	68.7	78.3	37.4	42.7	87.7	86.3	98.5	83.7	95.4	78.2	89.2	87.7	86.2	98.4	85.4	97.5	80.1	91.4
Hindi	83.3	77.6	93.1	68.7	82.4	40.8	48.9	83.3	81.3	97.6	80.4	96.5	76.2	91.5	83.3	81.4	97.7	80.9	97.1	75.8	90.9
T3	88.2	83.0	94.1	72.9	82.6	42.0	47.6	88.2	87.6	99.3	86.3	97.9	82.0	92.9	88.2	87.6	99.3	86.5	98.0	82.4	93.4
Afrikaans	92.0	85.9	93.4	76.2	82.9	39.0	42.4	92.0	91.3	99.3	90.9	98.8	86.0	93.5	92.0	91.0	98.9	90.7	98.6	85.7	93.1
Danish	91.6	84.3	92.1	76.8	83.9	47.3	51.7	91.6	91.7	100.1	90.3	98.7	87.4	95.5	91.6	91.4	99.9	90.1	98.4	88.6	96.7
Slovak	90.9	85.8	94.4	76.3	84.0	43.3	47.7	90.9	89.7	98.7	89.1	98.0	85.3	93.9	90.9	89.0	97.9	89.1	98.0	85.8	94.4
Slovenian	90.4	86.0	95.1	74.0	81.8	43.0	47.5	90.4	90.0	99.5	89.3	98.8	84.4	93.4	90.4	89.9	99.4	88.1	97.4	83.4	92.3
Malay	90.4	86.9	96.1	75.6	83.5	45.7	50.5	90.4	90.4	100.0	89.0	98.4	86.0	95.1	90.4	90.6	100.1	89.3	98.8	86.9	96.1
Bulgarian	90.2	87.0	96.4	77.6	86.0	48.2	53.4	90.2	90.7	100.5	89.8	99.5	87.4	96.9	90.2	91.2	101.1	90.3	100.1	86.6	95.9
Latvian	90.2	84.3	93.5	74.9	83.0	38.4	42.6	90.2	89.4	99.1	88.4	98.0	83.3	92.4	90.2	89.1	98.8	88.2	97.8	83.9	93.0
Romanian	90.2	85.3	94.6	76.4	84.7	47.8	53.0	90.2	90.2	100.0	88.6	98.2	85.2	94.5	90.2	88.9	98.5	88.9	98.5	84.3	93.5
Indonesian	90.1	85.7	95.1	77.6	86.1	48.6	53.9	90.1	89.6	99.4	89.2	99.0	87.6	97.2	90.1	90.2	100.1	89.0	98.8	87.3	96.9
Ukrainian	90.1	88.0	97.7	77.1	85.6	45.3	50.3	90.1	89.9	99.8	88.8	98.5	85.0	94.3	90.1	89.9	99.8	89.4	99.3	85.9	95.3
Greek	89.8	85.4	95.2	76.9	85.6	44.7	49.8	89.8	90.0	100.2	87.6	97.5	82.1	91.5	89.8	89.6	99.8	89.1	99.3	84.3	93.9
Filipino	89.7	83.4	93.1	71.9	80.2	40.8	45.5	89.7	88.8	99.0	87.0	97.0	82.7	92.2	89.7	88.1	98.3	87.7	97.8	82.4	91.9
Lithuanian	89.0	85.4	96.0	73.8	82.9	40.8	45.8	89.0	90.3	101.5	89.3	100.4	85.0	95.5	89.0	89.7	100.7	89.0	100.0	84.4	94.9
Estonian	88.9	82.7	93.0	70.4	79.2	34.8	39.1	88.9	88.7	99.8	87.9	98.9	81.1	91.3	88.9	88.2	99.2	87.7	98.6	81.8	92.0
Uzbek	88.2	80.1	90.8	69.0	78.2	39.0	44.2	88.2	86.2	97.7	85.1	96.5	79.6	90.2	88.2	87.2	98.9	84.2	95.5	78.4	88.9
Hebrew	86.8	82.6	95.1	74.4	85.8	42.0	48.4	86.8	86.9	100.1	85.8	98.8	83.8	96.5	86.8	87.2	100.5	86.2	99.4	83.7	96.4
Georgian	86.3	80.1	92.8	66.8	77.3	35.6	41.2	86.3	85.6	99.1	83.6	96.8	76.8	88.9	86.3	85.2	98.7	82.9	96.0	78.6	91.0
Thai	85.4	82.6	96.6	78.0	91.3	50.8	59.4	85.4	84.8	99.2	83.6	97.8	81.0	94.8	85.4	84.9	99.3	83.3	97.5	81.6	95.4
Bengali	84.9	78.7	92.7	69.2	81.5	40.7	47.9	84.9	83.4	98.3	81.8	96.3	75.8	89.3	84.9	84.4	99.5	82.7	97.4	78.6	92.5
Urdu	84.8	77.7	91.6	62.3	73.5	33.7	39.7	84.8	82.6	97.4	82.9	97.8	78.4	92.5	84.8	83.8	98.8	82.7	97.5	78.2	92.3
Kazakh	84.2	77.8	92.3	67.3	79.9	38.6	45.8	84.2	82.2	97.6	80.2	95.3	75.6	89.7	84.2	82.4	97.9	80.6	95.6	76.2	90.5
Cebuano	82.0	76.2	93.0	66.2	80.8	36.8	44.9	82.0	81.2	99.1	78.0	95.1	73.2	89.3	82.0	80.4	98.1	79.1	96.5	73.7	89.8
Tamil	82.0	76.8	93.6	67.4	82.2	40.8	49.7	82.0	80.3	98.0	79.8	97.3	72.7	88.6	82.0	81.3	99.2	80.4	98.1	74.8	91.2

Table 26: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4,3,2\}$ for Qwen3.5 on Belebele under top-k reduction, global α, and token α. Per-language mean within tier. Part 1: tiers T5–T3.

	Qwen3.5 – Belebele																					
	Standard							Global α							Token α							
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T2	73.6	65.9	89.1	54.1	73.0	31.7	43.7	73.6	71.9	97.4	69.3	93.6	62.4	84.1	73.6	72.1	97.6	70.0	94.8	63.6	85.7
Maltese	87.9	80.7	91.8	68.6	78.0	35.2	40.1	87.9	85.7	97.5	83.4	94.9	75.1	85.5	87.9	86.3	98.2	84.0	95.6	78.2	89.0
Swahili	87.0	78.2	89.9	67.1	77.1	37.7	43.3	87.0	84.0	96.6	82.7	95.0	76.2	87.6	87.0	84.8	97.4	82.8	95.1	77.9	89.5
Icelandic	85.8	82.4	96.1	69.8	81.3	34.6	40.3	85.8	86.4	100.8	84.8	98.8	79.4	92.6	85.8	86.6	100.9	85.8	100.0	80.9	94.3
Marathi	84.6	79.3	93.8	70.3	83.2	37.8	44.7	84.6	84.2	99.6	81.4	96.3	77.3	91.5	84.6	83.7	98.9	81.4	96.3	76.8	90.8
Punjabi	83.0	77.0	92.8	67.2	81.0	36.3	43.8	83.0	83.1	100.1	81.4	98.1	76.4	92.1	83.0	82.9	99.9	81.7	98.4	76.7	92.4
Haitian Creole	81.8	76.9	94.0	59.8	73.1	32.8	40.1	81.8	80.7	98.6	79.3	97.0	68.2	83.4	81.8	81.8	100.0	78.9	96.5	70.7	86.4
Lao	76.8	68.2	88.9	53.3	69.5	33.1	43.1	76.8	75.0	97.7	73.2	95.4	64.9	84.5	76.8	75.0	97.7	72.8	94.8	65.6	85.4
Amharic	76.2	65.1	85.4	52.3	68.7	30.7	40.2	76.2	76.2	100.0	70.7	92.7	62.2	81.6	76.2	74.8	98.1	72.6	95.2	65.2	85.6
Xhosa	71.3	63.8	89.4	48.8	68.4	29.6	41.4	71.3	69.3	97.2	65.1	91.3	59.1	82.9	71.3	69.1	96.9	66.0	92.5	58.6	82.1
Hausa	69.4	58.2	83.8	46.6	67.0	28.6	41.1	69.4	67.7	97.4	63.1	90.9	55.2	79.5	69.4	66.4	95.7	64.6	93.0	57.3	82.6
Tigrinya	68.3	60.1	88.0	45.0	65.9	28.7	42.0	68.3	65.9	96.4	63.2	92.5	53.2	77.9	68.3	66.2	96.9	64.1	93.8	56.1	82.1
Zulu	68.2	59.9	87.8	48.0	70.4	30.7	45.0	68.2	66.1	96.9	65.0	95.3	58.3	85.5	68.2	66.8	97.9	64.7	94.8	59.7	87.5
Tswana	63.4	54.2	85.5	42.0	66.2	27.1	42.7	63.4	61.2	96.5	56.7	89.3	51.7	81.4	63.4	62.1	97.9	59.2	93.3	53.4	84.2
Yoruba	56.0	46.9	83.7	37.0	66.1	28.9	51.6	56.0	53.1	94.8	51.3	91.7	43.9	78.4	56.0	53.2	95.0	51.9	92.7	43.8	78.2
Wolof	44.3	38.0	85.7	35.2	79.4	24.7	55.6	44.3	40.1	90.5	37.9	85.5	34.3	77.4	44.3	41.2	93.0	39.8	89.7	33.8	76.2
T1	73.4	66.1	89.6	54.9	74.2	32.4	44.9	73.4	71.4	97.1	68.9	93.2	62.7	84.5	73.4	71.7	97.4	69.8	94.7	63.8	86.2
Norwegian	91.2	84.2	92.3	72.9	79.9	43.3	47.5	91.2	90.7	99.4	89.7	98.3	85.4	93.7	91.2	90.4	99.1	89.1	97.7	86.7	95.0
Albanian	89.4	82.1	91.8	71.2	79.6	39.1	43.7	89.4	88.7	99.1	87.2	97.5	81.3	90.9	89.4	88.2	98.6	86.7	96.9	82.3	92.0
Macedonian	89.2	84.0	94.1	72.8	81.6	41.0	46.0	89.2	88.6	99.3	87.8	98.4	83.4	93.5	89.2	88.1	98.8	88.0	98.6	85.0	95.3
Kyrgyz	86.2	77.2	89.6	66.7	77.3	36.1	41.9	86.2	83.8	97.2	82.2	95.4	76.7	88.9	86.2	84.2	97.7	83.1	96.4	77.9	90.3
Armenian	85.9	79.3	92.4	66.2	77.1	37.2	43.3	85.9	85.4	99.5	84.4	98.3	80.3	93.5	85.9	85.9	100.0	84.6	98.4	79.3	92.4
Oriya	85.4	78.4	91.8	65.0	76.1	36.7	42.9	85.4	83.3	97.5	81.7	95.6	75.1	87.9	85.4	83.3	97.5	81.6	95.4	74.7	87.4
Malayalam	84.7	81.0	95.7	69.8	82.4	38.9	45.9	84.7	84.4	99.7	83.0	98.0	77.7	91.7	84.7	84.8	100.1	83.1	98.2	80.0	94.5
Kannada	84.1	77.1	91.7	64.1	76.2	37.9	45.0	84.1	82.1	97.6	80.9	96.2	73.1	86.9	84.1	82.9	98.5	81.2	96.6	75.4	89.7
Sinhala	83.7	77.7	92.8	68.7	82.1	37.0	44.2	83.7	82.1	98.1	80.6	96.3	76.0	90.8	83.7	83.6	99.9	80.9	96.7	78.6	93.9
Tajik	83.1	75.8	91.2	64.3	77.4	35.6	42.8	83.1	81.1	97.6	79.6	95.7	72.0	86.6	83.1	81.0	97.5	79.7	95.9	74.1	89.2
Javanese	82.8	77.7	93.8	60.7	73.3	35.7	43.1	82.8	82.7	99.9	80.9	97.7	73.7	89.0	82.8	83.0	100.3	81.1	98.0	75.9	91.7
Assamese	82.7	76.2	92.2	61.7	74.6	30.6	37.0	82.7	80.9	97.8	78.4	94.9	74.1	89.7	82.7	81.1	98.1	79.7	96.4	75.7	91.5
Nepali	82.6	77.8	94.2	66.9	81.0	38.3	46.4	82.6	82.0	99.3	80.4	97.4	75.3	91.3	82.6	82.9	100.4	81.4	98.7	74.1	89.8
Azerbaijani	82.0	75.9	92.5	63.3	77.2	34.8	42.4	82.0	79.9	97.4	77.2	94.2	69.2	84.4	82.0	79.3	96.7	78.2	95.4	69.3	84.6
Sundanese	82.0	72.9	88.9	58.1	70.9	31.2	38.1	82.0	78.3	95.5	77.2	94.2	72.4	88.3	82.0	80.2	97.8	77.7	94.7	72.7	88.6
Kurdish	81.9	73.1	89.3	57.8	70.6	30.2	36.9	81.9	79.3	96.9	75.9	92.7	68.6	83.7	81.9	79.3	96.9	77.4	94.6	68.2	83.3
Gujarati	81.9	74.3	90.8	63.8	77.9	38.6	47.1	81.9	80.7	98.5	78.3	95.7	70.2	85.8	81.9	80.9	98.8	78.7	96.1	70.8	86.4
Burmese	81.8	74.1	90.6	66.2	81.0	36.0	44.0	81.8	80.3	98.2	78.8	96.3	74.0	90.5	81.8	79.8	97.6	80.8	98.8	73.6	89.9
Telugu	80.7	73.0	90.5	62.2	77.1	37.9	47.0	80.7	77.9	96.6	76.2	94.5	70.0	86.8	80.7	78.8	97.7	76.7	95.0	72.4	89.8
Khmer	80.2	74.8	93.2	64.0	79.8	34.8	43.4	80.2	78.8	98.2	78.3	97.6	73.4	91.6	80.2	79.7	99.3	79.0	98.5	75.4	94.0
Pashto	79.0	72.1	91.3	60.7	76.8	31.4	39.8	79.0	77.4	98.0	73.9	93.5	67.8	85.8	79.0	77.8	98.5	74.3	94.1	66.8	84.5
Mongolian	78.2	71.8	91.8	58.1	74.3	30.6	39.1	78.2	77.7	99.3	75.9	97.0	67.2	85.9	78.2	76.2	97.4	75.6	96.6	69.2	88.5
Malagasy	77.3	67.7	87.5	53.8	69.5	29.0	37.5	77.3	73.8	95.4	68.3	88.4	62.4	80.7	77.3	74.1	95.8	71.8	92.8	64.8	83.8
Kinyarwanda	69.6	62.3	89.6	52.6	75.6	28.0	40.3	69.6	67.6	97.1	65.7	94.4	57.8	83.1	69.6	66.9	96.2	67.8	97.4	59.1	85.0
Southern Sotho	68.7	59.4	86.6	46.6	67.8	27.2	39.6	68.7	67.2	97.9	62.0	90.3	56.0	81.6	68.7	68.3	99.5	63.9	93.0	56.0	81.6
Shona	67.9	59.4	87.6	48.1	70.9	26.0	38.3	67.9	65.3	96.2	62.1	91.5	53.8	79.2	67.9	65.8	96.9	63.8	93.9	55.2	81.3
Somali	66.0	56.8	86.0	40.2	60.9	26.4	40.1	66.0	64.4	97.6	59.9	90.7	50.0	75.8	66.0	64.0	97.0	61.3	92.9	49.3	74.7
Maori	65.6	58.1	88.6	47.1	71.9	30.1	45.9	65.6	63.1	96.3	58.4	89.2	51.6	78.6	65.6	63.4	96.8	60.3	92.0	52.9	80.7
Chichewa	58.4	50.4	86.3	36.7	62.7	28.4	48.7	58.4	56.0	95.8	48.9	83.7	43.1	73.8	58.4	54.6	93.3	54.0	92.4	46.3	79.3
Igbo	56.6	48.4	85.7	38.3	67.8	25.3	44.8	56.6	53.6	94.7	49.8	88.0	44.2	78.2	56.6	54.4	96.3	49.8	88.0	43.8	77.4
Swati	55.9	44.3	79.3	37.2	66.6	26.6	47.5	55.9	50.1	89.7	46.9	83.9	39.2	70.2	55.9	50.4	90.3	46.8	83.7	40.8	73.0
Guarani	53.6	46.3	86.5	36.2	67.6	28.1	52.5	53.6	51.4	96.1	47.9	89.4	42.2	78.8	53.6	51.7	96.5	49.7	92.7	43.6	81.3
Lingala	51.7	43.7	84.5	36.4	70.5	27.2	52.7	51.7	47.6	92.0	44.6	86.2	40.1	77.6	51.7	48.4	93.8	47.1	91.2	39.7	76.8
Oromo	51.6	45.2	87.7	35.0	67.9	25.9	50.2	51.6	47.6	92.2	45.7	88.6	38.1	73.9	51.6	49.6	96.1	47.2	91.6	42.4	82.3
Tibetan	47.8	39.3	82.3	35.7	74.7	25.9	54.2	47.8	44.6	93.3	42.6	89.1	37.3	78.1	47.8	45.4	95.1	43.0	90.0	38.2	80.0
Luganda	47.8	41.7	87.2	34.4	72.1	26.0	54.4	47.8	46.1	96.5	41.7	87.2	35.4	74.2	47.8	46.6	97.4	43.9	91.9	40.0	83.7
Bambara	37.4	32.1	85.8	28.1	75.1	25.8	68.8	37.4	38.0	101.5	35.1	93.8	31.6	84.3	37.4	36.2	96.7	34.2	91.4	31.7	84.6
Avg	81.0	74.8	91.8	64.3	78.6	38.2	47.2	81.0	79.8	98.3	78.0	95.7	73.0	89.2	81.0	79.9	98.4	78.5	96.5	73.7	90.2

Table 27: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Qwen3.5 on Belebele under top-k reduction, global α, and token α. Per-language mean within tier. Part 2: tiers T2–T1 and overall mean.

Gemma4 – Belebele														
	Standard							Global α						
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	82.7	77.4	93.6	70.3	84.9	48.1	57.9	82.7	77.4	93.5	71.2	85.9	49.4	59.5
English	89.3	85.4	95.6	81.6	91.3	64.4	72.1	89.3	86.0	96.3	82.1	91.9	68.3	76.5
French	85.9	79.7	92.8	75.6	88.0	51.2	59.6	85.9	79.9	93.0	76.1	88.6	51.8	60.3
Chinese (S.)	83.1	77.2	92.9	68.1	82.0	44.9	54.0	83.1	77.4	93.2	68.1	82.0	43.4	52.3
German	82.9	77.4	93.4	70.9	85.5	46.6	56.2	82.9	78.2	94.4	72.4	87.4	48.1	58.0
Spanish	81.1	77.4	95.5	68.4	84.4	49.4	61.0	81.1	75.7	93.3	69.2	85.3	48.6	59.9
Arabic	80.3	74.1	92.3	62.7	78.0	36.4	45.4	80.3	72.9	90.7	64.1	79.8	40.4	50.3
Japanese	76.4	70.7	92.4	65.0	85.0	43.8	57.3	76.4	71.7	93.8	66.0	86.3	45.1	59.0
T4	80.7	74.6	92.5	68.0	84.2	44.1	54.6	80.7	75.0	92.9	68.8	85.3	46.6	57.7
Catalan	84.2	78.4	93.1	72.3	85.9	48.3	57.4	84.2	77.1	91.6	72.9	86.5	49.7	59.0
Italian	84.2	77.6	92.1	72.3	85.9	48.0	57.0	84.2	77.0	91.4	72.4	86.0	49.0	58.2
Czech	84.0	77.9	92.7	71.0	84.5	47.8	56.9	84.0	78.9	93.9	70.8	84.3	50.4	60.1
Dutch	83.8	76.8	91.6	72.1	86.1	47.0	56.1	83.8	75.6	90.2	70.0	83.6	52.0	62.1
Croatian	83.3	76.6	91.9	67.6	81.1	43.7	52.4	83.3	77.0	92.4	68.2	81.9	45.2	54.3
Swedish	83.1	78.0	93.9	72.1	86.8	46.4	55.9	83.1	78.0	93.9	71.9	86.5	50.1	60.3
Portuguese	82.9	78.8	95.0	71.3	86.1	49.1	59.2	82.9	80.1	96.6	73.6	88.7	52.9	63.8
Vietnamese	82.9	75.1	90.6	70.4	85.0	47.6	57.4	82.9	75.6	91.2	70.0	84.5	50.8	61.3
Serbian	81.8	73.8	90.2	66.9	81.8	43.8	53.5	81.8	74.2	90.8	68.3	83.6	44.6	54.5
Russian	81.4	77.1	94.7	72.1	88.5	49.8	61.1	81.4	77.1	94.7	70.9	87.0	49.7	61.0
Hungarian	80.6	74.3	92.3	68.7	85.2	40.9	50.8	80.6	75.2	93.4	70.0	86.9	41.9	52.0
Polish	80.3	73.0	90.9	68.2	84.9	41.6	51.7	80.3	75.6	94.1	69.4	86.4	44.8	55.7
Turkish	80.1	75.3	94.0	67.7	84.5	43.6	54.4	80.1	75.7	94.5	68.2	85.2	48.7	60.7
Finnish	79.9	74.4	93.2	66.1	82.8	41.6	52.0	79.9	74.8	93.6	69.4	86.9	44.9	56.2
Korean	77.4	71.4	92.3	64.7	83.5	42.2	54.5	77.4	70.9	91.5	64.8	83.6	44.4	57.4
Persian	76.7	71.3	93.0	64.1	83.6	39.6	51.6	76.7	70.2	91.6	65.6	85.5	40.2	52.5
Basque	73.2	66.0	90.1	57.4	78.5	36.1	49.3	73.2	68.4	93.5	61.2	83.6	37.7	51.4
Hindi	72.2	67.0	92.8	58.4	80.9	36.7	50.8	72.2	68.0	94.2	61.1	84.6	42.7	59.1
T3	78.2	71.5	91.4	64.1	81.9	40.4	51.6	78.2	71.9	92.0	65.2	83.3	43.2	55.2
Danish	83.3	77.7	93.2	71.3	85.6	47.2	56.7	83.3	78.9	94.7	72.2	86.7	50.0	60.0
Greek	82.7	78.9	95.4	72.0	87.1	46.1	55.8	82.7	78.3	94.8	73.0	88.3	53.2	64.4
Lithuanian	82.7	72.3	87.5	63.1	76.3	40.9	49.5	82.7	73.7	89.1	66.1	80.0	42.3	51.2
Romanian	82.7	76.6	92.6	70.2	84.9	45.2	54.7	82.7	75.2	91.0	71.7	86.7	47.4	57.4
Latvian	82.0	73.6	89.7	65.2	79.5	38.4	46.9	82.0	74.2	90.5	67.9	82.8	40.8	49.7
Bulgarian	81.8	75.3	92.1	67.9	83.0	43.7	53.4	81.8	76.6	93.6	68.7	84.0	46.8	57.2
Slovak	81.4	76.3	93.7	68.9	84.6	47.9	58.8	81.4	76.8	94.3	70.2	86.2	50.6	62.1
Ukrainian	81.4	74.2	91.1	66.2	81.3	41.7	51.2	81.4	76.0	93.3	67.7	83.1	45.4	55.8
Filipino	80.7	71.9	89.1	66.7	82.6	43.6	54.0	80.7	72.2	89.5	67.7	83.9	44.3	55.0
Afrikaans	80.4	74.6	92.7	67.0	83.3	39.6	49.2	80.4	74.2	92.3	66.4	82.6	42.7	53.0
Slovenian	80.3	75.2	93.6	68.2	84.9	43.1	53.7	80.3	74.7	92.9	70.1	87.3	45.9	57.1
Indonesian	80.1	75.7	94.5	68.6	85.6	45.6	56.9	80.1	74.9	93.5	69.2	86.4	48.0	59.9
Malay	79.9	73.2	91.7	65.0	81.4	41.9	52.4	79.9	72.8	91.1	65.2	81.6	43.7	54.7
Hebrew	79.0	72.4	91.7	64.6	81.7	39.6	50.1	79.0	71.1	90.0	64.8	82.0	42.7	54.0
Estonian	78.1	70.6	90.3	63.4	81.2	40.9	52.3	78.1	72.2	92.5	64.8	82.9	41.9	53.6
Uzbek	75.8	67.0	88.4	60.0	79.2	34.0	44.9	75.8	69.0	91.1	63.1	83.3	40.7	53.7
Georgian	74.7	67.4	90.3	58.9	78.9	33.8	45.2	74.7	67.7	90.6	61.2	82.0	35.6	47.6
Cebuano	74.2	65.2	87.9	57.4	77.4	38.1	51.3	74.2	64.7	87.1	57.1	76.9	38.8	52.2
Bengali	73.3	68.4	93.3	59.3	80.9	38.4	52.4	73.3	68.7	93.6	61.1	83.3	42.6	58.0
Tamil	72.0	65.7	91.2	59.4	82.6	37.4	52.0	72.0	66.1	91.8	58.8	81.6	38.8	53.9
Thai	71.7	67.8	94.6	59.7	83.3	39.1	54.6	71.7	68.7	95.8	60.4	84.3	36.2	50.5
Urdu	71.6	63.0	88.0	56.0	78.3	32.3	45.2	71.6	63.7	89.0	55.0	76.9	38.6	53.9
Kazakh	68.8	61.8	89.8	55.8	81.1	31.4	45.7	68.8	64.3	93.5	57.6	83.7	37.3	54.3

Table 28: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4,3,2\}$ for Gemma4 on Belebele under top-k reduction and global α. Per-language mean within tier. Part 1: tiers T5–T3.

	Gemma4 – Belebele													
	Standard							Global α						
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T2	61.2	53.2	86.9	46.6	76.1	30.6	51.4	61.2	54.2	88.6	47.6	78.1	32.1	53.4
Icelandic	76.7	68.7	89.6	63.0	82.2	38.7	50.4	76.7	68.3	89.1	62.7	81.7	41.4	54.1
Swahili	76.3	68.1	89.2	60.2	78.9	36.6	47.9	76.3	69.3	90.8	60.0	78.6	39.8	52.1
Maltese	75.9	66.6	87.7	55.7	73.4	35.2	46.4	75.9	66.6	87.7	58.7	77.3	38.8	51.1
Marathi	72.8	67.0	92.1	59.3	81.5	37.4	51.5	72.8	68.3	93.9	59.0	81.1	40.7	55.9
Punjabi	69.4	63.1	90.9	54.6	78.6	32.6	46.9	69.4	64.8	93.3	56.9	81.9	33.2	47.8
Haitian Creole	68.4	60.3	88.1	51.6	75.3	33.9	49.5	68.4	61.6	89.9	53.2	77.8	34.8	50.8
Lao	66.0	56.8	86.0	51.0	77.3	29.7	44.9	66.0	58.4	88.6	49.9	75.6	31.3	47.5
Amharic	65.0	54.3	83.6	48.6	74.7	29.4	45.3	65.0	56.7	87.2	47.6	73.2	32.6	50.1
Hausa	56.3	43.2	76.7	40.0	71.0	24.8	44.0	56.3	47.6	84.4	41.3	73.4	25.4	45.2
Tigrinya	55.0	46.9	85.3	40.2	73.1	24.7	44.8	55.0	47.3	86.1	43.0	78.2	25.7	46.7
Xhosa	54.7	45.2	82.7	38.4	70.3	27.8	50.8	54.7	47.7	87.2	41.9	76.6	27.9	51.0
Tswana	53.1	46.2	87.0	36.9	69.5	27.8	52.3	53.1	43.9	82.6	36.4	68.6	28.8	54.2
Zulu	51.8	44.6	86.1	40.2	77.7	26.4	51.1	51.8	43.8	84.5	37.4	72.3	27.8	53.6
Yoruba	44.6	38.2	85.8	33.7	75.6	26.9	60.3	44.6	38.0	85.3	36.6	82.0	28.4	63.8
Wolof	31.6	29.2	92.6	26.0	82.4	26.7	84.5	31.6	31.0	98.2	29.6	93.7	24.6	77.8
T1	62.1	54.9	88.2	48.1	77.4	31.5	51.6	62.1	55.3	89.0	48.7	78.2	32.4	52.9
Norwegian	80.2	75.0	93.5	67.4	84.1	44.8	55.8	80.2	75.2	93.8	70.7	88.1	49.4	61.6
Albanian	79.1	72.8	92.0	64.7	81.7	38.1	48.2	79.1	72.1	91.2	65.2	82.4	39.9	50.4
Macedonian	79.1	72.7	91.9	65.7	83.0	39.4	49.9	79.1	71.7	90.6	67.0	84.7	44.2	55.9
Armenian	76.2	68.7	90.1	58.2	76.4	34.8	45.6	76.2	68.7	90.1	58.2	76.4	37.6	49.3
Malayalam	75.3	64.0	85.0	58.0	77.0	36.3	48.2	75.3	66.7	88.5	56.8	75.4	39.3	52.2
Azerbaijani	72.0	65.6	91.0	57.0	79.2	37.8	52.5	72.0	67.3	93.5	58.0	80.6	36.3	50.5
Javanese	71.7	60.0	83.7	51.8	72.2	32.6	45.4	71.7	61.1	85.3	51.8	72.2	32.0	44.7
Gujarati	71.6	61.9	86.5	56.9	79.5	35.7	49.8	71.6	62.2	87.0	57.8	80.7	35.1	49.1
Sinhala	71.4	64.9	90.8	55.3	77.4	34.1	47.7	71.4	64.0	89.6	59.4	83.2	35.6	49.8
Tajik	71.4	62.4	87.4	51.7	72.3	33.8	47.3	71.4	63.1	88.3	55.3	77.4	35.0	49.0
Kannada	71.0	65.2	91.9	56.8	80.0	35.0	49.3	71.0	64.2	90.5	56.6	79.7	39.0	54.9
Telugu	69.8	64.2	92.0	56.0	80.3	34.1	48.9	69.8	63.9	91.6	57.6	82.5	35.3	50.6
Kyrgyz	69.6	62.1	89.3	56.8	81.6	35.3	50.8	69.6	63.0	90.6	55.8	80.2	33.3	47.9
Khmer	69.1	60.2	87.1	54.4	78.8	30.7	44.4	69.1	62.2	90.0	55.2	79.9	34.3	49.7
Nepali	69.1	62.9	91.0	55.2	79.9	35.1	50.8	69.1	62.6	90.5	56.1	81.2	37.6	54.3
Kurdish	68.9	58.1	84.4	48.6	70.5	31.4	45.6	68.9	59.1	85.8	49.3	71.6	30.0	43.5
Mongolian	67.2	56.6	84.1	49.6	73.7	33.7	50.1	67.2	58.3	86.8	51.6	76.7	34.7	51.6
Assamese	66.9	61.4	91.9	52.3	78.2	29.8	44.5	66.9	61.9	92.5	54.6	81.6	35.7	53.3
Sundanese	66.7	58.7	88.0	46.9	70.3	28.6	42.8	66.7	57.4	86.2	49.3	74.0	30.1	45.2
Oriya	66.6	59.0	88.6	49.3	74.1	29.9	44.9	66.6	61.2	92.0	51.2	77.0	32.1	48.2
Burmese	65.6	58.3	89.0	50.9	77.6	28.2	43.1	65.6	58.7	89.5	52.0	79.3	29.1	44.4
Malagasy	65.3	55.9	85.5	50.1	76.7	33.0	50.5	65.3	57.8	88.4	50.8	77.7	32.0	49.0
Pashto	64.9	55.0	84.8	45.9	70.7	33.2	51.2	64.9	54.3	83.7	45.8	70.5	31.1	47.9
Somali	57.4	48.2	83.9	43.4	75.6	29.9	52.0	57.4	49.6	86.3	44.8	77.9	30.7	53.4
Shona	56.6	49.6	87.6	43.7	77.2	27.2	48.1	56.6	49.3	87.2	42.9	75.8	28.9	51.1
Kinyarwanda	56.2	49.9	88.7	40.7	72.3	26.3	46.8	56.2	47.9	85.2	41.2	73.3	26.6	47.2
Southern Sotho	54.3	46.3	85.3	37.8	69.5	26.9	49.5	54.3	46.0	84.7	40.3	74.2	29.1	53.6
Chichewa	49.2	41.9	85.1	37.6	76.3	25.6	51.9	49.2	41.2	83.7	33.6	68.2	27.6	56.0
Guarani	48.3	41.8	86.4	38.1	78.9	27.8	57.5	48.3	42.0	86.9	38.3	79.3	27.1	56.1
Maori	48.1	39.6	82.2	37.3	77.6	27.9	58.0	48.1	39.9	82.9	33.4	69.5	26.8	55.7
Swati	47.2	42.4	89.9	36.4	77.2	28.2	59.8	47.2	42.7	90.4	37.2	78.8	26.9	56.9
Igbo	45.7	39.6	86.6	34.6	75.7	24.8	54.3	45.7	39.2	85.9	31.3	68.6	26.8	58.6
Oromo	44.8	42.9	95.8	40.6	90.6	29.9	66.7	44.8	44.1	98.5	38.4	85.9	26.7	59.6
Lingala	42.0	38.8	92.3	34.3	81.7	26.4	63.0	42.0	37.8	89.9	35.1	83.6	27.4	65.3
Tibetan	41.4	36.0	86.9	32.4	78.3	25.9	62.5	41.4	37.2	89.8	31.3	75.6	24.7	59.5
Luganda	41.0	36.6	89.2	31.7	77.2	27.3	66.7	41.0	39.2	95.7	33.9	82.7	24.4	59.6
Bambara	37.8	32.2	85.3	30.6	80.9	24.8	65.6	37.8	33.6	88.8	33.1	87.6	27.0	71.5
Avg	70.5	63.6	89.9	56.7	80.0	36.8	52.5	70.5	64.1	90.6	57.5	81.2	38.6	54.8

Table 29: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Gemma4 on Belebele under top-k reduction and global α. Per-language mean within tier. Part 2: tiers T2–T1 and overall mean.

N Detailed Results: GlobalMMLU

Qwen3 – GlobalMMLU																					
	Standard							Global α							Token α						
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	70.4	62.9	89.4	50.5	71.7	27.0	38.4	70.4	67.4	95.7	64.2	91.1	54.6	77.3	70.4	67.5	95.9	64.1	91.1	53.5	75.8
English	77.0	69.4	90.1	58.2	75.6	30.2	39.1	77.0	74.1	96.2	71.4	92.6	64.7	84.0	77.0	74.5	96.7	71.4	92.7	65.2	84.6
Chinese (S.)	72.4	64.7	89.3	54.5	75.3	30.8	42.5	72.4	69.1	95.4	66.4	91.7	58.9	81.3	72.4	69.2	95.6	66.3	91.7	55.9	77.3
French	71.6	63.0	88.0	48.1	67.3	25.8	36.0	71.6	68.7	96.0	65.6	91.7	53.6	74.9	71.6	68.4	95.6	65.3	91.3	52.9	74.0
German	69.4	63.5	91.5	50.4	72.7	25.1	36.1	69.4	67.0	96.6	63.6	91.7	53.9	77.7	69.4	67.3	97.0	63.7	91.8	52.7	76.0
Japanese	67.7	60.7	89.6	48.9	72.1	25.6	37.7	67.7	64.4	95.0	60.8	89.7	50.7	74.8	67.7	64.5	95.2	60.7	89.6	47.8	70.6
Arabic	64.0	56.4	88.0	43.0	67.1	24.7	38.6	64.0	60.9	95.1	57.2	89.3	45.6	71.2	64.0	61.1	95.4	57.2	89.4	46.3	72.2
T4	67.6	56.9	84.2	42.6	62.9	25.1	37.2	67.6	62.5	92.6	58.5	86.5	45.9	67.7	67.6	62.2	92.1	58.1	85.9	44.2	65.2
Russian	73.1	62.7	85.8	47.9	65.6	25.1	34.3	73.1	66.5	91.0	63.0	86.3	51.3	70.1	73.1	66.4	90.8	62.8	85.9	50.4	69.0
Portuguese	71.6	63.1	88.1	47.3	66.1	24.6	34.3	71.6	67.9	94.8	64.5	90.1	53.4	74.6	71.6	67.5	94.2	64.7	90.3	50.2	70.0
Swedish	71.1	59.5	83.6	44.3	62.3	25.4	35.8	71.1	64.4	90.5	60.7	85.3	47.6	66.9	71.1	63.7	89.6	60.0	84.4	46.4	65.3
Czech	70.7	57.9	81.8	42.9	60.7	24.6	34.8	70.7	63.3	89.5	59.7	84.4	46.5	65.8	70.7	63.3	89.5	59.5	84.1	45.3	64.1
Vietnamese	69.4	58.3	83.9	44.0	63.4	25.3	36.4	69.4	62.5	90.1	59.4	85.6	47.5	68.4	69.4	62.7	90.4	59.2	85.3	45.5	65.6
Dutch	68.9	60.5	87.7	46.3	67.2	25.6	37.2	68.9	65.3	94.8	62.4	90.5	49.5	71.8	68.9	65.3	94.7	61.4	89.0	47.7	69.2
Korean	66.5	58.6	88.0	45.7	68.7	25.4	38.2	66.5	63.0	94.7	59.6	89.6	48.8	73.3	66.5	62.8	94.4	59.1	88.9	47.2	70.9
Polish	65.7	57.5	87.5	42.0	63.9	25.3	38.5	65.7	62.2	94.7	58.7	89.3	45.0	68.5	65.7	62.3	94.8	58.2	88.6	43.6	66.3
Persian	63.9	43.2	67.7	32.1	50.2	24.2	38.0	63.9	56.1	87.8	49.7	77.8	37.1	58.1	63.9	55.3	86.5	49.4	77.4	35.6	55.7
Serbian	63.6	54.0	84.9	38.7	60.9	25.5	40.2	63.6	60.0	94.3	54.4	85.6	41.5	65.2	63.6	59.3	93.2	54.3	85.4	39.1	61.4
Hindi	59.2	51.3	86.6	37.3	63.1	24.8	41.9	59.2	56.8	95.9	51.5	87.1	36.6	61.8	59.2	56.0	94.6	50.4	85.1	35.3	59.6
T3	65.2	53.9	82.6	39.1	59.9	25.0	38.5	65.2	60.0	92.0	55.1	84.4	41.6	63.6	65.2	59.6	91.5	54.9	84.1	40.0	61.3
Romanian	71.8	59.7	83.1	44.0	61.2	24.8	34.6	71.8	64.3	89.5	60.4	84.1	46.4	64.6	71.8	64.2	89.3	60.4	84.0	43.8	61.0
Ukrainian	70.1	57.7	82.4	41.3	58.9	24.9	35.5	70.1	62.7	89.5	57.6	82.2	45.2	64.5	70.1	62.4	89.1	57.4	81.9	42.3	60.3
Indonesian	69.4	59.5	85.8	43.6	62.9	25.6	36.8	69.4	66.2	95.3	61.1	88.1	45.7	65.9	69.4	66.0	95.1	60.9	87.7	44.0	63.4
Greek	67.8	55.9	82.4	40.6	59.8	24.9	36.6	67.8	60.1	88.7	56.3	83.1	42.8	63.1	67.8	60.4	89.0	56.2	82.9	42.7	62.9
Filipino	65.3	56.1	85.9	40.3	61.8	25.4	38.8	65.3	61.3	93.9	56.5	86.5	44.2	67.8	65.3	61.0	93.5	56.7	86.9	40.8	62.5
Malay	65.1	54.8	84.2	39.3	60.3	25.6	39.2	65.1	62.3	95.6	55.8	85.7	40.6	62.4	65.1	62.1	95.3	55.3	84.9	39.3	60.4
Lithuanian	59.6	48.9	82.0	34.8	58.4	24.5	41.1	59.6	55.5	93.1	50.7	85.1	38.7	64.9	59.6	54.9	92.1	50.9	85.4	38.6	64.7
Hebrew	59.5	45.7	76.8	33.3	56.1	24.3	40.9	59.5	53.5	90.1	49.3	83.0	37.4	62.9	59.5	53.4	89.8	49.6	83.4	37.8	63.6
Bengali	58.2	46.9	80.6	34.9	59.9	24.8	42.5	58.2	53.7	92.2	47.7	81.9	33.0	56.7	58.2	52.6	90.3	46.7	80.1	31.0	53.3
T2	40.5	36.0	88.9	29.4	72.6	24.4	60.2	40.5	38.3	94.7	35.5	87.7	29.4	72.5	40.5	38.0	93.9	35.6	87.9	28.3	69.7
Swahili	40.5	36.0	88.9	29.4	72.6	24.4	60.2	40.5	38.3	94.7	35.5	87.7	29.4	72.5	40.5	38.0	93.9	35.6	87.9	28.3	69.7
T1	53.7	42.3	78.8	31.9	59.6	24.3	45.8	53.7	47.4	88.4	43.2	80.5	31.6	59.3	53.7	46.8	87.2	42.3	78.9	30.0	56.2
Nepali	59.0	46.5	78.7	34.8	58.9	24.8	42.0	59.0	51.4	87.2	46.6	79.0	34.0	57.7	59.0	50.9	86.3	46.0	77.9	32.7	55.5
Telugu	56.7	44.9	79.2	32.2	56.8	24.0	42.4	56.7	50.6	89.2	46.0	81.1	31.2	55.1	56.7	50.2	88.5	45.0	79.5	29.0	51.2
Sinhala	45.4	35.6	78.5	28.6	63.1	24.1	53.2	45.4	40.3	88.8	36.9	81.4	29.5	65.1	45.4	39.3	86.7	36.0	79.3	28.1	62.0
Avg	65.1	55.1	84.4	41.6	63.8	25.3	39.5	65.1	60.4	92.7	56.3	86.2	44.3	67.7	65.1	60.2	92.2	56.0	85.8	42.8	65.4

Table 30: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Qwen3 on GlobalMMLU under top-k reduction, global α, and token α. Per-language mean within tier. Qwen3 is restricted to its 73 supported languages.

Qwen3.5 – GlobalMMLU																					
	Standard							Global α							Token α						
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	78.2	74.2	94.9	69.0	88.3	52.0	66.4	78.2	76.6	98.0	75.0	95.9	71.3	91.2	78.2	76.5	97.8	75.1	96.0	70.8	90.5
English	83.1	79.9	96.2	76.6	92.2	64.6	77.8	83.1	82.2	99.0	81.0	97.5	78.6	94.6	83.1	81.9	98.6	81.2	97.7	78.3	94.3
French	78.8	75.1	95.4	69.0	87.6	47.8	60.7	78.8	77.3	98.1	76.1	96.6	72.3	91.8	78.8	77.3	98.2	76.3	96.8	71.9	91.4
German	78.4	73.2	93.4	68.3	87.1	53.6	68.3	78.4	76.5	97.5	74.3	94.8	69.9	89.1	78.4	76.2	97.2	74.3	94.7	69.7	88.8
Chinese (S.)	77.6	74.0	95.3	69.3	89.3	54.3	70.0	77.6	76.4	98.5	75.1	96.8	72.6	93.6	77.6	76.0	98.0	74.8	96.4	71.8	92.5
Japanese	76.2	72.3	94.9	66.1	86.7	49.9	65.5	76.2	74.4	97.6	72.6	95.2	68.1	89.3	76.2	74.5	97.8	72.9	95.6	67.2	88.2
Arabic	75.0	70.5	94.0	65.0	86.7	41.9	55.9	75.0	72.7	97.1	70.9	94.6	66.4	88.6	75.0	72.9	97.3	71.1	94.8	65.8	87.8
T4	75.9	70.7	93.1	64.1	84.5	44.0	57.9	75.9	73.8	97.2	72.0	94.9	66.9	88.2	75.9	74.0	97.5	72.2	95.2	66.3	87.3
Dutch	78.0	71.9	92.1	65.6	84.1	48.4	62.0	78.0	76.2	97.7	74.2	95.1	70.2	89.9	78.0	76.6	98.1	74.5	95.5	70.2	90.0
Portuguese	78.0	67.8	87.0	62.3	79.9	45.8	58.7	78.0	75.0	96.2	74.5	95.5	71.0	91.0	78.0	76.8	98.5	75.7	97.1	71.2	91.3
Russian	77.9	73.7	94.5	67.4	86.5	46.9	60.1	77.9	76.1	97.6	74.1	95.1	69.7	89.4	77.9	76.2	97.7	74.5	95.6	69.0	88.5
Swedish	77.7	73.6	94.6	66.3	85.2	43.5	56.0	77.7	75.8	97.5	73.2	94.2	66.2	85.2	77.7	75.7	97.3	73.0	93.9	65.4	84.2
Czech	76.6	71.5	93.3	63.7	83.2	44.0	57.5	76.6	74.3	97.0	72.9	95.2	66.8	87.2	76.6	74.9	97.8	73.0	95.4	65.2	85.1
Polish	76.5	72.5	94.7	66.5	86.9	48.8	63.8	76.5	75.0	97.9	72.8	95.2	67.0	87.5	76.5	74.7	97.7	72.5	94.7	65.2	85.1
Vietnamese	75.6	71.7	94.8	66.1	87.4	46.5	61.5	75.6	73.7	97.5	71.9	95.1	66.2	87.6	75.6	74.0	97.8	71.8	95.0	66.4	87.8
Serbian	75.5	70.8	93.8	64.0	84.8	39.7	52.6	75.5	73.1	96.9	71.0	94.0	64.9	86.0	75.5	73.1	96.9	71.5	94.8	63.6	84.2
Korean	75.1	71.3	94.9	65.7	87.4	45.8	61.0	75.1	73.2	97.4	71.7	95.4	68.0	90.6	75.1	73.5	97.8	72.2	96.1	67.2	89.4
Persian	72.2	67.3	93.2	59.8	82.9	37.2	51.5	72.2	69.7	96.6	68.0	94.2	63.0	87.2	72.2	69.5	96.2	68.0	94.2	62.6	86.7
Hindi	71.3	65.3	91.5	58.0	81.3	37.1	52.0	71.3	69.2	97.0	67.7	94.9	63.0	88.4	71.3	69.2	97.0	67.8	95.1	62.9	88.2
T3	74.5	69.5	93.4	62.4	83.7	41.9	56.1	74.5	72.4	97.2	70.3	94.4	65.0	87.2	74.5	72.4	97.3	70.4	94.5	64.5	86.6
Romanian	78.4	74.4	94.9	68.5	87.4	49.9	63.7	78.4	76.7	97.8	74.9	95.6	69.5	88.7	78.4	76.5	97.6	74.6	95.2	68.7	87.6
Indonesian	77.4	73.6	95.1	68.1	87.9	50.5	65.2	77.4	75.6	97.6	73.7	95.1	69.9	90.3	77.4	75.6	97.6	74.1	95.7	69.7	90.0
Ukrainian	76.5	70.7	92.4	64.0	83.6	41.1	53.7	76.5	74.4	97.2	71.9	94.0	65.7	85.9	76.5	74.6	97.4	72.1	94.2	65.7	85.8
Greek	75.9	68.6	90.4	59.9	79.0	37.6	49.5	75.9	73.3	96.6	71.5	94.3	66.1	87.1	75.9	73.6	97.1	71.8	94.6	66.4	87.5
Malay	75.5	71.7	95.0	65.1	86.2	44.2	58.5	75.5	73.3	97.1	71.1	94.1	65.7	87.0	75.5	73.5	97.3	71.3	94.4	66.0	87.5
Filipino	73.4	68.9	93.8	61.8	84.2	43.5	59.3	73.4	71.2	96.9	69.1	94.2	63.9	87.1	73.4	71.1	96.8	69.3	94.4	64.8	88.3
Lithuanian	71.9	67.4	93.7	60.6	84.2	38.4	53.4	71.9	70.0	97.3	67.6	94.0	62.1	86.3	71.9	69.9	97.1	67.5	93.8	60.4	84.0
Hebrew	70.9	66.8	94.2	58.8	82.9	37.5	52.9	70.9	69.3	97.7	67.3	94.9	61.3	86.4	70.9	69.2	97.5	66.4	93.6	59.0	83.2
Bengali	70.3	63.7	90.7	54.8	78.0	34.2	48.6	70.3	68.1	96.9	65.8	93.7	60.6	86.3	70.3	68.1	96.9	66.3	94.3	60.2	85.6
T2	57.6	51.5	89.5	44.6	77.4	30.4	52.8	57.6	54.8	95.2	52.2	90.6	46.6	81.0	57.6	54.8	95.3	52.8	91.8	47.1	81.8
Swahili	64.2	59.5	92.6	52.9	82.5	34.2	53.3	64.2	61.4	95.6	59.3	92.3	52.9	82.4	64.2	61.5	95.7	59.7	93.0	53.1	82.7
Amharic	58.6	51.2	87.4	44.1	75.3	29.4	50.1	58.6	55.5	94.7	52.8	90.1	47.1	80.4	58.6	55.8	95.1	53.2	90.7	47.6	81.2
Hausa	55.2	48.4	87.7	41.0	74.4	28.8	52.1	55.2	52.1	94.4	48.8	88.4	43.7	79.2	55.2	52.1	94.4	49.9	90.5	44.0	79.7
Yoruba	52.2	47.1	90.2	40.4	77.5	29.1	55.7	52.2	50.0	95.9	47.8	91.6	42.8	82.0	52.2	50.0	95.8	48.5	93.0	43.7	83.8
T1	61.5	55.7	90.4	48.0	77.9	31.8	51.8	61.5	58.8	95.5	56.3	91.4	50.4	81.7	61.5	58.9	95.6	56.7	92.0	50.7	82.4
Nepali	69.2	61.5	88.9	53.7	77.6	34.1	49.3	69.2	66.3	95.8	63.5	91.7	57.0	82.4	69.2	66.3	95.8	63.8	92.3	57.2	82.7
Telugu	68.8	63.4	92.1	54.9	79.8	35.0	50.8	68.8	66.4	96.6	64.3	93.6	58.2	84.7	68.8	66.2	96.3	64.2	93.4	57.9	84.2
Kyrgyz	65.7	59.4	90.4	50.9	77.4	32.3	49.2	65.7	63.1	96.0	61.0	92.8	55.1	83.9	65.7	63.2	96.2	60.8	92.5	53.3	81.1
Sinhala	65.4	59.5	91.0	50.0	76.4	32.1	49.1	65.4	62.7	95.8	60.1	91.8	54.9	84.0	65.4	62.7	95.8	61.1	93.4	55.5	84.8
Igbo	58.7	53.5	91.2	46.3	79.0	30.4	51.8	58.7	56.1	95.5	53.7	91.5	46.9	79.9	58.7	56.3	96.0	54.8	93.3	48.2	82.1
Somali	57.3	51.3	89.4	43.3	75.5	30.1	52.5	57.3	54.1	94.4	51.7	90.2	45.3	79.0	57.3	54.3	94.8	51.8	90.5	45.8	79.9
Shona	57.3	51.0	88.9	43.8	76.4	31.0	54.1	57.3	53.7	93.8	50.9	88.9	45.5	79.4	57.3	54.2	94.6	51.9	90.5	46.9	81.8
Malagasy	57.1	51.9	90.9	45.3	79.4	30.7	53.7	57.1	54.6	95.6	52.4	91.7	46.5	81.4	57.1	54.5	95.4	52.8	92.4	47.5	83.1
Chichewa	54.3	49.4	90.9	43.5	80.0	30.2	55.6	54.3	52.0	95.7	49.1	90.3	43.9	80.9	54.3	52.1	95.8	49.0	90.2	44.4	81.7
Avg	70.7	65.5	92.5	58.8	82.7	40.5	56.9	70.7	68.5	96.7	66.4	93.7	61.2	86.2	70.7	68.6	96.8	66.7	94.1	61.0	85.9

Table 31: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Qwen3.5 on GlobalMMLU under top-k reduction, global α, and token α. Per-language mean within tier.

Gemma4 – GlobalMMLU														
	Standard							Global α						
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	67.2	62.6	93.1	57.4	85.4	43.5	64.5	67.2	62.8	93.5	57.9	86.2	45.5	67.6
English	73.1	69.7	95.3	65.6	89.7	54.5	74.5	73.1	69.7	95.4	65.3	89.3	55.3	75.7
French	68.6	64.1	93.5	58.4	85.2	44.0	64.1	68.6	64.5	94.1	59.2	86.3	46.2	67.4
German	67.5	62.4	92.5	57.8	85.7	42.9	63.5	67.5	62.8	93.0	58.3	86.4	44.7	66.3
Chinese (S.)	65.9	60.7	92.1	55.8	84.6	42.0	63.6	65.9	61.4	93.1	56.5	85.7	44.0	66.7
Japanese	64.9	60.8	93.7	55.8	86.0	40.7	62.7	64.9	60.6	93.4	56.3	86.8	43.5	67.1
Arabic	62.8	57.7	91.8	51.2	81.5	36.9	58.8	62.8	58.0	92.3	51.9	82.6	39.1	62.2
T4	65.0	60.3	92.7	54.7	84.0	39.9	61.3	65.0	60.5	93.0	55.3	85.0	41.9	64.4
Portuguese	67.7	63.3	93.5	56.7	83.7	42.3	62.4	67.7	63.7	94.0	57.4	84.7	44.2	65.3
Dutch	67.2	63.3	94.3	58.0	86.3	43.0	64.0	67.2	63.1	94.0	58.6	87.3	44.3	65.9
Swedish	67.0	62.2	92.8	56.0	83.6	40.4	60.3	67.0	62.3	93.0	56.7	84.6	42.9	64.1
Russian	67.0	62.7	93.6	57.3	85.5	42.3	63.2	67.0	62.7	93.6	57.6	86.1	44.5	66.5
Czech	66.0	61.1	92.5	55.8	84.4	40.2	60.9	66.0	61.3	92.9	56.1	84.9	42.7	64.6
Polish	65.1	60.7	93.3	55.6	85.4	40.5	62.2	65.1	60.8	93.4	55.6	85.4	42.3	65.0
Korean	64.6	60.1	92.9	55.0	85.1	41.6	64.4	64.6	59.8	92.6	55.9	86.4	42.7	66.1
Serbian	64.0	58.3	91.1	52.4	82.0	37.3	58.3	64.0	59.1	92.3	53.6	83.7	39.4	61.6
Vietnamese	63.9	59.4	92.9	53.6	83.9	39.5	61.8	63.9	59.4	92.9	54.4	85.0	41.3	64.5
Persian	61.9	56.5	91.3	50.5	81.6	35.4	57.2	61.9	56.6	91.5	51.3	82.9	37.3	60.2
Hindi	60.8	55.9	91.9	50.3	82.7	36.3	59.7	60.8	56.6	93.1	51.3	84.4	39.0	64.1
T3	63.7	58.9	92.4	53.5	83.9	38.6	60.5	63.7	59.2	92.8	54.0	84.7	40.6	63.6
Romanian	66.8	62.1	92.9	56.6	84.8	41.8	62.6	66.8	62.0	92.8	57.1	85.5	43.9	65.7
Ukrainian	65.5	60.6	92.6	55.4	84.5	39.7	60.6	65.5	60.6	92.6	55.8	85.1	41.9	64.0
Indonesian	65.1	60.5	93.0	55.1	84.7	41.4	63.6	65.1	60.9	93.6	55.9	85.9	43.4	66.7
Greek	64.9	60.3	92.9	54.7	84.2	37.7	58.1	64.9	60.5	93.2	55.5	85.5	40.2	61.9
Filipino	64.3	59.6	92.8	55.0	85.6	40.9	63.7	64.3	60.1	93.6	55.2	86.0	41.4	64.5
Malay	63.6	58.6	92.2	53.2	83.6	38.7	60.9	63.6	59.0	92.7	53.9	84.7	41.1	64.6
Hebrew	62.6	57.4	91.6	51.5	82.2	35.7	57.1	62.6	57.7	92.2	51.5	82.2	38.1	60.9
Lithuanian	62.2	57.0	91.6	50.7	81.4	35.4	56.9	62.2	57.2	92.0	51.0	82.0	37.3	60.0
Bengali	58.6	53.6	91.5	49.1	83.8	35.9	61.3	58.6	54.4	92.9	49.9	85.3	37.8	64.6
T2	49.9	45.2	90.6	40.0	80.2	29.8	60.1	49.9	45.7	91.7	40.7	81.5	31.4	63.2
Swahili	56.7	52.2	92.1	46.3	81.6	33.3	58.7	56.7	52.2	92.1	46.6	82.1	34.9	61.5
Amharic	52.9	48.1	90.9	42.3	79.9	30.3	57.3	52.9	48.2	91.1	43.3	81.8	32.7	61.7
Hausa	49.5	44.3	89.6	39.0	78.9	29.4	59.3	49.5	45.1	91.1	39.4	79.7	30.6	61.7
Yoruba	40.4	36.3	89.8	32.4	80.2	26.3	65.1	40.4	37.4	92.5	33.3	82.4	27.4	67.7
T1	53.4	48.6	90.9	43.3	80.9	31.9	59.7	53.4	48.9	91.4	44.0	82.3	33.3	62.3
Nepali	59.4	54.5	91.6	48.7	82.0	34.8	58.6	59.4	54.4	91.5	49.4	83.0	36.8	61.9
Telugu	57.5	52.2	90.7	47.2	82.1	34.2	59.4	57.5	53.2	92.4	48.4	84.1	35.9	62.4
Kyrgyz	56.5	52.3	92.6	46.2	81.8	32.1	56.9	56.5	52.5	92.8	46.2	81.7	34.7	61.4
Sinhala	55.8	50.1	89.9	45.1	80.9	32.9	59.1	55.8	50.8	91.0	45.7	82.0	34.4	61.6
Malagasy	53.6	47.9	89.3	41.9	78.2	31.8	59.4	53.6	48.6	90.7	43.8	81.7	32.4	60.4
Somali	50.6	46.0	90.8	40.3	79.6	30.1	59.5	50.6	46.0	90.9	41.3	81.5	30.8	60.9
Shona	50.5	45.7	90.4	40.7	80.6	30.8	60.9	50.5	45.9	90.9	41.3	81.7	31.9	63.2
Igbo	48.6	44.0	90.7	39.1	80.6	29.5	60.7	48.6	44.0	90.6	39.4	81.1	31.2	64.3
Chichewa	48.5	44.7	92.3	40.0	82.5	30.5	62.9	48.5	44.3	91.5	40.8	84.1	31.2	64.4
Avg	60.8	56.1	92.1	50.7	83.1	37.3	61.1	60.8	56.3	92.5	51.3	84.1	39.1	64.1

Table 32: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Gemma4 on GlobalMMLU under top-k reduction and global α. Per-language mean within tier.

O Detailed Results: XCOPA

Qwen3 – XCOPA																					
	Standard							Global α							Token α						
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	79.6	75.6	95.0	69.0	86.7	55.4	69.6	79.6	77.0	96.7	74.0	93.0	67.8	85.2	79.6	76.8	96.5	77.0	96.7	68.0	85.4
Chinese (S.)	79.6	75.6	95.0	69.0	86.7	55.4	69.6	79.6	77.0	96.7	74.0	93.0	67.8	85.2	79.6	76.8	96.5	77.0	96.7	68.0	85.4
T4	75.9	71.9	94.9	63.7	83.9	54.7	72.3	75.9	72.7	95.7	68.8	90.6	62.3	82.2	75.9	72.0	94.8	68.5	90.2	61.9	81.5
Italian	79.6	73.8	92.7	67.4	84.7	54.0	67.8	79.6	76.6	96.2	72.8	91.5	64.8	81.4	79.6	75.6	95.0	72.4	91.0	65.4	82.2
Vietnamese	78.6	73.4	93.4	65.6	83.5	55.2	70.2	78.6	74.6	94.9	71.2	90.6	63.6	80.9	78.6	75.0	95.4	70.4	89.6	64.2	81.7
Turkish	69.6	68.6	98.6	58.2	83.6	54.8	78.7	69.6	66.8	96.0	62.4	89.7	58.6	84.2	69.6	65.4	94.0	62.6	89.9	56.2	80.7
T3	66.2	62.6	94.8	58.1	87.8	54.2	82.9	66.2	62.8	95.0	60.8	91.8	56.8	85.8	66.2	63.4	95.7	61.1	92.3	56.7	85.7
Indonesian	77.4	72.0	93.0	67.0	86.6	52.6	68.0	77.4	73.0	94.3	71.6	92.5	65.8	85.0	77.4	74.2	95.9	71.8	92.8	66.0	85.3
Tamil	65.6	58.4	89.0	58.0	88.4	57.6	87.8	65.6	62.0	94.5	59.0	89.9	56.4	86.0	65.6	62.2	94.8	59.6	90.9	55.4	84.5
Thai	63.6	63.4	99.7	55.6	87.4	55.8	87.7	63.6	61.4	96.5	61.4	96.5	55.8	87.7	63.6	62.4	98.1	59.6	93.7	56.8	89.3
Estonian	58.0	56.6	97.6	51.6	89.0	51.0	87.9	58.0	54.8	94.5	51.2	88.3	49.0	84.5	58.0	54.6	94.1	53.2	91.7	48.6	83.8
T2	55.3	53.0	95.9	52.5	95.0	54.5	98.6	55.3	53.5	96.7	52.1	94.3	52.4	94.8	55.3	53.9	97.4	51.3	92.8	50.6	91.6
Haitian Creole	56.8	53.6	94.4	53.4	94.0	54.2	95.4	56.8	55.6	97.9	52.4	92.3	52.4	92.3	56.8	55.8	98.2	52.6	92.6	50.2	88.4
Swahili	53.8	52.4	97.4	51.6	95.9	54.8	101.9	53.8	51.4	95.5	51.8	96.3	52.4	97.4	53.8	52.0	96.7	50.0	92.9	51.0	94.8
T1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Avg	68.3	64.8	95.1	59.7	88.0	54.5	81.5	68.3	65.3	95.7	62.8	92.0	58.7	86.5	68.3	65.4	95.9	62.9	92.2	58.2	85.6

Table 33: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4,3,2\}$ for Qwen3 on XCOPA under top-k reduction, global α, and token α. Per-language mean within tier. Qwen3 is restricted to its 73 supported languages.

Qwen3.5 – XCOPA																					
	Standard							Global α							Token α						
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	83.0	81.2	97.8	78.8	94.9	72.0	86.7	83.0	82.8	99.8	82.8	99.8	79.6	95.9	83.0	83.0	100.0	84.0	101.2	79.6	95.9
Chinese (S.)	83.0	81.2	97.8	78.8	94.9	72.0	86.7	83.0	82.8	99.8	82.8	99.8	79.6	95.9	83.0	83.0	100.0	84.0	101.2	79.6	95.9
T4	79.6	77.7	97.7	73.7	92.5	64.5	81.0	79.6	78.3	98.3	75.9	95.4	72.4	90.8	79.6	77.7	97.5	75.0	94.1	71.3	89.5
Vietnamese	83.2	81.2	97.6	78.4	94.2	67.8	81.5	83.2	82.0	98.6	79.2	95.2	79.6	95.7	83.2	80.8	97.1	80.0	96.2	77.4	93.0
Italian	82.0	79.8	97.3	75.6	92.2	64.8	79.0	82.0	82.0	100.0	78.4	95.6	73.6	89.8	82.0	81.8	99.8	78.0	95.1	73.6	89.8
Turkish	73.6	72.2	98.1	67.0	91.0	60.8	82.6	73.6	71.0	96.5	70.2	95.4	64.0	87.0	73.6	70.4	95.7	67.0	91.0	63.0	85.6
T3	72.0	70.2	97.5	66.2	92.0	59.8	83.1	72.0	70.6	98.1	69.5	96.5	66.4	92.2	72.0	69.8	96.9	69.0	95.9	65.8	91.3
Indonesian	80.6	79.6	98.8	76.4	94.8	69.0	85.6	80.6	78.8	97.8	78.0	96.8	76.4	94.8	80.6	78.4	97.3	77.6	96.3	74.8	92.8
Estonian	73.4	71.8	97.8	65.4	89.1	55.2	75.2	73.4	72.2	98.4	72.0	98.1	66.4	90.5	73.4	71.6	97.5	71.2	97.0	65.2	88.8
Tamil	68.6	65.0	94.8	61.8	90.1	59.4	86.6	68.6	67.6	98.5	65.2	95.0	61.6	89.8	68.6	66.8	97.4	64.6	94.2	61.2	89.2
Thai	65.4	64.4	98.5	61.4	93.9	55.6	85.0	65.4	63.8	97.6	62.8	96.0	61.2	93.6	65.4	62.4	95.4	62.8	96.0	61.8	94.5
T2	66.1	62.2	94.1	59.5	90.0	52.9	80.0	66.1	63.8	96.5	62.5	94.5	59.4	89.8	66.1	62.9	95.1	62.3	94.2	60.6	91.7
Haitian Creole	66.8	62.4	93.4	60.6	90.7	54.4	81.4	66.8	65.6	98.2	64.2	96.1	61.8	92.5	66.8	64.8	97.0	63.4	94.9	61.8	92.5
Swahili	65.4	62.0	94.8	58.4	89.3	51.4	78.6	65.4	62.0	94.8	60.8	93.0	57.0	87.2	65.4	61.0	93.3	61.2	93.6	59.4	90.8
T1	55.0	52.8	96.0	51.4	93.5	51.0	92.7	55.0	54.0	98.2	52.0	94.5	51.8	94.2	55.0	54.0	98.2	52.0	94.5	52.2	94.9
Guarani	55.0	52.8	96.0	51.4	93.5	51.0	92.7	55.0	54.0	98.2	52.0	94.5	51.8	94.2	55.0	54.0	98.2	52.0	94.5	52.2	94.9
Avg	72.5	70.2	96.8	66.8	92.2	60.1	83.2	72.5	71.1	98.0	69.6	96.0	66.6	91.9	72.5	70.5	97.1	69.3	95.5	66.4	91.6

Table 34: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4,3,2\}$ for Qwen3.5 on XCOPA under top-k reduction, global α, and token α. Per-language mean within tier.

Gemma4 – XCOPA														
	Standard							Global α						
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	83.2	82.8	99.5	81.6	98.1	73.6	88.5	83.2	82.0	98.6	82.0	98.6	76.2	91.6
Chinese (S.)	83.2	82.8	99.5	81.6	98.1	73.6	88.5	83.2	82.0	98.6	82.0	98.6	76.2	91.6
T4	82.0	80.0	97.6	78.0	95.1	71.3	87.0	82.0	80.3	97.9	77.9	95.1	70.6	86.2
Italian	85.0	82.8	97.4	80.0	94.1	71.6	84.2	85.0	82.2	96.7	79.2	93.2	69.6	81.9
Vietnamese	84.2	81.0	96.2	80.4	95.5	74.4	88.4	84.2	83.2	98.8	79.8	94.8	74.4	88.4
Turkish	76.8	76.2	99.2	73.6	95.8	68.0	88.5	76.8	75.4	98.2	74.8	97.4	67.8	88.3
T3	76.2	74.2	97.5	72.4	95.1	66.3	86.8	76.2	74.5	97.8	72.8	95.5	67.0	88.0
Indonesian	82.8	79.2	95.7	76.6	92.5	74.8	90.3	82.8	80.2	96.9	78.8	95.2	74.4	89.9
Estonian	78.8	76.6	97.2	74.8	94.9	65.8	83.5	78.8	76.4	97.0	73.6	93.4	64.8	82.2
Thai	73.4	71.2	97.0	70.4	95.9	64.4	87.7	73.4	71.6	97.5	70.6	96.2	65.8	89.6
Tamil	70.0	70.0	100.0	67.8	96.9	60.0	85.7	70.0	69.8	99.7	68.0	97.1	63.2	90.3
T2	70.9	69.2	97.6	65.8	92.8	59.3	83.4	70.9	69.6	98.1	67.1	94.7	60.4	85.2
Swahili	75.4	73.6	97.6	69.8	92.6	65.2	86.5	75.4	74.4	98.7	70.4	93.4	64.2	85.1
Haitian Creole	66.4	64.8	97.6	61.8	93.1	53.4	80.4	66.4	64.8	97.6	63.8	96.1	56.6	85.2
T1	54.2	52.4	96.7	51.8	95.6	52.8	97.4	54.2	51.8	95.6	53.2	98.2	52.0	95.9
Guarani	54.2	52.4	96.7	51.8	95.6	52.8	97.4	54.2	51.8	95.6	53.2	98.2	52.0	95.9
Avg	75.5	73.7	97.6	71.7	95.0	65.8	87.4	75.5	73.8	97.7	72.2	95.8	66.3	88.0

Table 35: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4,3,2\}$ for Gemma4 on XCOPA under top-k reduction and global α. Per-language mean within tier.

P Detailed Results: XStoryCloze

Qwen3 – XStoryCloze																					
	Standard							Global α							Token α						
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	71.1	67.1	94.3	62.0	87.2	52.2	73.5	71.1	68.2	95.8	66.3	93.2	61.4	86.2	71.1	68.6	96.5	66.0	92.8	60.5	84.9
English	79.4	75.5	95.1	69.8	87.8	58.3	73.4	79.4	76.6	96.4	74.7	94.1	70.4	88.6	79.4	77.4	97.5	74.3	93.6	70.3	88.5
Spanish	71.5	67.4	94.2	61.2	85.6	51.0	71.3	71.5	68.4	95.6	67.0	93.7	60.8	84.9	71.5	69.0	96.4	66.4	92.9	59.8	83.5
Chinese (S.)	70.9	65.4	92.2	60.2	84.9	50.4	71.0	70.9	68.3	96.3	65.9	92.9	61.2	86.3	70.9	68.0	95.8	65.5	92.4	60.4	85.1
Arabic	62.5	60.0	96.0	56.7	90.6	48.9	78.2	62.5	59.4	95.0	57.6	92.2	53.3	85.2	62.5	60.2	96.3	57.7	92.3	51.5	82.3
T4	63.6	62.3	98.1	57.8	91.3	49.9	79.1	63.6	62.3	97.9	60.1	94.6	56.1	88.5	63.6	62.6	98.4	60.5	95.2	55.8	88.0
Russian	70.5	68.2	96.6	61.6	87.3	50.9	72.1	70.5	68.9	97.7	66.9	94.8	60.9	86.3	70.5	69.4	98.3	67.0	94.9	60.7	86.0
Hindi	65.2	63.3	97.1	58.6	89.8	49.7	76.2	65.2	63.5	97.5	60.1	92.2	55.9	85.7	65.2	63.5	97.4	61.1	93.7	54.5	83.7
Basque	55.2	55.5	100.6	53.3	96.6	49.0	88.8	55.2	54.4	98.6	53.4	96.8	51.7	93.6	55.2	55.0	99.6	53.5	97.0	52.1	94.4
T3	69.5	65.7	94.6	63.3	91.1	49.0	70.5	69.5	67.3	96.9	64.9	93.3	59.6	85.8	69.5	67.3	96.9	64.8	93.2	57.6	83.0
Indonesian	69.5	65.7	94.6	63.3	91.1	49.0	70.5	69.5	67.3	96.9	64.9	93.3	59.6	85.8	69.5	67.3	96.9	64.8	93.2	57.6	83.0
T2	54.2	52.6	97.1	50.3	92.8	46.7	86.2	54.2	53.3	98.3	52.0	95.8	50.4	92.9	54.2	53.3	98.3	52.3	96.6	50.6	93.4
Swahili	54.2	52.6	97.1	50.3	92.8	46.7	86.2	54.2	53.3	98.3	52.0	95.8	50.4	92.9	54.2	53.3	98.3	52.3	96.6	50.6	93.4
T1	60.1	57.7	95.9	54.8	91.3	50.8	84.6	60.1	58.1	96.6	56.0	93.1	53.7	89.6	60.1	57.5	95.7	55.9	92.9	52.9	88.1
Telugu	63.1	61.9	98.1	57.2	90.7	52.5	83.3	63.1	62.5	99.2	59.5	94.3	55.5	88.0	63.1	61.9	98.2	59.4	94.1	55.1	87.4
Burmese	57.0	53.5	93.7	52.4	91.9	49.0	85.8	57.0	53.6	94.0	52.4	91.9	52.0	91.1	57.0	53.1	93.2	52.3	91.8	50.6	88.7
Avg	65.4	62.6	95.9	58.6	89.9	50.5	77.9	65.4	63.3	96.8	61.3	93.8	57.4	88.0	65.4	63.5	97.1	61.3	93.9	56.7	86.9

Table 36: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4,3,2\}$ for Qwen3 on XStoryCloze under top-k reduction, global α, and token α. Per-language mean within tier. Qwen3 is restricted to its 73 supported languages.

Qwen3.5 – XStoryCloze																								
	Standard							Global α							Token α									
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	73.2	70.2	95.9	67.0	91.5	59.5	81.3	73.2	71.8	98.1	70.8	96.6	67.8	92.5	73.2	71.6	97.8	70.7	96.5	68.0	92.8
English	79.6	77.2	97.0	73.9	92.9	67.2	84.4	79.6	79.2	99.4	78.6	98.8	76.4	95.9	79.6	79.0	99.3	78.6	98.7	77.1	96.8
Spanish	74.7	70.5	94.5	65.7	87.9	57.0	76.3	74.7	73.3	98.1	71.1	95.3	67.0	89.8	74.7	71.9	96.4	71.2	95.4	66.9	89.6
Arabic	69.6	67.2	96.6	64.6	92.8	54.7	78.6	69.6	67.4	96.8	66.4	95.3	62.5	89.8	69.6	67.5	97.0	66.2	95.2	63.0	90.5
Chinese (S.)	69.0	65.9	95.5	63.8	92.4	59.2	85.7	69.0	67.6	97.9	67.0	97.0	65.3	94.6	69.0	68.0	98.6	66.9	96.9	65.1	94.2
T4	69.4	67.2	96.9	63.4	91.5	56.1	81.0	69.4	68.0	98.0	66.6	96.0	63.2	91.2	69.4	68.0	98.0	65.8	94.9	63.5	91.6
Russian	75.2	72.5	96.3	68.2	90.7	58.7	78.0	75.2	73.9	98.2	72.5	96.4	67.0	89.1	75.2	73.8	98.1	70.9	94.3	67.7	90.0
Hindi	68.8	66.5	96.6	61.4	89.2	53.5	77.7	68.8	67.6	98.2	65.5	95.1	63.4	92.1	68.8	67.6	98.2	65.5	95.1	64.3	93.5
Basque	64.1	62.7	97.7	60.7	94.6	56.1	87.4	64.1	62.6	97.6	61.9	96.6	59.3	92.5	64.1	62.6	97.6	61.1	95.3	58.5	91.2
T3	72.1	70.2	97.2	66.1	91.7	58.6	81.2	72.1	71.5	99.2	69.8	96.7	66.9	92.8	72.1	71.4	99.0	69.4	96.1	65.8	91.2
Indonesian	72.1	70.2	97.2	66.1	91.7	58.6	81.2	72.1	71.5	99.2	69.8	96.7	66.9	92.8	72.1	71.4	99.0	69.4	96.1	65.8	91.2
T2	62.7	60.6	96.5	58.3	92.9	52.6	83.9	62.7	61.1	97.4	59.2	94.4	57.7	92.0	62.7	61.1	97.4	58.8	93.8	57.1	91.0
Swahili	62.7	60.6	96.5	58.3	92.9	52.6	83.9	62.7	61.1	97.4	59.2	94.4	57.7	92.0	62.7	61.1	97.4	58.8	93.8	57.1	91.0
T1	65.3	63.1	96.7	60.2	92.3	54.9	84.2	65.3	64.0	98.1	62.9	96.4	60.7	93.0	65.3	64.2	98.3	62.3	95.5	60.1	92.0
Telugu	68.7	66.6	96.9	62.8	91.4	57.0	82.9	68.7	67.8	98.7	66.3	96.5	64.3	93.5	68.7	68.0	98.9	65.7	95.6	63.6	92.6
Burmese	61.8	59.6	96.5	57.6	93.1	52.8	85.4	61.8	60.3	97.5	59.6	96.4	57.1	92.4	61.8	60.4	97.8	59.0	95.5	56.5	91.4
Avg	69.7	67.2	96.5	63.9	91.8	57.0	82.0	69.7	68.4	98.1	67.1	96.2	64.3	92.2	69.7	68.3	98.0	66.7	95.6	64.1	92.0

Table 37: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Qwen3.5 on XStoryCloze under top-k reduction, global α, and token α. Per-language mean within tier.

Gemma4 – XStoryCloze														
	Standard							Global α						
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	77.0	74.7	97.0	72.1	93.7	65.0	84.4	77.0	75.0	97.5	72.3	93.9	65.6	85.2
English	84.2	82.3	97.7	79.8	94.8	73.5	87.3	84.2	82.7	98.3	80.1	95.2	73.4	87.2
Spanish	78.9	76.1	96.5	73.1	92.7	66.6	84.4	78.9	76.0	96.4	73.5	93.1	66.8	84.7
Chinese (S.)	73.3	71.3	97.2	68.3	93.1	59.6	81.3	73.3	71.9	98.1	68.0	92.8	61.4	83.7
Arabic	71.5	69.2	96.8	67.2	94.0	60.4	84.5	71.5	69.4	97.1	67.4	94.4	61.0	85.3
T4	73.2	70.9	96.7	68.0	92.9	62.5	85.4	73.2	70.8	96.7	68.5	93.5	62.4	85.2
Russian	78.9	77.4	98.1	73.1	92.7	66.5	84.3	78.9	77.0	97.7	74.8	94.8	66.4	84.2
Hindi	70.8	67.8	95.7	67.0	94.6	61.8	87.3	70.8	68.9	97.3	66.8	94.3	62.4	88.1
Basque	70.0	67.5	96.4	63.9	91.3	59.2	84.5	70.0	66.6	95.1	63.9	91.3	58.3	83.3
T3	76.6	74.1	96.8	70.7	92.3	65.0	84.9	76.6	74.0	96.6	71.3	93.1	65.7	85.7
Indonesian	76.6	74.1	96.8	70.7	92.3	65.0	84.9	76.6	74.0	96.6	71.3	93.1	65.7	85.7
T2	71.0	67.2	94.7	66.2	93.3	59.4	83.6	71.0	67.7	95.3	66.6	93.8	61.2	86.2
Swahili	71.0	67.2	94.7	66.2	93.3	59.4	83.6	71.0	67.7	95.3	66.6	93.8	61.2	86.2
T1	67.7	65.6	96.9	63.0	93.1	57.8	85.3	67.7	65.9	97.4	62.8	92.8	58.4	86.3
Telugu	70.7	67.2	95.1	65.0	91.9	60.7	85.9	70.7	67.9	96.1	64.9	91.8	60.4	85.4
Burmese	64.7	63.9	98.7	61.0	94.3	54.9	84.8	64.7	63.9	98.8	60.8	93.9	56.5	87.2
Avg	73.7	71.3	96.7	68.7	93.2	62.5	84.8	73.7	71.5	97.0	68.9	93.5	63.0	85.5

Table 38: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Gemma4 on XStoryCloze under top-k reduction and global α. Per-language mean within tier.

Q Detailed Results: Translation

	Qwen3 – Translation																						
	Standard							Global α							Token α								
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	63.6	60.2	94.5	50.9	79.9	12.4	19.5	63.6	62.7	98.5	61.4	96.4	56.7	89.0	63.6	62.6	98.3	61.2	96.1	56.6	88.8
French	68.1	66.2	97.2	59.7	87.6	16.7	24.5	68.1	67.7	99.4	67.2	98.7	63.3	93.0	68.1	67.9	99.7	66.9	98.2	64.0	93.9
German	68.1	65.1	95.6	56.3	82.6	12.5	18.3	68.1	67.5	99.1	66.8	98.0	61.8	90.7	68.1	67.3	98.8	66.5	97.7	61.3	90.0
Arabic	66.6	61.7	92.5	46.1	69.2	10.6	15.8	66.6	64.3	96.5	63.0	94.5	55.1	82.7	66.6	65.0	97.6	62.9	94.4	56.4	84.6
Spanish	61.5	58.9	95.8	52.9	86.0	13.4	21.8	61.5	60.9	99.0	58.9	95.8	56.7	92.3	61.5	60.2	98.0	59.1	96.2	56.0	91.2
Chinese (S.)	59.6	55.7	93.4	46.4	77.9	12.2	20.4	59.6	59.1	99.2	57.5	96.5	53.1	89.1	59.6	58.1	97.5	56.6	94.9	52.3	87.7
Japanese	57.9	53.5	92.3	44.2	76.3	9.4	16.2	57.9	56.6	97.7	54.9	94.7	49.9	86.1	57.9	56.9	98.1	55.1	95.2	49.5	85.5
T4	63.7	59.2	93.0	49.0	76.9	11.8	18.5	63.7	62.3	97.8	60.4	94.8	53.5	83.8	63.7	62.4	97.9	60.7	95.2	53.2	83.3
Portuguese	71.7	67.6	94.2	60.8	84.8	16.7	23.3	71.7	71.1	99.1	70.1	97.8	66.7	92.9	71.7	71.1	99.1	70.4	98.1	66.6	92.9
Catalan	71.0	66.4	93.5	56.2	79.1	13.0	18.3	71.0	69.4	97.6	67.9	95.6	62.8	88.3	71.0	69.8	98.3	68.1	95.9	62.2	87.5
Swedish	69.8	66.5	95.3	56.4	80.8	13.5	19.4	69.8	68.9	98.7	67.3	96.4	62.6	89.6	69.8	69.0	98.8	67.6	96.8	59.9	85.7
Serbian	67.6	63.0	93.2	52.4	77.5	11.0	16.3	67.6	66.2	98.0	64.8	95.9	54.4	80.5	67.6	66.3	98.1	64.8	95.9	58.2	86.1
Czech	65.8	62.9	95.7	52.5	79.8	11.5	17.4	65.8	64.3	97.8	63.5	96.5	56.5	85.9	65.8	64.6	98.2	63.7	96.9	58.9	89.6
Hindi	64.4	59.4	92.3	43.9	68.2	9.0	14.0	64.4	63.2	98.3	60.0	93.2	50.8	78.9	64.4	63.1	98.1	61.3	95.2	43.4	67.3
Croatian	64.2	59.6	92.9	48.5	75.6	10.6	16.5	64.2	63.1	98.4	61.3	95.5	55.1	85.8	64.2	63.3	98.7	61.0	95.1	53.7	83.7
Turkish	63.8	57.3	89.9	44.0	69.0	10.9	17.1	63.8	62.6	98.2	60.3	94.6	48.0	75.2	63.8	62.2	97.6	59.9	94.0	50.3	78.8
Hungarian	63.3	58.7	92.8	46.5	73.4	12.1	19.1	63.3	62.2	98.3	59.7	94.3	51.9	82.0	63.3	62.0	98.0	58.0	91.7	52.6	83.1
Italian	63.1	61.0	96.7	55.5	87.9	13.8	21.9	63.1	62.5	99.0	62.1	98.3	58.0	91.8	63.1	62.0	98.1	61.9	98.0	58.2	92.2
Vietnamese	63.0	58.9	93.4	50.0	79.3	10.1	16.0	63.0	61.3	97.2	59.9	95.1	54.4	86.3	63.0	61.7	97.9	60.2	95.4	56.2	89.1
Russian	63.0	60.8	96.5	51.6	82.0	10.7	17.0	63.0	62.0	98.4	61.0	96.9	56.4	89.5	63.0	62.3	98.9	61.5	97.6	55.8	88.6
Persian	62.1	55.4	89.1	42.8	68.9	10.7	17.2	62.1	60.0	96.5	57.2	92.0	46.8	75.2	62.1	60.4	97.3	57.8	93.0	45.8	73.7
Finnish	60.8	55.8	91.8	44.6	73.4	10.7	17.6	60.8	58.9	96.9	56.9	93.6	48.8	80.2	60.8	59.2	97.4	57.1	93.9	46.0	75.6
Korean	59.6	53.3	89.5	46.2	77.5	13.1	22.0	59.6	57.8	97.0	55.0	92.4	49.3	82.8	59.6	57.9	97.2	56.5	94.8	48.4	81.2
Dutch	59.1	55.9	94.5	47.2	79.7	12.5	21.2	59.1	58.4	98.8	57.1	96.6	54.1	91.5	59.1	58.5	99.0	57.8	97.8	55.0	93.0
Polish	58.5	56.5	96.5	48.1	82.3	11.7	19.9	58.5	57.6	98.4	55.9	95.5	51.8	88.4	58.5	57.1	97.6	56.6	96.7	52.0	88.8
Basque	54.9	47.4	86.4	35.5	64.6	10.1	18.4	54.9	51.5	94.0	47.3	86.2	35.2	64.2	54.9	52.0	94.7	48.1	87.7	34.1	62.1
T3	63.7	57.7	90.4	45.0	70.0	10.1	15.8	63.7	61.8	97.0	58.8	92.1	48.7	75.8	63.7	61.8	97.0	58.9	92.2	45.2	70.0
Afrikaans	75.7	70.0	92.4	57.9	76.4	12.4	16.4	75.7	74.3	98.1	71.5	94.4	64.0	84.5	75.7	74.3	98.2	71.9	94.9	61.0	80.6
Danish	71.9	68.1	94.7	56.2	78.1	12.1	16.8	71.9	70.2	97.6	69.6	96.8	62.5	87.0	71.9	70.1	97.5	69.5	96.7	61.8	85.9
Malay	69.3	64.2	92.6	55.3	79.7	11.8	17.1	69.3	68.0	98.1	65.1	93.9	57.3	82.6	69.3	68.1	98.3	65.6	94.6	58.0	83.7
Romanian	68.6	65.5	95.4	55.3	80.6	12.9	18.8	68.6	67.8	98.8	67.1	97.8	61.1	89.1	68.6	67.9	98.9	66.9	97.5	60.6	88.3
Hebrew	67.7	61.3	90.5	49.7	73.4	12.2	18.1	67.7	64.9	95.9	62.6	92.4	55.0	81.2	67.7	64.6	95.4	62.6	92.4	43.5	64.3
Indonesian	67.4	63.0	93.5	56.0	83.2	12.9	19.2	67.4	66.1	98.1	65.1	96.7	60.5	89.8	67.4	66.0	98.0	64.6	95.9	60.2	89.4
Bulgarian	67.2	63.7	94.8	52.7	78.5	9.3	13.8	67.2	66.5	99.0	64.7	96.4	55.7	82.9	67.2	66.1	98.4	64.8	96.6	57.5	85.6
Filipino	67.1	60.6	90.4	50.2	74.9	12.0	17.9	67.1	65.6	97.8	63.0	93.9	52.1	77.8	67.1	65.8	98.0	62.3	92.9	51.6	77.0
Slovak	65.2	61.1	93.8	50.5	77.5	12.4	19.1	65.2	64.6	99.1	62.3	95.6	55.4	85.0	65.2	64.5	98.9	62.8	96.4	54.5	83.7
Cebuano	64.7	59.8	92.3	42.9	66.3	11.8	18.2	64.7	62.9	97.2	58.1	89.8	46.5	71.8	64.7	63.1	97.5	58.6	90.5	48.9	75.5
Ukrainian	64.6	60.7	93.9	50.5	78.1	9.2	14.3	64.6	63.2	97.9	62.0	96.0	55.9	86.5	64.6	63.6	98.4	61.9	95.8	31.9	49.4
Latvian	63.5	57.3	90.3	45.7	72.0	11.3	17.7	63.5	62.3	98.1	59.1	93.1	49.0	77.2	63.5	62.8	99.0	59.1	93.1	48.5	76.4
Greek	62.0	58.5	94.3	44.2	71.3	7.8	12.6	62.0	60.9	98.2	58.6	94.5	49.8	80.3	62.0	60.5	97.5	58.2	93.9	45.3	73.1
Slovenian	61.9	57.5	92.8	46.9	75.7	12.6	20.3	61.9	60.3	97.4	58.3	94.1	50.0	80.7	61.9	60.1	97.0	58.0	93.6	50.2	81.0
Estonian	61.4	55.8	90.8	42.1	68.6	11.0	17.9	61.4	59.0	96.1	56.7	92.4	48.2	78.5	61.4	59.0	96.2	56.9	92.7	48.5	79.1
Lithuanian	60.8	55.4	91.2	44.8	73.6	10.6	17.3	60.8	59.5	97.8	56.3	92.6	48.3	79.4	60.8	58.6	96.4	56.2	92.4	45.8	75.3
Bengali	60.3	51.5	85.3	38.8	64.2	7.5	12.4	60.3	57.6	95.6	53.6	88.8	41.2	68.3	60.3	57.0	94.5	54.5	90.4	36.5	60.5
Thai	60.2	50.3	83.6	37.9	63.0	5.4	9.0	60.2	58.1	96.6	54.9	91.2	41.4	68.8	60.2	58.8	97.7	55.2	91.8	43.2	71.8
Urdu	59.8	52.4	87.6	36.8	61.4	7.3	12.3	59.8	56.3	94.2	51.2	85.6	38.4	64.2	59.8	56.7	94.8	53.0	88.5	38.0	63.6
Kazakh	58.3	50.2	86.0	32.4	55.6	5.8	10.0	58.3	54.4	93.3	48.4	83.0	35.8	61.3	58.3	54.5	93.4	49.5	84.8	32.0	54.9
Tamil	56.7	46.7	82.4	30.0	53.0	7.2	12.7	56.7	54.2	95.6	48.8	86.1	31.9	56.2	56.7	54.5	96.0	49.2	86.8	23.1	40.7
Uzbek	56.1	45.7	81.5	31.6	56.3	10.1	18.1	56.1	53.7	95.7	49.6	88.5	33.1	59.1	56.1	53.0	94.5	50.1	89.3	15.3	27.2
Georgian	53.8	47.3	87.9	26.7	49.6	7.6	14.2	53.8	50.7	94.3	45.2	84.0	27.1	50.5	53.8	51.3	95.4	42.3	78.7	22.9	42.7

Table 39: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Qwen3 on Translation under top-k reduction, global α, and token α. Per-language mean within tier. Qwen3 is restricted to its 73 supported languages. Part 1: tiers T5–T3.

Qwen3 – Translation																								
	Standard							Global α							Token α									
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T2	56.8	48.2	84.9	31.7	55.7	9.6	17.1	56.8	53.3	93.8	49.2	86.6	34.2	60.0	56.8	53.0	93.4	48.9	86.0	32.8	57.7
Maltese	62.5	53.8	86.1	35.6	56.9	10.5	16.8	62.5	59.0	94.4	54.2	86.7	38.1	60.9	62.5	57.6	92.2	53.5	85.6	34.9	55.8
Punjabi	61.6	51.6	83.7	35.1	57.0	9.1	14.7	61.6	59.2	96.1	54.1	87.7	39.2	63.6	61.6	58.7	95.2	54.9	89.1	32.1	52.1
Marathi	60.2	53.5	88.8	37.2	61.8	10.1	16.8	60.2	57.0	94.7	53.5	88.8	39.5	65.6	60.2	57.6	95.7	52.8	87.6	36.9	61.3
Lao	56.6	47.4	83.8	26.2	46.3	4.0	7.0	56.6	53.2	94.1	50.3	88.9	31.2	55.2	56.6	53.4	94.5	49.5	87.4	32.2	56.9
Icelandic	54.2	47.7	88.0	34.4	63.4	11.5	21.2	54.2	50.5	93.1	46.4	85.5	36.9	67.9	54.2	50.5	93.0	47.2	87.1	39.0	71.9
Haitian Creole	51.8	43.8	84.6	28.7	55.5	11.1	21.4	51.8	48.6	93.8	44.6	86.1	30.7	59.2	51.8	47.9	92.4	44.4	85.6	31.8	61.3
Swahili	50.2	39.7	79.1	24.6	49.1	10.8	21.5	50.2	45.4	90.3	41.4	82.4	24.0	47.8	50.2	45.6	90.7	39.9	79.4	22.5	44.8
T1	57.3	48.8	84.7	32.9	56.1	7.7	13.3	57.3	54.3	94.8	50.2	87.3	34.8	59.4	57.3	54.4	94.8	50.2	87.1	30.8	52.7
Norwegian	74.9	70.6	94.2	56.3	75.2	13.0	17.4	74.9	73.2	97.8	72.2	96.5	66.6	88.9	74.9	74.2	99.1	72.2	96.4	67.4	90.0
Macedonian	66.5	61.5	92.5	49.3	74.1	10.4	15.6	66.5	64.6	97.1	62.4	93.7	53.8	80.9	66.5	64.5	96.9	62.6	94.1	52.8	79.3
Albanian	63.6	58.2	91.5	41.9	65.9	13.0	20.4	63.6	60.5	95.1	58.0	91.1	50.9	80.1	63.6	61.3	96.3	58.3	91.6	35.8	56.2
Nepali	62.1	53.2	85.7	38.9	62.7	8.4	13.5	62.1	58.3	93.8	53.6	86.4	39.9	64.3	62.1	58.3	93.8	54.0	86.9	26.3	42.4
Gujarati	62.1	53.4	86.1	31.7	51.1	6.3	10.1	62.1	59.3	95.6	55.9	90.0	36.8	59.2	62.1	58.9	94.9	55.9	90.1	21.9	35.3
Armenian	61.6	52.2	84.7	34.5	55.9	4.6	7.5	61.6	59.7	96.9	54.9	89.0	39.3	63.8	61.6	59.1	95.9	54.5	88.5	34.0	55.3
Telugu	60.9	47.1	77.3	32.7	53.7	5.0	8.2	60.9	57.0	93.6	52.4	86.2	21.7	35.6	60.9	55.9	91.8	53.3	87.5	20.5	33.7
Malayalam	59.5	49.9	83.8	32.8	55.2	7.0	11.8	59.5	55.7	93.6	52.1	87.5	32.0	53.7	59.5	57.4	96.4	53.4	89.8	31.1	52.2
Javanese	59.1	51.3	86.8	38.9	65.8	11.0	18.7	59.1	57.1	96.5	53.4	90.4	41.9	70.9	59.1	56.1	94.9	52.1	88.1	39.4	66.7
Kannada	57.6	49.6	86.1	35.0	60.7	6.7	11.7	57.6	54.5	94.6	50.8	88.1	30.2	52.4	57.6	54.4	94.3	50.2	87.1	31.1	53.9
Sundanese	56.0	48.6	86.8	35.5	63.3	10.0	17.8	56.0	52.0	92.8	48.8	87.1	35.8	63.9	56.0	51.5	91.9	47.8	85.3	34.3	61.1
Oriya	54.3	44.2	81.4	27.9	51.3	5.5	10.1	54.3	50.7	93.3	45.4	83.6	27.5	50.7	54.3	50.2	92.4	46.1	84.9	26.8	49.3
Assamese	53.2	42.6	80.1	28.7	54.0	7.2	13.6	53.2	49.1	92.3	43.5	81.8	26.9	50.5	53.2	49.4	92.9	44.6	83.9	23.8	44.8
Azerbaijani	52.4	45.3	86.5	34.6	66.1	8.7	16.6	52.4	50.5	96.4	47.6	90.8	34.3	65.5	52.4	49.9	95.2	46.9	89.5	30.5	58.3
Khmer	50.4	41.9	83.2	16.0	31.7	4.0	8.0	50.4	47.8	94.9	40.2	79.7	26.0	51.6	50.4	50.0	99.2	44.3	87.9	24.5	48.6
Tajik	49.9	40.9	82.1	20.3	40.7	7.2	14.5	49.9	44.8	89.9	39.5	79.2	21.5	43.0	49.9	44.0	88.3	37.6	75.5	16.3	32.7
Sinhala	48.6	37.0	76.0	24.0	49.4	8.1	16.7	48.6	44.0	90.5	38.0	78.1	21.5	44.1	48.6	44.4	91.3	37.3	76.6	18.8	38.7
Burmese	38.5	31.1	80.6	12.9	33.5	2.6	6.9	38.5	39.1	101.5	35.5	92.2	19.3	50.0	38.5	39.0	101.1	32.2	83.6	19.3	50.2
Avg	61.4	55.1	89.4	42.2	67.7	10.1	16.3	61.4	59.3	96.5	56.3	91.4	45.7	73.3	61.4	59.3	96.4	56.4	91.4	43.3	69.4

Table 40: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Qwen3 on Translation under top-k reduction, global α, and token α. Per-language mean within tier. Qwen3 is restricted to its 73 supported languages. Part 2: tiers T2–T1 and overall mean.

Qwen3.5 – Translation																							
	Standard							Global α							Token α								
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	63.6	59.4	93.3	51.9	81.6	31.7	49.7	63.6	63.3	99.5	62.7	98.6	60.6	95.2	63.6	63.4	99.7	63.0	99.0	61.5	96.7
German	69.5	65.6	94.4	57.2	82.3	36.0	51.7	69.5	69.2	99.6	68.5	98.5	67.2	96.6	69.5	69.3	99.8	68.6	98.7	67.8	97.6
French	68.4	66.7	97.5	58.8	85.9	43.1	63.1	68.4	68.4	100.0	67.9	99.3	66.7	97.6	68.4	68.3	99.8	67.3	98.5	66.6	97.4
Arabic	66.8	63.2	94.6	51.9	77.7	26.1	39.0	66.8	66.1	98.9	66.3	99.1	64.2	96.0	66.8	66.5	99.6	66.2	99.0	63.7	95.3
Spanish	60.2	58.9	97.8	55.6	92.4	34.3	57.0	60.2	60.3	100.2	60.1	99.8	59.7	99.2	60.2	60.3	100.2	60.2	100.0	59.7	99.1
Chinese (S.)	59.8	52.2	87.2	44.4	74.1	23.8	39.8	59.8	58.5	97.7	56.9	95.1	54.9	91.8	59.8	58.8	98.3	58.2	97.3	57.8	96.6
Japanese	56.8	50.0	88.2	43.8	77.2	26.9	47.3	56.8	57.2	100.8	56.7	99.9	51.0	89.9	56.8	57.2	100.7	57.2	100.7	53.6	94.3
T4	64.5	60.9	94.3	52.5	81.1	31.3	48.3	64.5	64.3	99.7	63.8	98.9	61.9	95.9	64.5	64.3	99.6	63.8	98.8	62.4	96.6
Portuguese	72.5	69.7	96.2	56.4	77.8	39.5	54.6	72.5	72.1	99.5	71.5	98.6	69.8	96.4	72.5	72.3	99.7	71.5	98.6	70.7	97.6
Catalan	71.7	69.9	97.5	63.5	88.5	37.9	52.9	71.7	71.2	99.4	70.5	98.3	69.1	96.5	71.7	71.1	99.2	70.7	98.6	69.4	96.8
Swedish	70.1	67.3	96.0	63.6	90.8	39.6	56.5	70.1	70.5	100.6	70.1	99.9	67.8	96.7	70.1	70.1	100.0	69.6	99.2	68.8	98.1
Serbian	69.3	67.0	96.7	55.2	79.7	30.6	44.2	69.3	68.8	99.3	68.8	99.3	66.7	96.3	69.3	68.5	98.8	68.1	98.3	66.2	95.6
Czech	67.1	64.3	95.8	56.4	84.2	35.9	53.5	67.1	66.8	99.6	66.4	99.0	65.2	97.2	67.1	66.9	99.8	66.2	98.7	65.0	97.0
Persian	64.8	60.9	93.9	55.3	85.4	23.6	36.4	64.8	63.3	97.7	62.6	96.7	61.0	94.2	64.8	63.8	98.4	62.7	96.8	61.3	94.7
Hindi	64.7	62.0	96.0	53.1	82.1	31.6	48.8	64.7	65.0	100.5	63.7	98.5	62.3	96.4	64.7	65.0	100.6	63.5	98.2	62.3	96.4
Croatian	64.6	59.1	91.6	54.5	84.4	29.1	45.0	64.6	64.7	100.1	64.8	100.3	63.0	97.6	64.6	64.6	100.0	64.4	99.7	62.7	97.1
Turkish	64.4	58.9	91.4	50.0	77.7	27.9	43.3	64.4	64.8	100.5	64.7	100.4	62.2	96.5	64.4	64.1	99.5	64.4	100.0	61.8	95.9
Russian	63.6	61.2	96.2	49.8	78.3	32.3	50.8	63.6	63.5	99.9	63.4	99.7	61.5	96.7	63.6	63.5	99.9	62.7	98.6	61.9	97.4
Hungarian	63.6	57.4	90.3	52.3	82.3	31.6	49.6	63.6	63.5	99.9	63.3	99.6	60.8	95.7	63.6	63.9	100.5	63.0	99.1	61.6	96.9
Vietnamese	63.2	58.9	93.1	44.4	70.2	28.9	45.7	63.2	63.5	100.4	62.9	99.4	61.4	97.1	63.2	63.6	100.6	62.7	99.2	61.6	97.5
Italian	62.7	60.4	96.3	55.1	88.0	37.0	59.1	62.7	62.5	99.6	62.2	99.3	61.7	98.4	62.7	62.7	100.1	62.6	99.9	61.3	97.8
Finnish	62.3	54.9	88.2	50.3	80.7	28.1	45.1	62.3	62.2	100.0	60.9	97.8	57.8	92.9	62.3	60.7	97.4	61.5	98.7	59.2	95.2
Korean	60.3	54.5	90.3	43.5	72.1	28.9	48.0	60.3	60.2	99.8	59.2	98.1	55.7	92.4	60.3	59.8	99.1	59.3	98.3	57.2	94.9
Dutch	59.6	58.0	97.3	48.4	81.2	34.3	57.6	59.6	59.4	99.7	59.2	99.3	58.6	98.3	59.6	59.3	99.5	59.3	99.5	58.7	98.4
Polish	59.4	57.9	97.5	47.4	79.8	19.7	33.2	59.4	58.8	98.9	58.4	98.3	56.5	95.1	59.4	58.6	98.7	58.3	98.1	57.6	96.9
Basque	58.1	54.2	93.3	44.9	77.3	26.4	45.4	58.1	57.3	98.5	56.9	98.0	53.8	92.5	58.1	58.2	100.1	57.8	99.4	55.3	95.2
T3	65.0	60.2	92.6	53.0	81.4	27.7	42.2	65.0	64.6	99.4	63.9	98.4	61.6	94.8	65.0	64.5	99.2	63.8	98.2	62.0	95.4
Afrikaans	76.8	74.7	97.3	69.0	89.9	33.8	44.0	76.8	76.6	99.7	76.3	99.3	74.4	96.8	76.8	76.7	99.8	76.1	99.0	75.5	98.2
Danish	72.5	68.2	94.1	61.1	84.4	42.7	58.9	72.5	71.9	99.2	71.2	98.3	70.0	96.6	72.5	70.8	97.7	70.7	97.5	69.7	96.1
Malay	70.5	68.7	97.4	63.9	90.7	38.2	54.2	70.5	70.1	99.5	69.1	98.0	67.9	96.3	70.5	69.8	98.9	69.2	98.1	68.1	96.6
Romanian	69.5	61.0	87.7	56.9	81.8	27.8	40.0	69.5	69.3	99.7	68.5	98.5	67.5	97.1	69.5	69.4	99.9	68.6	98.6	67.6	97.3
Hebrew	68.6	63.7	92.8	56.4	82.2	31.9	46.6	68.6	68.5	99.9	67.9	98.9	65.6	95.6	68.6	68.1	99.2	67.7	98.7	66.1	96.4
Filipino	68.4	63.2	92.5	55.2	80.7	24.0	35.2	68.4	67.0	97.9	65.4	95.7	63.0	92.2	68.4	67.2	98.3	66.1	96.7	64.2	93.9
Indonesian	68.2	66.1	96.9	58.5	85.8	37.3	54.7	68.2	68.2	99.9	67.2	98.4	66.8	97.9	68.2	68.0	99.7	67.5	98.9	66.8	97.9
Bulgarian	67.8	65.6	96.7	59.4	87.6	35.0	51.6	67.8	68.4	100.9	67.6	99.7	66.5	98.1	67.8	68.0	100.2	67.7	99.9	65.6	96.8
Slovak	66.3	62.4	94.2	54.3	82.0	31.5	47.5	66.3	66.2	100.0	66.1	99.8	63.8	96.2	66.3	65.3	98.5	65.7	99.2	64.1	96.8
Cebuano	65.5	61.5	93.9	53.3	81.3	30.0	45.8	65.5	64.4	98.4	63.1	96.4	60.3	92.0	65.5	64.4	98.3	63.2	96.4	61.5	93.8
Ukrainian	65.5	56.5	86.3	53.9	82.3	33.9	51.7	65.5	64.7	98.8	64.4	98.3	62.2	95.0	65.5	65.3	99.8	64.5	98.5	62.4	95.3
Latvian	65.1	58.4	89.7	50.9	78.1	26.4	40.5	65.1	64.4	98.9	64.4	98.9	61.5	94.5	65.1	64.2	98.6	63.9	98.1	61.5	94.4
Estonian	63.9	56.3	88.2	52.3	81.8	20.6	32.2	63.9	63.2	98.9	63.6	99.6	59.2	92.8	63.9	63.6	99.7	62.6	98.0	60.6	94.9
Slovenian	63.8	60.0	94.1	52.2	81.9	23.4	36.6	63.8	63.0	98.8	62.3	97.6	60.7	95.1	63.8	63.9	100.2	62.9	98.7	60.5	94.8
Greek	62.4	57.1	91.4	52.2	83.7	26.3	42.2	62.4	62.5	100.1	61.8	99.0	61.6	98.7	62.4	62.5	100.1	61.4	98.3	61.0	97.7
Bengali	61.7	59.1	95.7	50.1	81.2	26.1	42.4	61.7	60.7	98.3	60.2	97.5	57.2	92.7	61.7	60.2	97.6	59.2	96.0	58.3	94.4
Lithuanian	61.5	58.8	95.6	50.7	82.4	24.4	39.7	61.5	61.3	99.5	60.3	98.0	58.1	94.4	61.5	60.8	98.8	59.4	96.5	57.6	93.5
Urdu	60.8	55.7	91.6	43.9	72.1	20.7	34.0	60.8	60.0	98.7	58.7	96.4	55.4	91.1	60.8	60.4	99.3	59.3	97.4	54.9	90.3
Thai	60.0	50.3	83.8	48.2	80.4	25.8	43.0	60.0	60.1	100.2	59.8	99.6	58.0	96.7	60.0	59.9	99.8	59.7	99.5	58.2	97.1
Uzbek	59.9	56.5	94.4	44.0	73.6	24.8	41.5	59.9	60.2	100.6	60.1	100.4	55.4	92.5	59.9	59.8	99.8	58.9	98.3	56.4	94.2
Kazakh	59.4	55.6	93.6	43.2	72.8	18.3	30.8	59.4	59.2	99.7	58.5	98.6	55.2	93.0	59.4	59.0	99.4	58.6	98.7	56.2	94.7
Georgian	58.2	52.9	90.9	45.7	78.5	17.0	29.2	58.2	57.6	99.1	56.6	97.4	53.5	91.9	58.2	58.0	99.6	57.1	98.1	55.2	94.8
Tamil	57.6	52.4	91.0	44.0	76.4	16.6	28.8	57.6	57.5	99.8	56.5	98.2	53.8	93.5	57.6	57.1	99.2	56.7	98.5	54.3	94.3

Table 41: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Qwen3.5 on Translation under top-k reduction, global α, and token α. Per-language mean within tier. Part 1: tiers T5–T3.

	Qwen3.5 – Translation																					
	Standard							Global α							Token α							
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T2	56.0	51.4	91.5	43.2	76.5	19.5	34.4	56.0	55.0	97.8	53.7	95.5	49.3	87.2	56.0	55.1	98.2	54.0	96.0	50.6	89.5
Maltese	74.7	73.0	97.7	62.8	84.1	29.6	39.6	74.7	75.0	100.4	74.6	99.9	71.6	95.8	74.7	74.5	99.8	74.3	99.4	72.1	96.6
Swahili	64.6	60.1	93.0	52.5	81.2	25.8	40.0	64.6	64.5	99.7	63.4	98.1	58.8	91.0	64.6	63.2	97.8	63.5	98.3	59.6	92.3
Punjabi	62.4	60.2	96.5	53.7	86.1	18.8	30.2	62.4	62.1	99.6	61.6	98.7	57.9	92.8	62.4	62.4	100.0	61.4	98.5	58.5	93.8
Haitian Creole	61.6	57.1	92.6	49.2	79.9	25.5	41.4	61.6	61.7	100.1	60.2	97.8	56.9	92.4	61.6	60.8	98.6	60.6	98.3	56.5	91.8
Marathi	61.2	57.9	94.6	48.8	79.6	23.8	38.9	61.2	60.9	99.5	59.9	97.8	56.8	92.7	61.2	60.6	99.0	59.5	97.1	57.4	93.8
Icelandic	60.2	55.6	92.2	50.1	83.2	24.0	39.9	60.2	59.7	99.0	58.2	96.7	55.4	91.9	60.2	59.6	99.0	58.4	97.0	57.0	94.6
Lao	59.1	53.1	89.8	40.4	68.3	13.6	23.0	59.1	58.7	99.3	58.5	99.0	55.0	93.0	59.1	58.9	99.7	58.3	98.6	55.5	93.9
Zulu	58.1	54.9	94.5	47.9	82.3	23.3	40.1	58.1	57.1	98.2	55.8	96.0	51.4	88.4	58.1	57.4	98.7	55.5	95.5	52.4	90.2
Xhosa	56.5	47.6	84.3	38.0	67.2	15.0	26.5	56.5	54.5	96.5	53.5	94.6	48.0	84.9	56.5	54.9	97.1	53.9	95.5	51.2	90.6
Amharic	54.5	47.5	87.0	37.6	68.9	12.2	22.4	54.5	53.1	97.4	51.0	93.5	44.4	81.4	54.5	54.0	99.0	51.7	94.9	46.4	85.1
Hausa	54.0	50.2	93.1	41.8	77.5	23.3	43.1	54.0	52.7	97.7	50.1	92.9	44.7	82.9	54.0	52.6	97.6	52.3	97.0	49.1	91.0
Tswana	48.2	42.1	87.2	33.8	70.0	16.3	33.7	48.2	46.9	97.2	44.1	91.3	39.6	82.1	48.2	47.1	97.7	45.7	94.8	41.1	85.2
Tigrinya	44.7	39.6	88.7	32.2	72.1	12.7	28.3	44.7	42.3	94.6	41.0	91.7	36.4	81.4	44.7	43.6	97.7	41.0	91.9	36.5	81.7
Yoruba	43.3	39.1	90.5	30.2	69.9	14.7	33.9	43.3	40.7	94.1	40.3	93.2	35.2	81.5	43.3	41.4	95.6	39.4	91.1	36.5	84.4
Wolof	37.1	33.6	90.7	28.4	76.6	13.2	35.6	37.1	34.8	93.9	33.7	90.8	28.0	75.6	37.1	35.5	95.6	34.1	91.9	29.0	78.2
T1	53.9	49.0	90.5	40.1	73.8	18.4	34.3	53.9	52.9	97.8	51.4	94.7	47.4	86.7	53.9	53.0	97.8	51.6	95.2	48.2	88.1
Norwegian	75.7	71.0	93.9	65.9	87.1	37.7	49.9	75.7	75.6	100.0	75.7	100.0	74.0	97.8	75.7	75.1	99.2	75.4	99.6	74.0	97.8
Macedonian	67.5	64.7	95.9	59.9	88.7	30.9	45.8	67.5	67.5	100.1	66.5	98.6	65.3	96.8	67.5	67.0	99.2	67.0	99.3	65.1	96.5
Albanian	66.6	62.3	93.6	53.8	80.8	27.1	40.6	66.6	66.0	99.1	65.9	98.9	63.8	95.8	66.6	66.6	100.0	65.8	98.9	64.0	96.0
Armenian	64.1	61.5	95.9	51.7	80.6	27.0	42.2	64.1	63.5	99.0	62.8	97.9	59.2	92.3	64.1	63.7	99.5	63.1	98.5	61.2	95.5
Gujarati	63.7	59.9	94.0	51.9	81.4	14.8	23.1	63.7	63.1	99.0	61.4	96.3	58.8	92.2	63.7	62.4	97.9	61.5	96.5	59.1	92.7
Javanese	63.3	58.0	91.7	48.5	76.6	24.0	37.9	63.3	61.2	96.8	60.6	95.8	56.9	90.0	63.3	62.3	98.5	60.8	96.1	58.0	91.7
Telugu	62.9	56.7	90.1	43.7	69.5	15.8	25.1	62.9	61.8	98.2	60.3	95.9	57.3	91.1	62.9	61.9	98.4	60.3	95.9	59.1	93.9
Nepali	62.3	59.4	95.4	47.3	76.0	21.9	35.1	62.3	63.2	101.4	61.7	99.1	59.1	94.9	62.3	62.9	100.9	61.3	98.4	58.4	93.8
Malayalam	61.9	56.6	91.4	47.8	77.3	13.7	22.2	61.9	61.9	100.1	60.2	97.3	56.6	91.5	61.9	61.7	99.8	60.2	97.4	57.6	93.1
Sundanese	61.0	55.4	90.7	50.2	82.2	21.2	34.7	61.0	60.3	98.7	58.8	96.3	56.6	92.7	61.0	59.8	98.0	58.9	96.4	56.9	93.3
Oriya	59.5	57.6	96.7	45.0	75.6	20.8	35.0	59.5	60.2	101.2	58.5	98.4	54.8	92.0	59.5	60.6	101.9	59.0	99.1	56.3	94.6
Sinhala	59.4	51.8	87.3	43.3	73.0	14.3	24.1	59.4	58.2	98.1	57.0	96.0	52.5	88.5	59.4	59.1	99.6	56.4	95.0	53.6	90.4
Tajik	59.4	56.6	95.4	49.3	83.0	20.2	34.0	59.4	59.1	99.6	57.6	97.1	53.6	90.3	59.4	59.6	100.5	57.5	97.0	54.9	92.5
Kannada	58.9	54.6	92.8	32.6	55.3	10.7	18.2	58.9	58.6	99.5	57.7	98.0	55.2	93.7	58.9	58.4	99.2	57.5	97.7	55.6	94.4
Khmer	56.9	53.1	93.5	35.0	61.5	15.8	27.8	56.9	56.6	99.6	55.2	97.1	50.8	89.3	56.9	56.7	99.8	56.4	99.2	52.3	92.0
Kurdish	56.6	50.7	89.5	36.7	64.8	15.8	27.9	56.6	55.3	97.6	52.3	92.5	48.8	86.3	56.6	55.0	97.2	53.8	95.1	48.6	85.9
Pashto	56.1	49.2	87.7	38.4	68.5	18.6	33.2	56.1	54.9	97.9	53.9	96.0	48.4	86.3	56.1	55.7	99.3	53.2	94.8	49.4	88.0
Southern Sotho	55.9	50.5	90.3	42.4	75.9	20.0	35.9	55.9	54.1	96.9	50.9	91.2	44.0	78.7	55.9	53.9	96.5	52.5	94.0	47.3	84.6
Assamese	55.4	52.3	94.4	43.1	77.8	21.2	38.2	55.4	53.3	96.2	52.9	95.4	47.3	85.4	55.4	53.9	97.2	52.7	95.1	47.7	86.1
Mongolian	55.1	48.0	87.1	31.0	56.3	14.0	25.4	55.1	55.2	100.1	52.6	95.3	49.5	89.8	55.1	55.2	100.2	52.6	95.5	50.1	90.9
Azerbaijani	54.8	52.8	96.3	46.0	83.8	22.8	41.6	54.8	54.5	99.4	52.9	96.4	50.5	92.1	54.8	54.5	99.3	53.5	97.6	51.2	93.4
Kyrgyz	53.0	47.8	90.1	42.0	79.2	18.2	34.3	53.0	51.2	96.6	50.8	95.9	48.8	92.2	53.0	50.6	95.5	51.1	96.5	47.8	90.2
Kinyarwanda	52.4	49.6	94.7	39.7	75.7	18.7	35.7	52.4	52.1	99.3	50.9	97.1	46.2	88.2	52.4	52.0	99.3	50.8	96.9	47.4	90.3
Burmese	52.2	47.1	90.3	37.3	71.4	15.8	30.3	52.2	52.3	100.2	50.3	96.5	46.2	88.5	52.2	52.3	100.3	50.6	97.0	48.5	92.9
Malagasy	51.3	46.3	90.3	40.6	79.1	18.5	36.1	51.3	50.4	98.3	48.8	95.3	43.4	84.6	51.3	51.0	99.5	49.4	96.4	46.8	91.3
Somali	50.3	45.0	89.5	37.1	73.9	18.1	35.9	50.3	50.9	101.3	49.1	97.6	45.3	90.1	50.3	50.4	100.3	49.8	99.1	45.5	90.5
Maori	50.2	40.1	79.9	35.1	69.8	17.7	35.2	50.2	49.3	98.1	47.6	94.7	42.9	85.4	50.2	49.3	98.2	48.1	95.8	44.2	88.1
Igbo	49.2	43.2	87.8	37.0	75.1	17.2	35.0	49.2	48.0	97.5	46.2	93.9	40.8	82.9	49.2	47.9	97.3	46.5	94.5	41.0	83.2
Shona	47.6	43.1	90.5	34.0	71.5	17.1	35.9	47.6	46.6	98.0	45.6	95.8	39.6	83.2	47.6	45.4	95.3	44.8	94.1	41.6	87.3
Chichewa	46.7	43.4	92.8	32.6	69.8	17.6	37.6	46.7	46.2	99.0	45.0	96.3	40.2	86.1	46.7	47.3	101.3	44.6	95.5	42.8	91.7
Swati	45.2	38.5	85.1	33.3	73.6	15.1	33.4	45.2	43.2	95.6	41.4	91.6	34.4	76.1	45.2	43.6	96.3	42.0	92.8	37.0	81.9
Luganda	40.3	36.1	89.7	31.3	77.7	15.1	37.4	40.3	37.3	92.7	35.4	87.8	31.4	77.9	40.3	38.7	96.0	36.0	89.3	30.6	76.0
Lingala	39.8	36.0	90.3	30.7	77.0	15.4	38.8	39.8	37.8	95.0	35.9	90.2	31.4	79.0	39.8	37.6	94.5	36.6	92.0	32.9	82.5
Guarani	39.6	31.3	79.1	24.1	61.0	13.5	34.0	39.6	38.6	97.4	35.9	90.6	31.3	79.1	39.6	37.4	94.5	36.4	91.8	29.3	74.0
Oromo	38.7	33.3	86.2	25.2	65.3	12.2	31.5	38.7	35.0	90.6	31.9	82.5	25.7	66.5	38.7	35.5	91.7	33.5	86.7	26.3	68.1
Tibetan	36.1	29.6	81.9	21.7	60.2	12.3	34.2	36.1	33.7	93.3	31.2	86.5	26.1	72.2	36.1	34.2	94.8	31.7	87.9	24.5	67.9
Bambara	24.2	20.6	85.1	17.8	73.4	11.1	46.0	24.2	20.9	86.1	19.9	82.1	16.9	69.7	24.2	20.1	82.9	19.1	78.8	16.4	67.9
Avg	59.3	54.8	92.0	46.5	77.8	23.9	39.6	59.3	58.6	98.6	57.6	96.7	54.4	90.9	59.3	58.6	98.6	57.7	96.9	55.1	92.1

Table 42: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Qwen3.5 on Translation under top-k reduction, global α, and token α. Per-language mean within tier. Part 2: tiers T2–T1 and overall mean.

	Gemma4 – Translation													
	Standard							Global α						
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T5	64.7	64.3	99.3	63.5	98.0	57.9	89.4	64.7	64.4	99.4	63.5	98.1	59.3	91.7
French	69.8	69.4	99.5	69.4	99.4	65.1	93.2	69.8	69.5	99.5	68.5	98.0	66.4	95.1
German	69.4	69.0	99.5	69.0	99.5	64.1	92.4	69.4	69.1	99.6	68.8	99.2	65.0	93.6
Arabic	68.5	68.2	99.4	66.7	97.2	59.2	86.3	68.5	68.5	100.0	67.5	98.4	58.9	86.0
Spanish	60.8	60.9	100.0	59.2	97.4	55.1	90.6	60.8	60.6	99.6	60.1	98.8	57.3	94.2
Chinese (S.)	60.5	59.4	98.3	58.1	96.1	51.4	85.1	60.5	59.7	98.7	58.3	96.4	54.9	90.8
Japanese	59.3	58.7	99.0	58.3	98.3	52.5	88.5	59.3	58.7	99.1	57.9	97.6	53.5	90.2
T4	65.7	65.0	98.9	63.8	97.1	57.6	87.6	65.7	65.0	99.0	64.1	97.5	59.1	89.8
Catalan	73.0	72.8	99.8	72.0	98.6	66.4	91.0	73.0	72.7	99.7	72.0	98.7	67.7	92.8
Portuguese	72.8	72.5	99.6	71.7	98.5	66.3	91.1	72.8	72.5	99.7	71.8	98.6	69.1	95.0
Swedish	72.0	71.0	98.6	70.8	98.3	67.3	93.4	72.0	71.3	99.0	70.5	97.8	67.7	94.0
Serbian	70.7	69.6	98.4	68.8	97.3	59.7	84.4	70.7	69.7	98.5	68.7	97.1	63.3	89.5
Czech	68.0	67.5	99.4	65.8	96.8	62.9	92.5	68.0	67.2	98.9	66.6	97.9	62.8	92.4
Turkish	67.1	65.8	98.2	64.5	96.1	53.9	80.4	67.1	65.3	97.3	64.4	96.0	57.1	85.1
Hindi	66.6	65.9	98.9	64.0	96.0	56.6	85.0	66.6	66.2	99.4	64.4	96.7	56.6	84.9
Croatian	66.2	65.0	98.2	64.5	97.5	57.0	86.1	66.2	65.5	98.9	65.0	98.2	60.5	91.4
Hungarian	65.2	64.6	99.0	63.0	96.6	57.9	88.8	65.2	65.2	99.9	64.2	98.4	59.6	91.3
Russian	64.7	64.0	99.0	62.9	97.3	58.5	90.5	64.7	63.9	98.9	62.4	96.5	58.8	91.0
Vietnamese	64.4	64.1	99.5	62.5	97.0	52.7	81.8	64.4	64.0	99.4	62.9	97.6	56.1	87.1
Persian	64.3	62.1	96.6	61.4	95.5	53.0	82.5	64.3	62.8	97.7	61.9	96.2	54.6	85.0
Finnish	63.7	63.0	98.9	61.1	96.0	55.7	87.5	63.7	63.3	99.4	62.5	98.1	56.6	88.9
Italian	63.2	62.4	98.7	61.7	97.6	60.5	95.8	63.2	62.6	99.0	62.1	98.3	60.7	96.1
Korean	61.0	60.2	98.6	58.9	96.6	50.7	83.0	61.0	60.1	98.5	58.6	96.0	50.8	83.3
Dutch	60.2	59.9	99.5	59.2	98.3	56.4	93.6	60.2	59.8	99.3	58.9	97.8	57.3	95.2
Basque	60.2	59.5	98.9	57.6	95.6	46.0	76.5	60.2	59.3	98.6	58.6	97.3	49.0	81.4
Polish	59.2	59.2	99.9	57.9	97.7	55.0	92.8	59.2	59.0	99.5	58.4	98.5	54.9	92.6
T3	66.6	65.8	98.7	64.4	96.5	55.4	82.8	66.6	66.0	99.0	64.5	96.7	57.8	86.5
Afrikaans	78.9	78.0	98.9	75.7	96.0	69.5	88.2	78.9	77.8	98.7	76.7	97.3	71.2	90.2
Danish	72.5	72.4	99.8	72.0	99.3	68.3	94.2	72.5	72.6	100.2	71.4	98.5	69.3	95.6
Filipino	71.6	70.3	98.3	69.7	97.4	61.4	85.7	71.6	70.5	98.5	69.7	97.4	64.2	89.6
Malay	71.4	71.0	99.4	69.5	97.4	63.1	88.4	71.4	70.8	99.1	70.0	98.0	65.0	91.0
Hebrew	71.1	70.5	99.1	69.8	98.3	56.0	78.8	71.1	70.6	99.3	69.4	97.7	60.5	85.2
Romanian	70.3	70.0	99.6	69.6	99.1	63.8	90.8	70.3	69.8	99.3	68.9	98.0	65.2	92.7
Bulgarian	69.9	69.1	98.9	68.9	98.6	63.0	90.2	69.9	69.4	99.4	69.0	98.7	65.5	93.8
Cebuano	69.5	67.9	97.8	66.8	96.1	56.4	81.3	69.5	68.4	98.5	67.3	96.9	60.4	87.0
Indonesian	69.1	68.2	98.6	66.9	96.7	59.8	86.5	69.1	69.0	99.8	67.0	96.8	62.0	89.6
Slovak	67.3	66.7	99.2	65.9	98.0	61.3	91.1	67.3	67.0	99.6	65.3	97.1	62.5	93.0
Ukrainian	66.3	65.4	98.6	64.7	97.5	59.2	89.3	66.3	66.1	99.6	64.3	97.0	60.9	91.8
Latvian	66.2	66.0	99.7	65.2	98.5	56.5	85.3	66.2	66.0	99.7	65.0	98.3	59.2	89.5
Estonian	65.8	64.7	98.3	63.8	96.9	56.1	85.2	65.8	65.2	99.0	63.9	97.1	59.2	89.9
Slovenian	64.5	63.4	98.2	62.6	97.1	55.3	85.7	64.5	63.5	98.4	62.1	96.2	57.5	89.2
Greek	64.3	64.2	99.9	62.4	97.1	55.5	86.3	64.3	63.7	99.2	62.2	96.7	59.2	92.1
Kazakh	63.1	61.2	97.1	57.9	91.9	41.4	65.7	63.1	61.6	97.7	59.5	94.3	48.1	76.3
Lithuanian	62.8	62.9	100.2	61.0	97.1	54.4	86.6	62.8	62.5	99.6	61.1	97.4	54.9	87.6
Bengali	62.4	61.2	98.0	59.4	95.2	49.2	78.8	62.4	61.6	98.7	59.8	95.9	48.9	78.4
Urdu	62.1	61.0	98.1	58.7	94.5	44.9	72.2	62.1	61.1	98.5	59.2	95.4	41.6	66.9
Thai	62.0	60.5	97.6	59.4	95.8	50.3	81.2	62.0	61.2	98.7	60.0	96.8	53.2	85.9
Uzbek	62.0	59.9	96.7	56.1	90.4	40.7	65.6	62.0	60.7	98.0	58.2	93.9	46.4	74.8
Georgian	60.0	59.6	99.3	57.7	96.1	46.4	77.4	60.0	59.5	99.1	56.7	94.5	48.8	81.3
Tamil	59.8	59.5	99.5	56.6	94.6	42.1	70.4	59.8	59.2	99.0	56.0	93.6	46.2	77.2

Table 43: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Gemma4 on Translation under top-k reduction and global α. Per-language mean within tier. Part 1: tiers T5–T3.

Gemma4 – Translation														
	Standard							Global α						
Tier/Language	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2	$r{=}8$	$r{=}4$	Ret_4	$r{=}3$	Ret_3	$r{=}2$	Ret_2
T2	58.4	56.5	96.3	53.7	91.1	40.7	68.3	58.4	56.8	97.0	53.9	91.5	44.1	74.2
Maltese	78.0	77.2	98.9	76.2	97.7	61.6	79.0	78.0	77.6	99.4	76.5	98.1	63.5	81.3
Swahili	68.3	66.5	97.3	64.6	94.6	56.1	82.2	68.3	66.7	97.6	63.3	92.7	58.0	85.0
Punjabi	65.1	63.1	97.0	60.4	92.8	44.5	68.5	65.1	63.3	97.3	61.8	94.9	48.2	74.1
Haitian Creole	63.6	62.0	97.5	60.0	94.3	48.3	76.0	63.6	62.1	97.6	60.0	94.3	52.1	81.9
Marathi	63.2	62.7	99.1	59.5	94.1	40.9	64.7	63.2	62.6	99.0	60.3	95.4	49.1	77.7
Lao	62.8	61.5	97.8	59.6	94.8	47.5	75.7	62.8	61.8	98.4	59.9	95.4	48.9	77.9
Icelandic	62.7	62.1	99.1	60.4	96.5	54.6	87.1	62.7	62.2	99.2	60.0	95.8	55.4	88.3
Zulu	61.7	60.6	98.2	57.9	93.8	44.8	72.5	61.7	60.0	97.2	57.0	92.3	46.1	74.8
Amharic	59.8	56.9	95.2	54.7	91.6	38.3	64.1	59.8	58.3	97.6	54.0	90.3	44.0	73.7
Xhosa	58.8	56.4	95.9	52.3	89.0	38.0	64.6	58.8	56.5	96.0	52.8	89.8	42.4	72.1
Hausa	57.7	55.8	96.7	51.8	89.7	38.7	67.1	57.7	56.2	97.3	52.8	91.4	44.5	77.1
Tswana	49.9	47.7	95.6	44.2	88.5	32.5	65.1	49.9	47.8	95.8	43.8	87.6	34.4	69.0
Tigrinya	46.2	42.0	91.0	36.2	78.4	22.1	47.7	46.2	43.5	94.3	38.8	83.9	24.0	51.9
Yoruba	44.1	41.8	94.7	38.1	86.3	25.1	56.9	44.1	41.9	94.9	39.0	88.4	27.1	61.5
Wolof	34.2	31.1	91.1	28.8	84.3	18.1	53.1	34.2	31.7	92.8	28.0	81.9	22.9	67.1
T1	56.6	54.7	96.2	51.8	90.8	37.7	65.7	56.6	55.0	96.9	52.4	91.9	41.4	72.1
Norwegian	76.8	75.4	98.1	76.5	99.5	69.4	90.3	76.8	75.8	98.7	75.5	98.3	73.0	95.0
Macedonian	68.8	68.7	99.9	67.2	97.7	61.0	88.7	68.8	68.2	99.2	67.9	98.7	63.5	92.3
Albanian	68.1	67.3	98.9	65.8	96.6	58.3	85.6	68.1	67.2	98.6	66.3	97.4	60.6	89.0
Nepali	66.9	65.6	98.0	61.9	92.6	49.7	74.3	66.9	65.5	97.9	62.8	93.8	52.1	77.8
Armenian	66.6	66.2	99.4	64.5	96.8	52.3	78.5	66.6	65.8	98.8	64.5	96.8	54.1	81.2
Gujarati	64.9	63.7	98.1	60.6	93.3	40.0	61.6	64.9	63.9	98.4	62.1	95.6	48.1	74.0
Javanese	64.3	62.9	97.9	59.8	93.0	48.2	74.9	64.3	62.9	97.9	59.7	92.8	51.9	80.7
Telugu	64.2	62.7	97.7	61.3	95.4	39.2	61.0	64.2	62.5	97.3	61.9	96.3	43.7	68.1
Malayalam	63.4	61.5	97.0	59.7	94.2	42.4	66.9	63.4	61.8	97.6	60.1	94.9	48.5	76.6
Tajik	61.9	59.1	95.6	55.1	89.1	38.5	62.3	61.9	59.5	96.1	56.7	91.7	45.7	73.9
Sundanese	61.8	60.1	97.2	57.5	93.0	39.5	63.9	61.8	60.3	97.6	57.8	93.4	45.6	73.7
Oriya	61.5	59.9	97.3	56.5	91.7	33.7	54.7	61.5	60.9	98.9	56.9	92.5	38.8	63.0
Sinhala	61.4	60.3	98.2	56.6	92.2	41.6	67.7	61.4	60.1	97.9	56.7	92.4	41.7	68.0
Kannada	60.6	59.1	97.4	57.7	95.2	39.6	65.3	60.6	59.5	98.1	57.2	94.4	45.6	75.2
Khmer	60.3	59.7	98.9	57.3	94.9	46.5	77.0	60.3	60.4	100.1	56.9	94.2	48.7	80.7
Kurdish	59.7	58.4	97.8	53.7	90.0	36.4	61.0	59.7	58.4	97.8	55.1	92.3	41.3	69.2
Mongolian	59.0	57.3	97.1	53.7	91.0	41.9	70.9	59.0	57.9	98.2	56.3	95.4	42.5	72.0
Southern Sotho	58.8	56.2	95.6	53.5	91.0	35.4	60.2	58.8	56.0	95.3	53.8	91.4	38.6	65.6
Pashto	57.5	55.5	96.6	53.4	92.9	35.5	61.8	57.5	56.2	97.9	53.2	92.6	37.0	64.3
Assamese	56.9	54.6	96.0	51.8	91.0	32.4	56.9	56.9	55.2	97.0	52.4	92.1	39.2	69.0
Malagasy	56.5	54.1	95.8	52.4	92.8	37.1	65.7	56.5	54.5	96.6	53.3	94.3	42.1	74.6
Burmese	56.4	53.8	95.4	49.5	87.7	35.9	63.7	56.4	54.3	96.3	51.3	90.9	38.6	68.4
Kinyarwanda	55.7	53.6	96.3	50.0	89.7	29.6	53.1	55.7	54.1	97.2	51.4	92.2	37.4	67.1
Azerbaijani	55.6	54.5	98.0	50.5	90.8	39.1	70.4	55.6	54.2	97.6	51.3	92.3	39.5	71.0
Somali	53.9	52.3	97.1	49.9	92.6	35.8	66.3	53.9	52.2	96.9	51.5	95.5	40.9	75.9
Kyrgyz	53.4	52.3	97.9	49.5	92.7	37.1	69.3	53.4	53.2	99.6	50.4	94.3	39.6	74.0
Igbo	53.0	49.5	93.4	45.2	85.3	34.0	64.2	53.0	50.5	95.2	46.2	87.2	38.3	72.2
Chichewa	51.6	50.0	96.7	48.6	94.2	36.8	71.2	51.6	50.5	97.8	48.4	93.8	39.2	75.9
Shona	51.5	49.8	96.7	47.1	91.5	34.9	67.9	51.5	50.0	97.3	47.7	92.8	37.7	73.3
Maori	50.0	48.8	97.5	45.8	91.5	32.8	65.6	50.0	49.3	98.5	45.2	90.4	36.2	72.4
Swati	47.5	43.3	91.3	37.4	78.8	22.2	46.8	47.5	44.3	93.3	40.1	84.4	26.8	56.4
Luganda	44.8	42.0	93.6	37.5	83.7	26.1	58.3	44.8	42.3	94.4	38.3	85.3	29.1	64.9
Lingala	44.5	42.3	95.1	39.0	87.7	28.3	63.5	44.5	42.3	95.0	39.5	88.7	33.2	74.5
Guarani	44.0	41.8	94.9	38.5	87.5	27.0	61.3	44.0	41.7	94.8	38.6	87.7	26.7	60.8
Oromo	43.8	39.4	89.9	36.8	84.0	22.2	50.8	43.8	40.4	92.3	36.6	83.7	27.0	61.7
Tibetan	40.9	35.8	87.6	31.1	76.1	17.9	43.9	40.9	37.0	90.7	32.4	79.2	20.5	50.1
Bambara	28.7	25.3	88.0	23.6	82.0	18.2	63.5	28.7	26.4	91.9	23.4	81.4	18.4	64.2
Avg	61.4	60.0	97.5	57.9	93.7	47.1	75.5	61.4	60.2	97.9	58.2	94.4	49.9	80.2

Table 44: Performance (%) and retention (Ret_r) at $r{=}k{=}8$ and $r{\in}\{4, 3, 2\}$ for Gemma4 on Translation under top-k reduction and global α. Per-language mean within tier. Part 2: tiers T2–T1 and overall mean.