# Rethinking the Role of Temperature in Large Language Model Distillation

**Hoang-Chau Luong** **Lingwei Chen**
Golisano College of Computing and Information Sciences
Rochester Institute of Technology
Rochester, NY, United States
cl6300@rit.edu, lwcics@rit.edu

## Abstract

Reverse Kullback–Leibler (RKL) divergence is widely favored over forward KL (FKL) in large language models (LLM) distillation, yet this preference is largely based on comparisons that omit the temperature $\tau$, overlooking its central role in softening teacher distributions and improving knowledge transfer. In this work, we revisit temperature in LLM distillation and show that it fundamentally changes the comparison between FKL and RKL. Our analysis reveals an asymmetric effect: temperature substantially enriches FKL with non-dominant token signals, whereas it mainly rescales RKL gradients, causing FKL to benefit much more from $\tau$ scaling than RKL. This asymmetry overturns the standard empirical conclusion: although RKL outperforms FKL at $\tau = 1$, FKL consistently surpasses RKL at higher temperatures across instruction-following benchmarks. Moreover, the impact of temperature is not limited to FKL; it improves a broader family of distillation objectives, enabling simple KL-based methods to achieve competitive performance against recent state-of-the-art LLM distillation approaches.

## 1 Introduction

Knowledge distillation (KD) (Hinton et al., 2015) is a widely used paradigm for model compression that transfers knowledge from a large teacher model to a smaller student model (Romero et al., 2015; Cho and Hariharan, 2019; Gou et al., 2021). A central component of classical KD is the distillation temperature $\tau$, which softens teacher distribution and reveals informative "dark knowledge" beyond the top prediction (Hinton et al., 2015; Tang et al., 2020; Zhao et al., 2022). By redistributing probability mass from dominant classes to lower-probability alternatives, KD temperature prevents the distillation signal from collapsing onto the teacher's top prediction, and provides the student with richer inter-class relational information.

Although temperature has been extensively studied in vision-based KD (Beyer et al., 2022; Li et al., 2023; Sun et al., 2024, 2025), it is often omitted in LLM distillation. Early sequence-level KD work for language tasks (Kim and Rush, 2016) reported that $\tau = 1$ achieved the best empirical performance and therefore adopted the default softmax distribution without further analyzing the role of temperature. Subsequent LLM distillation methods similarly formulate and evaluate their distillation frameworks under the default $\tau = 1$ setting (Gu et al., 2024; Agarwal et al., 2024; Ko et al., 2024; Wang et al., 2025; Song and Zheng, 2026), implicitly removing temperature from the design space of KL-based distillation objectives.

This convention directly affects how recent work compares forward KL (FKL) and reverse KL (RKL) objectives, where the commonly reported superiority of RKL is largely established through evaluations conducted at $\tau = 1$ (Gu et al., 2024; Ko et al., 2024; Wu et al., 2025). Unlike RKL, whose optimization is weighted by the student distribution, FKL directly weights each sample loss by the teacher probabilities. As a result, low-entropy teacher soft targets suppress the non-target probability mass that FKL relies on, making it appear inferior not due to an inherent limitation of the objective itself, but because it is evaluated under insufficiently softened teacher distributions. This raises an important question: *Is RKL intrinsically superior for LLM distillation or has FKL been systematically underestimated?*

### 1.1 Related Work

**Temperature in KD**. Given teacher logits $z^t$ and student logits $z^s$, temperature-scaled distributions are defined as

$$p^\tau = \mathrm{softmax}(z^t/\tau),\ q^\tau = \mathrm{softmax}(z^s/\tau), \quad (1)$$

where $\tau \geq 1$. Increasing $\tau$ smooths the distribution by reducing probability mass on dominant tokens



*Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing*, pages 27292–27303
October 24-29, 2026 ©2026 Association for Computational Linguistics

and amplifying lower-probability alternatives. This smoothing has been widely used in vision-based distillation (Hinton et al., 2015; Zhao et al., 2022; Li et al., 2023; Jin et al., 2023; Sun et al., 2024; Cui et al., 2024; Wei et al., 2024), as well as in early NLP distillation methods such as Seq-KD (Kim and Rush, 2016), DistilBERT (Sanh et al., 2019), and TinyBERT (Jiao et al., 2020). In LLM distillation, however, most methods omit KD temperature and evaluate objectives at $\tau = 1$ (Gu et al., 2024; Ko et al., 2024; Wu et al., 2025; Song and Zheng, 2026). Although recent work has revisited temperature for FKL (Song and Zheng, 2026) or proposed adaptive temperature strategies (Xie et al., 2026; Luong et al., 2026b), its objective-dependent effect on FKL and RKL remains poorly understood.

**KL objectives for LLM distillation**. FKL and RKL are two standard distribution-matching objectives. FKL minimizes $D_{\mathrm{KL}}(p^\tau \| q^\tau)$, encouraging the student to cover the teacher distribution, while RKL minimizes $D_{\mathrm{KL}}(q^\tau \| p^\tau)$ and is associated with mode-seeking behavior. Recent LLM distillation methods favor RKL due to its strong empirical performance (Gu et al., 2024; Ko et al., 2024), leading to variants such as AKL (Wu et al., 2025), AB-KD (Wang et al., 2025), SFKL/SRKL (Ko et al., 2024), and DRKL (Luong et al., 2026a). However, these comparisons are typically conducted under $\tau = 1$ setting, leaving the interaction between temperature and KL objective design underexplored.

### 1.2 Contributions

We revisit the role of temperature in LLM distillation and address three key questions:

**RQ1: How does temperature distinguish FKL from RKL?** We theoretically show that both FKL and RKL converge to the same logit-matching behavior in the high-temperature regime. At practical temperatures, KD temperature reshapes the teacher soft targets and enriches the non-target knowledge transfer for FKL, while primarily rescaling RKL gradients. Thus, KD temperature changes *what* FKL learns, but mainly changes *how* RKL learns.

**RQ2: When does FKL outperform RKL?** We demonstrate that the FKL–RKL comparison is highly temperature-dependent. While RKL often performs better at $\tau = 1$, FKL benefits substantially more from softened teacher targets and consistently surpasses RKL at higher KD temperatures. This finding challenges the common view that RKL is intrinsically superior for LLM distillation.

**RQ3: Does temperature broadly improve KL-based distillation objectives?** We further show that KD temperature improves not only FKL, but also a broader family of KL-based objectives, including Sym-KL, JS, SFKL, and AKL. With appropriate KD temperature, these simple objectives achieve performance competitive with existing LLM distillation methods. This suggests that KD temperature is not merely a correction for FKL, but a general objective-dependent factor for designing and fairly comparing LLM distillation methods.

## 2 How Does Temperature Distinguish FKL from RKL?

This section theoretically analyzes the behavior of FKL and RKL under KD temperature, which shows that the two objectives become equivalent in the high-temperature regime, but diverge substantially at low temperatures, explaining why FKL can outperform RKL in LLM distillation.

Let $z^t, z^s \in \mathbb{R}^V$ denote the teacher and student logits over a vocabulary of size $V$. The temperature-scaled teacher and student distributions, denoted by $p^\tau$ and $q^\tau$, are defined in Eq. (1). We study $\mathcal{L}_{\mathrm{FKL}} = D_{\mathrm{KL}}(p^\tau \| q^\tau) = \sum_{i=1}^{V} p_i^\tau \log(p_i^\tau / q_i^\tau)$ and $\mathcal{L}_{\mathrm{RKL}} = D_{\mathrm{KL}}(q^\tau \| p^\tau)$. Their gradients with respect to student logit $z_i^s$ are

$$\nabla_{z_i^s} \mathcal{L}_{\mathrm{FKL}} = \frac{1}{\tau}(q_i^\tau - p_i^\tau), \tag{2}$$

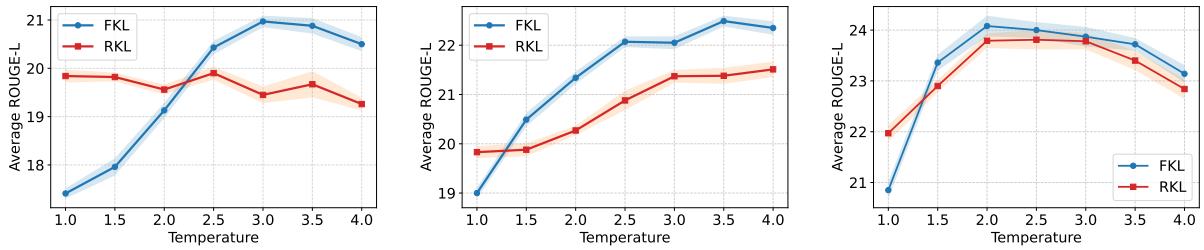
$$\nabla_{z_i^s} \mathcal{L}_{\mathrm{RKL}} = \frac{1}{\tau} q_i^\tau \left( \log \frac{q_i^\tau}{p_i^\tau} - \mathcal{L}_{\mathrm{RKL}} \right). \tag{3}$$

**Proposition 2.1.** *Assume the teacher and student logits are centered, i.e., $\sum_i z_i^t = \sum_i z_i^s = 0$. As $\tau \to \infty$, FKL and RKL gradients satisfy*

$$\nabla_{z_i^s} \mathcal{L}_{\mathrm{FKL}} = \nabla_{z_i^s} \mathcal{L}_{\mathrm{RKL}} = \frac{1}{V\tau^2}(z_i^s - z_i^t) + O(\tau^{-3}).$$

**High-temperature regime: FKL and RKL become logit matching**. Recent studies (Wu et al., 2025; Luong et al., 2026a) suggest that FKL and RKL reach the same solution under idealized assumptions, such as sufficient student capacity and exact global optimization. Proposition 2.1 (proof is provided in Appendix B) shows that temperature alone is sufficient to recover this equivalence: as $\tau \to \infty$, both objectives share the same gradient and reduce to teacher–student logit matching. Thus, the practical distinction between FKL and RKL must come from their low-temperature regime.

**Low-temperature regime: temperature separates FKL from RKL**. At practical temperatures,




(a) GPT-2 XL → GPT-2 Base (b) GPT-2 XL → GPT-2 Medium (c) OPT 6.7B → OPT 1.3B

Figure 1: Effect of temperature on FKL and RKL across model scales. Temperature consistently improves FKL while providing limited benefit for RKL, leading to a reversal in their performance at higher temperatures.

FKL directly matches the student distribution to the softened teacher distribution, whereas RKL still operates through student-weighted teacher–student logit gaps. For FKL, $p^\tau$ serves as the probability target, and changing $\tau$ changes the supervision itself: $\frac{\partial p_i^\tau}{\partial \tau} = \frac{p_i^\tau}{\tau^2}(\mathbb{E}_{j\sim p^\tau}[z_j^t] - z_i^t)$. Increasing temperature reduces the dominance of high-logit tokens and redistributes probability mass to lower-logit tokens. Together with Eq. (2), this shows temperature changes the strength of the FKL update, and teacher distribution the student is trained to match.

**Proposition 2.2.** *Let $\Delta_i = z_i^s - z_i^t$ denote the student–teacher logit gap at index $i$, the RKL gradient can be rewritten as*

$$\nabla_{z_i^s}\mathcal{L}_{\mathrm{RKL}} = \frac{q_i^\tau}{\tau^2}\left(\Delta_i - \mathbb{E}_{j\sim q^\tau}[\Delta_j]\right), \qquad (4)$$

For RKL, temperature plays a different role. Proposition 2.2 (see Appendix C for the detailed proof) shows that the RKL gradient is driven by teacher–student logit gaps rather than direct probability matching between student and softened teacher distributions as in FKL. Thus, temperature mainly conditions RKL optimization by rescaling the gradient by $1/\tau^2$ and smoothing the student-side weights $q_i^\tau$. Hinton et al. (2015) counteract this rescaling by multiplying the gradient by $\tau^2$, while the smoothing effect of $q_i^\tau$ still remains.

**Insight.** Temperature changes *what* FKL learns, but mainly *how* RKL learns. At low-regime temperatures, FKL receives richer non-target supervision from softened teacher distribution, while RKL gets gradient rescaling.

## 3 When Does FKL Outperform RKL?

Our analysis leads to a natural question: *Can temperature change which KL objective is better?* We answer this question empirically on five instruction-following benchmarks including Dolly Eval, Self-Instruct (Wang et al., 2023), Vicuna Eval (Chiang et al., 2023), Super-Natural Instructions (Super-NI) (Wang et al., 2022), and Unnatural Instructions (UnNI) (Honovich et al., 2023) using GPT-2 (Radford et al., 2019) and OPT models (Zhang et al., 2022), with implementation details provided in Appendix A. Figure 1 presents the comparative average ROUGE-L across five datasets, and reveals a clear reversal. Under $\tau = 1$, RKL outperforms FKL, consistent with prior LLM distillation results. However, once temperature is introduced, FKL consistently surpasses RKL, showing that its apparent weakness at $\tau = 1$ stems from insufficient access to softened non-target teacher signals. In particular, we draw the following insights.

Table 1: Distillation under on-policy setting for GPT-2 XL → GPT-2 Base. We report ROUGE-L mean and standard deviation over five random seeds. Bold: Avg. Rouge-L exceeds the default ($\tau = 1.0$).

| Loss | $\tau$ | Dolly | SelfInst | Super-NI | UnNI | Vicuna | Avg. |
|---|---|---|---|---|---|---|---|
| FKL | 1.0 | 22.03± 0.86 | 10.29± 0.40 | 17.42± 0.18 | 20.11± 0.10 | 14.89± 0.43 | 16.95 |
| | 1.5 | 23.62± 0.49 | 10.57± 0.55 | 19.01± 0.08 | 21.40± 0.09 | 14.27± 0.40 | **17.77** |
| | 2.0 | 24.16± 0.11 | 11.81± 0.62 | 20.96± 0.05 | 23.64± 0.14 | 15.20± 0.43 | **19.15** |
| | 2.5 | 26.00± 0.54 | 12.62± 0.54 | 22.35± 0.34 | 25.83± 0.10 | 16.06± 0.56 | **20.57** |
| | 3.0 | 25.30± 0.30 | 12.11± 0.25 | 23.70± 0.20 | 26.79± 0.26 | 16.62± 0.58 | **20.90** |
| | 3.5 | 25.32± 0.07 | 12.40± 0.22 | 24.92± 0.44 | 27.20± 0.02 | 16.35± 0.29 | **21.24** |
| | 4.0 | 25.60± 0.48 | 12.36± 0.44 | 24.28± 0.26 | 26.63± 0.09 | 16.42± 0.50 | **21.06** |
| RKL | 1.0 | 24.12± 0.25 | 12.20± 0.38 | 22.43± 0.09 | 24.93± 0.06 | 16.14± 0.37 | 19.97 |
| | 1.5 | 23.96± 0.13 | 12.00± 0.62 | 23.16± 0.26 | 26.78± 0.16 | 16.14± 0.37 | **20.41** |
| | 2.0 | 24.56± 0.45 | 12.33± 0.09 | 22.76± 0.18 | 25.86± 0.15 | 15.64± 0.72 | **20.23** |
| | 2.5 | 24.48± 0.13 | 10.96± 0.46 | 21.88± 0.22 | 24.94± 0.01 | 15.90± 0.44 | 19.63 |
| | 3.0 | 24.36± 0.37 | 11.06± 0.13 | 19.92± 0.48 | 23.63± 0.13 | 16.01± 0.12 | 19.00 |
| | 3.5 | 24.00± 0.18 | 10.66± 0.14 | 20.41± 0.38 | 23.48± 0.22 | 15.78± 0.30 | 18.87 |
| | 4.0 | 23.85± 0.31 | 11.26± 0.03 | 19.88± 0.43 | 23.15± 0.15 | 16.07± 0.39 | 18.84 |

**Temperature turns FKL from weak to outperforming RKL.** Increasing temperature softens the distribution, exposes richer teacher information, and yields large gains. For GPT-2 Base, FKL improves from 19.41 to 20.97 average ROUGE-L at $\tau = 3.0$. Similar gains appear for larger students: GPT-2 Medium rises from around 19 to above 22, and OPT-1.3B from below 21 to above 24, with best results around $\tau = 2.5$–$3.5$. Across model scales, these gains reverse the ranking: FKL shifts from trailing RKL at $\tau = 1$ to outperforming it at proper temperatures. For example, FKL exceeds RKL by 1.5 points at $\tau = 3.0$ on GPT-2 Base, by over 1.0 point at $\tau = 3.5$ on GPT-2 Medium, and by 0.27 points at $\tau = 2.0$ on OPT 1.3B.

**RKL benefits less from temperature.** For GPT-2

Base, RKL remains nearly flat across temperatures and can even degrade at larger $\tau$. For larger students, the gains are modest, about 1 point on GPT-2 Medium and 2 points on OPT-1.3B, compared with consistent gains above 3 points for FKL. This agrees with our theoretical analysis: temperature mainly smooths and reconditions the RKL gradient without giving RKL direct access to the softened teacher distribution as a probability target.

**On-policy distillation**. To examine whether our findings extend beyond the off-policy setting, we run on-policy GPT-2 XL $\rightarrow$ GPT-2 Base distillation with both FKL and RKL (Table 1). The same asymmetric effect of temperature persists: FKL improves from 16.95 at $\tau = 1$ to 21.24 at $\tau = 3.5$ ($+4.29$ average Rouge-L), whereas RKL improves by only $+0.44$ before degrading at higher $\tau$. The relative ranking also reverses: although RKL outperforms FKL at $\tau = 1$, FKL surpasses RKL at an appropriate KD temperature. These results show that our finding is not specific to the off-policy setting, but holds under on-policy distillation.

**Insight.** FKL outperforms RKL when temperature exposes useful non-target teacher information. At $\tau = 1$, FKL underuses the teacher because the target is too sharp. At higher temperature, the softened distribution becomes sufficiently informative for FKL to surpass RKL.

## 4 Does Temperature Broadly Improve KL-based Distillation Objectives?

We now ask whether temperature is broadly beneficial for KL-based distillation objectives. To answer this, we evaluate FKL, RKL, and Sym-KL, together with recent objectives such as JS (Agarwal et al., 2024), SFKL/SRKL (Ko et al., 2024), AKL (Wu et al., 2025), AB (Wang et al., 2025), and DRKL (Luong et al., 2026a). Experiments are conducted on three instruction-following benchmarks including Dolly, Super-NI, and UnNI, with two additional benchmarks reported in Appendix D and implementation detailed in Appendix A.

**Temperature broadly improves KD objectives**. Table 2 compares each objective at default $\tau = 1$ with its temperature-tuned variant. Temperature improves all objectives across GPT-2 and OPT models, although the gains vary substantially across objectives. For GPT-2 Base, FKL improves from 19.70 to 25.53 average ROUGE-L ($+5.83$), while AKL achieves the largest gain ($+6.62$). JS, SFKL, and Sym-KL also improve notably, whereas RKL and SRKL gain only $+0.49$ and $+0.79$, supporting our analysis that temperature is especially beneficial for FKL-objectives that directly exploit softened teacher probabilities. For OPT-1.3B, the gains become broadly substantial, ranging from $+3.46$ for RKL to $+4.88$ for AKL. Figure 2 visualizes ROUGE-L gains across temperatures and objectives. FKL-based objectives (FKL, Sym-KL, JS, and AKL) exhibit consistent gains, while RKL-based objectives (RKL, SRKL, and SFKL) are less stable and can even degrade. Thus, temperature is more robustly beneficial for FKL-based objectives but requires careful tuning for RKL-based ones.

Table 2: Effect of temperature scaling on distillation losses. We report ROUGE-L mean and standard deviation over five random seeds.

| Method | Dolly | Super-NI | UnNI | Avg. | Δ Avg. |
|---|---|---|---|---|---|
| ***GPT-2 XL (1.5B) → GPT-2 Base (0.1B)*** | | | | | |
| GPT-2 XL | 27.00± 0.19 | 26.46± 0.41 | 31.10± 0.06 | 28.19 | – |
| FKL | 23.80± 0.55 | 16.27± 0.24 | 19.03± 0.09 | 19.70 | – |
| FKL + $\tau$ | **25.83**± 0.24 | 23.78± 0.26 | 26.97± 0.12 | 25.53 | ↑5.83 |
| RKL | 24.67± 0.13 | 21.03± 0.21 | 23.94± 0.12 | 23.21 | – |
| RKL + $\tau$ | 25.00± 0.24 | 21.41± 0.34 | 24.68± 0.07 | 23.70 | ↑0.49 |
| Sym-KL | 24.38± 0.14 | 19.92± 0.25 | 22.21± 0.03 | 22.17 | – |
| Sym-KL + $\tau$ | 25.54± 0.14 | 21.69± 0.10 | 25.29± 0.14 | 24.17 | ↑2.00 |
| JS | 24.08± 0.26 | 19.66± 0.23 | 22.17± 0.21 | 21.97 | – |
| JS + $\tau$ | 25.51± 0.40 | 24.34± 0.16 | 27.03± 0.11 | **25.63** | ↑3.66 |
| SFKL | 24.44± 0.42 | 22.01± 0.23 | 23.47± 0.12 | 23.31 | – |
| SFKL + $\tau$ | 25.33± 0.28 | **24.40**± 0.27 | **27.07**± 0.16 | 25.60 | ↑2.29 |
| SRKL | 24.48± 0.37 | 23.25± 0.36 | 24.01± 0.05 | 23.91 | – |
| SRKL + $\tau$ | 25.03± 0.42 | 23.95± 0.14 | 25.13± 0.08 | 24.70 | ↑0.79 |
| AKL | 21.83± 0.19 | 15.40± 0.20 | 18.06± 0.13 | 18.43 | – |
| AKL + $\tau$ | 25.35± 0.27 | 23.63± 0.16 | 26.18± 0.16 | 25.05 | **↑6.62** |
| AB | 24.32± 0.29 | 23.08± 0.19 | 24.32± 0.14 | 23.91 | – |
| DRKL | 25.51± 0.42 | 23.39± 0.11 | 26.57± 0.11 | 25.16 | – |
| ***OPT 6.7B → OPT 1.3B*** | | | | | |
| OPT 6.7B | 27.52± 0.29 | 30.41± 0.46 | 31.39± 0.20 | 29.77 | – |
| FKL | 26.07± 0.65 | 22.11± 0.38 | 26.82± 0.10 | 25.00 | – |
| FKL + $\tau$ | 28.20± 0.48 | 27.62± 0.45 | 31.43± 0.17 | 29.08 | ↑4.08 |
| RKL | 26.58± 0.11 | 22.64± 0.16 | 26.29± 0.10 | 25.17 | – |
| RKL + $\tau$ | 27.64± 0.48 | 26.96± 0.27 | 31.29± 0.13 | 28.63 | ↑3.46 |
| Sym-KL | 25.73± 0.40 | 23.40± 0.22 | 25.44± 0.07 | 24.86 | – |
| Sym-KL + $\tau$ | 27.74± 0.40 | 28.21± 0.16 | 31.73± 0.15 | 29.23 | ↑4.37 |
| JS | 26.39± 0.71 | 24.02± 0.31 | 26.59± 0.11 | 25.67 | – |
| JS + $\tau$ | 27.97± 0.39 | **28.27**± 0.26 | 32.20± 0.09 | 29.48 | ↑3.81 |
| SFKL | 25.98± 0.44 | 23.62± 0.19 | 26.15± 0.11 | 25.25 | – |
| SFKL + $\tau$ | 27.52± 0.64 | 27.39± 0.20 | 32.03± 0.13 | 28.98 | ↑3.73 |
| SRKL | 26.04± 0.36 | 22.96± 0.22 | 24.71± 0.22 | 24.57 | – |
| SRKL + $\tau$ | 27.83± 0.38 | 28.20± 0.37 | 31.68± 0.14 | 29.24 | ↑4.67 |
| AKL | 26.20± 0.27 | 22.12± 0.18 | 25.89± 0.20 | 24.74 | – |
| AKL + $\tau$ | **28.35**± 0.19 | 28.10± 0.23 | **32.40**± 0.15 | **29.62** | **↑4.88** |
| AB | 26.86± 0.17 | 24.39± 0.11 | 26.62± 0.09 | 25.96 | – |
| DRKL | 27.72± 0.36 | 25.56± 0.36 | 29.84± 0.15 | 27.71 | – |

**Temperature changes objective rankings**. Applying KD temperature substantially strengthens standard KL-based objectives and changes their relative ranking against stronger baselines. On GPT-2 Base, JS+$\tau$ achieves the best average score (25.63), while FKL+$\tau$ reaches 25.53, surpassing state-of-the-art AB (23.91) and DRKL (25.16). On OPT-1.3B, FKL+$\tau$ improves to 29.08, outperforming

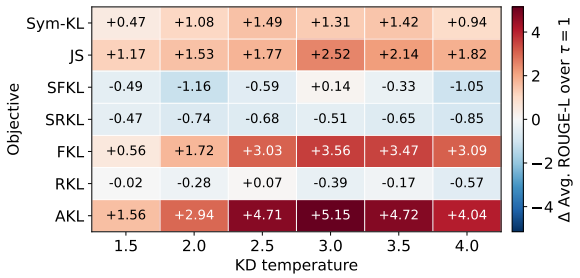



Figure 2: Temperature-sensitivity heatmap on GPT-2 XL → GPT-2 Base.

Table 3: Pointwise LLM-as-a-judge 1-10 score of best-$\tau$ vs. $\tau = 1$ on OPT 6.7B → OPT 1.3B, per method (Claude Haiku 4.5, MT-Bench single-answer rubric, score ×10 shown), across all five instruction-following evaluation sets.

| Method | Dolly | Vicuna | Self-Inst | Super-NI | UnNI | Avg. | ΔAvg. |
|---|---|---|---|---|---|---|---|
| FKL | 40.3 | 23.0 | 22.7 | 16.0 | 26.8 | 25.8 | – |
| FKL + $\tau$ | **44.3** | **29.0** | 27.7 | 20.2 | **29.2** | **30.1** | +4.3 |
| RKL | 36.7 | 27.3 | 25.0 | 19.6 | 17.5 | 25.2 | – |
| RKL + $\tau$ | 37.3 | 27.6 | 26.2 | 20.2 | 19.0 | 26.1 | +0.9 |
| Sym-KL | 38.0 | 20.3 | 26.0 | 16.4 | 17.2 | 23.6 | – |
| Sym-KL + $\tau$ | 38.7 | 27.0 | 26.0 | 16.4 | 18.4 | 25.3 | +1.7 |
| SFKL | 34.0 | 26.3 | 23.7 | 15.6 | 21.6 | 24.2 | – |
| SFKL + $\tau$ | 39.7 | 28.7 | 24.0 | 17.6 | 22.4 | 26.5 | +2.3 |
| SRKL | 37.3 | 23.7 | 26.0 | 16.8 | 18.8 | 24.5 | – |
| SRKL + $\tau$ | 40.3 | 27.7 | 26.3 | 18.4 | 19.2 | 26.4 | +1.9 |
| AKL | 40.7 | 23.7 | 23.7 | 16.4 | 18.0 | 24.5 | – |
| AKL + $\tau$ | 42.0 | 25.0 | **28.0** | **21.2** | 21.6 | 27.6 | +3.1 |

both AB (25.96) and DRKL (27.71), while AKL+$\tau$ achieves the best overall score (29.62). Thus, temperature is not a minor implementation detail: it can change which objective appears strongest and is necessary for fair comparison.

**LLM-as-a-judge**. We additionally conduct an LLM-as-a-judge evaluation on OPT 6.7B → OPT 1.3B across five instruction-following benchmarks, using Claude Haiku 4.5 with the MT-Bench pointwise and pairwise protocols (Zheng et al., 2023). Results are reported in Tables 3 and 4. This evaluation independently corroborates the ROUGE-L findings. Pointwise evaluation shows that FKL+$\tau$ achieves the highest average score (30.1 vs. 25.8 at $\tau = 1$, +4.3), while RKL+$\tau$ improves by only +0.9. Pairwise evaluation reaches a similar conclusion: FKL+$\tau$ wins 57.7% of head-to-head comparisons against its $\tau = 1$ counterpart, whereas RKL+$\tau$ wins only 42.3%, indicating that temperature tuning does not consistently benefit RKL. The same trend extends to other FKL-based objectives (Sym-KL, SFKL, and AKL), all of which are consistently preferred after temperature tuning. These results provide complementary evidence that the asymmetric effect of KD temperature is not specific to ROUGE-L, further strengthening that KD temperature benefits FKL-based objectives substantially more than RKL-based objectives.

Table 4: Pairwise LLM-as-a-judge win rates for best-$\tau$ vs. $\tau = 1$ on OPT 6.7B → OPT 1.3B. Evaluation uses Claude Haiku 4.5, MT-Bench pairwise rubric with position-swap bias control. We report $\tau$-adjusted win rate: win% + 0.5 × tie%, where 50% indicates parity, >50% favors the temperature-tuned variant, and <50% prefers $\tau = 1$.

| Method | Dolly | Vicuna | Self-Inst | Super-NI | UnNI | Avg. |
|---|---|---|---|---|---|---|
| FKL | 60.0 | 60.0 | 61.7 | 53.3 | 53.3 | 57.7 |
| RKL | 45.0 | 43.3 | 43.3 | 33.3 | 46.7 | 42.3 |
| Sym-KL | 51.7 | 66.7 | 53.3 | 53.3 | 53.3 | 55.7 |
| SFKL | 56.7 | 68.3 | 51.7 | 53.3 | 58.3 | 57.7 |
| SRKL | 55.0 | 51.7 | 56.7 | 53.3 | 60.0 | 55.3 |
| AKL | 51.7 | 53.3 | 66.7 | 65.0 | 60.0 | 59.3 |

**Modern model families**. To further strengthen the empirical evidence on modern model families, we additionally evaluated two teacher–student pairs: Llama 3.2-3B (Grattafiori et al., 2024) → Llama 3.2-1B and Qwen3-4B (Yang et al., 2025) → Qwen3-1.7B. The results are reported in Appendix D. These results indicate that the asymmetric effect of temperature is not specific to any particular LLM architecture.

**Case study**. Appendix E provides case studies on the Unnatural Instructions test set, comparing model outputs under different objectives.

**Insight.** Temperature broadly improves KL-based distillation. Its gains are largest for FKL-based objectives, but weaker for RKL-based objectives. Thus, KD temperature should be considered essential for fair comparison, rather than a minor implementation detail.

## 5 Conclusion

In this work, we revisit the role of KD temperature in LLM distillation and show that it can change the relative behavior of FKL and RKL. KD temperature makes non-target teacher probabilities more informative, allowing FKL to transfer richer knowledge and surpass RKL despite underperforming at $\tau = 1$. Beyond that, KD temperature also improves a broad range of KL-based distillation objectives, showing that it is essential for fair comparison.

## Acknowledgments

This work is partially supported by the NSF under grant CNS-2549995. The authors also acknowledge Research Computing at the Rochester Institute of Technology for providing computational resources and support that have contributed to the research results reported in this publication.

## Limitations

Due to computational constraints, our language experiments use teacher models with up to 7B parameters, and evaluating larger teachers remains future work. For the Qwen3 and Llama3.2 teacher–student pairs, we adopt LoRA rather than full fine-tuning. This follows common practice for teacher models at this scale and enables us to evaluate a broader range of model families and sizes within a fixed computational budget. As with most work in this setting, our evaluation centers on English-language instruction following. Extending it to other languages, task types, and training distributions is a natural next step.

## References


Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos Garea, Matthieu Geist, and Olivier Bachem. 2024. On-policy distillation of language models: Learning from self-generated mistakes. In *The twelfth international conference on learning representations*.

Lucas Beyer, Xiaohua Zhai, Amélie Royer, Larisa Markeeva, Rohan Anil, and Alexander Kolesnikov. 2022. Knowledge distillation: A good teacher is patient and consistent. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10925–10934.

Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E. Gonzalez, Ion Stoica, and Eric P. Xing. 2023. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality.

Jang Hyun Cho and Bharath Hariharan. 2019. On the efficacy of knowledge distillation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4794–4802.

Mike Conover, Matt Hayes, Ankit Mathur, Jianwei Xie, Jun Wan, Sam Shah, Ali Ghodsi, Patrick Wendell, Matei Zaharia, and Reynold Xin. 2023. Free dolly: Introducing the world's first truly open instruction-tuned llm.

Jiequan Cui, Zhuotao Tian, Zhisheng Zhong, Xiaojuan Qi, Bei Yu, and Hanwang Zhang. 2024. Decoupled kullback-leibler divergence loss. *Advances in Neural Information Processing Systems*, 37:74461–74486.

Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. 2021. Knowledge distillation: A survey. *International journal of computer vision*, 129(6):1789–1819.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. 2024. Minillm: Knowledge distillation of large language models. In *International Conference on Learning Representations*.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Or Honovich, Thomas Scialom, Omer Levy, and Timo Schick. 2023. Unnatural instructions: Tuning language models with (almost) no human labor. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14409–14428.

Xiaoqi Jiao, Yichun Yin, Lifeng Shang, Xin Jiang, Xiao Chen, Linlin Li, Fang Wang, and Qun Liu. 2020. Tinybert: Distilling bert for natural language understanding. In *Findings of the association for computational linguistics: EMNLP 2020*, pages 4163–4174.

Ying Jin, Jiaqi Wang, and Dahua Lin. 2023. Multi-level logit distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24276–24285.

Yoon Kim and Alexander M Rush. 2016. Sequence-level knowledge distillation. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 1317–1327.

Jongwoo Ko, Sungnyun Kim, Tianyi Chen, and Se-Young Yun. 2024. Distillm: Towards streamlined distillation for large language models. In *Forty-first International Conference on Machine Learning*.

Zheng Li, Xiang Li, Lingfeng Yang, Borui Zhao, Renjie Song, Lei Luo, Jun Li, and Jian Yang. 2023. Curriculum temperature for knowledge distillation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 1504–1512.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Hoang-Chau Luong, Dat Ba Tran, and Lingwei Chen. 2026a. Diversity-aware reverse kullback-leibler divergence for large language model distillation. *arXiv preprint arXiv:2604.00223*.

Hoang-Chau Luong, Nghia Van Vo, Kaiqi Zhao, and Lingwei Chen. 2026b. Consistently informative soft-label temperature for knowledge distillation. *arXiv preprint arXiv:2605.20357*.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, and 1 others. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio. 2015. Fitnets: Hints for thin deep nets. In *In Proceedings of ICLR*.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Mingyang Song and Mao Zheng. 2026. A survey of on-policy distillation for large language models. *arXiv preprint arXiv:2604.00626*.

Shangquan Sun, Wenqi Ren, Jingzhi Li, Rui Wang, and Xiaochun Cao. 2024. Logit standardization in knowledge distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15731–15740.

Wujie Sun, Defang Chen, Siwei Lyu, Genlang Chen, Chun Chen, and Can Wang. 2025. Knowledge distillation with refined logits. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1110–1119.

Jiaxi Tang, Rakesh Shivanna, Zhe Zhao, Dong Lin, Anima Singh, Ed H Chi, and Sagar Jain. 2020. Understanding and improving knowledge distillation. *arXiv preprint arXiv:2002.03532*.

Guanghui Wang, Zhiyong Yang, Zitai Wang, Shi Wang, Qianqian Xu, and Qingming Huang. 2025. ABKD: Pursuing a proper allocation of the probability mass in knowledge distillation via $\alpha$-$\beta$-divergence. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 65167–65212. PMLR.

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi. 2023. Self-instruct: Aligning language models with self-generated instructions. In *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: long papers)*, pages 13484–13508.

Yizhong Wang, Swaroop Mishra, Pegah Alipoormolabashi, Yeganeh Kordi, Amirreza Mirzaei, Atharva Naik, Arjun Ashok, Arut Selvan Dhanasekaran, Anjana Arunkumar, David Stap, and 1 others. 2022. Super-naturalinstructions: Generalization via declarative instructions on 1600+ nlp tasks. In *Proceedings of the 2022 conference on empirical methods in natural language processing*, pages 5085–5109.

Shicai Wei, Chunbo Luo, and Yang Luo. 2024. Scaled decoupled distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15975–15983.

Taiqiang Wu, Chaofan Tao, Jiahao Wang, Runming Yang, Zhe Zhao, and Ngai Wong. 2025. Rethinking kullback-leibler divergence in knowledge distillation for large language models. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 5737–5755.

Xurong Xie, Zhucun Xue, Jiafu Wu, Jian Li, Yabiao Wang, Xiaobin Hu, Yong Liu, and Jiangning Zhang. 2026. Llm-oriented token-adaptive knowledge distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 34070–34078.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, and 41 others. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona Diab, Xian Li, Xi Victoria Lin, and 1 others. 2022. Opt: Open pre-trained transformer language models. *arXiv preprint arXiv:2205.01068*.

Borui Zhao, Quan Cui, Renjie Song, Yiyu Qiu, and Jiajun Liang. 2022. Decoupled knowledge distillation. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 11953–11962.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, and 1 others. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in neural information processing systems*, 36:46595–46623.

This appendix provides additional implementation details and supporting theoretical and empirical materials for the main paper:

- Appendix A presents implementation, training, and evaluation details, including datasets, teacher–student settings, compared objectives, and temperature tuning protocols.
- Appendix B provides proof for Proposition 2.1.
- Appendix C provides proof for Proposition 2.2.
- Appendix D reports additional experimental results on the Self-Instruct and Vicuna benchmarks across all objectives, as well as results on the modern model families Llama3.2 and Qwen3.
- Appendix E presents qualitative case studies illustrating how KD temperature improves instruction following and output quality under different distillation objectives.
- Appendix F describes the limited use of AI assistants for language refinement during manuscript preparation.

## A Implementation Details

The implementation is publicly available at `https://github.com/lhchau/rethinking-kd-temperature-llm`.

**Training resources**. All experiments were conducted on a GPU cluster with four NVIDIA A100 GPUs, each with 40GB memory.

**Experimental setup**. We evaluate LLM distillation in an off-policy instruction-following setting, where student models are trained on fixed teacher-generated responses. Following Gu et al. (2024), we use the instruction–response dataset constructed from `databricks-dolly-15k` (Conover et al., 2023), containing 14k training examples, 500 validation examples, and 500 test examples. We first fine-tune the teacher models on this dataset and then distill their responses into smaller students. Our teacher–student pairs include GPT-2 XL (Radford et al., 2019) (1.5B) $\rightarrow$ GPT-2 Base (120M) and OPT 6.7B (Zhang et al., 2022) $\rightarrow$ OPT 1.3B.

**Training protocol**. To isolate the effect of the distillation objective and KD temperature, all methods are trained under the same protocol. For both GPT-2 and OPT families, we use a batch size of 32, train for 20 epochs, and set the maximum input length to 512 tokens. Following Gu et al. (2024), the learning rate is set to $5 \times 10^{-4}$ for GPT-2 Base and $5 \times 10^{-5}$ for OPT-1.3B. All compared methods use the same training data, teacher checkpoint, student initialization, and hyperparameter budget, ensuring that performance differences are attributable to the distillation objective and temperature design rather than implementation or training discrepancies.

**Evaluation protocol**. We evaluate the distilled students on five instruction-following benchmarks, including Dolly Eval, Self-Instruct (Wang et al., 2023), Vicuna Eval (Chiang et al., 2023), Super-Natural Instructions (Super-NI) (Wang et al., 2022), and Unnatural Instructions (UnNI) (Honovich et al., 2023). We report ROUGE-L (Lin, 2004) averaged over five random seeds $\{10, 20, 30, 40, 50\}$. Model checkpoints are saved after each epoch, and final results are reported using the checkpoint with the best validation ROUGE-L. The decoding temperature is set to 1 during evaluation.

**Compared objectives**. We compare KL-family and recent LLM distillation objectives under the same off-policy setting. The KL-family baselines include FKL, RKL (Gu et al., 2024), symmetric KL (Sym-KL), Jensen–Shannon divergence (JS) (Agarwal et al., 2024), skewed FKL (SFKL) (Ko et al., 2024), skewed RKL (SRKL) (Ko et al., 2024). For Sym-KL, we use $0.5\,\mathrm{FKL} + 0.5\,\mathrm{RKL}$ and for SFKL and SRKL, we set the smoothing parameter to $\alpha = 0.1$. We also compare with stronger recent objectives, including Adaptive KL (AKL) (Wu et al., 2025), $\alpha$-$\beta$ divergence (AB) (Wang et al., 2025) with $(\alpha = 0.2, \beta = 0.7)$, and diversity-aware RKL (DRKL) (Luong et al., 2026a). Since our focus is off-policy distillation from fixed teacher responses, we exclude methods that require on-policy sampling or additional external datasets. For all baselines, we follow the hyperparameter settings from the original papers and use the implementation provided by DistilLM (Ko et al., 2024).

For the temperature-tuned variants reported in Tables 2 and 5, including FKL+$\tau$, RKL+$\tau$, Sym-KL+$\tau$, JS+$\tau$, SFKL+$\tau$, SRKL+$\tau$, and AKL+$\tau$, we apply KD temperature on top of each original distillation objective while keeping its implementation and default hyperparameters unchanged. The only tuned factor is the KD temperature, for which we perform a grid search over $\tau \in \{1.5, 2.0, 2.5, 3.0, 3.5, 4.0\}$ and report the best-performing setting for comparison. Thus, the gains reported in Tables 2 and 5 are attributable solely to KD temperature, not to additional objective-specific hyperparameter tuning. Further tuning of each objective's own hyperparameters could poten-

tially yield additional improvements.

## B Proof for Proposition 2.1

*Proof.* For large $\tau$, we use the first-order expansion $\exp(z_i/\tau) = 1 + z_i/\tau + O(\tau^{-2})$. For the teacher distribution, this gives

$$\sum_j \exp(z_j^t/\tau) = V + \frac{1}{\tau}\sum_j z_j^t + O(\tau^{-2}) = V + O(\tau^{-2}),$$

where the last equality follows from the centered-logit assumption $\sum_j z_j^t = 0$. Thus,

$$p_i^\tau = \frac{1 + z_i^t/\tau + O(\tau^{-2})}{V + O(\tau^{-2})} = \frac{1}{V} + \frac{z_i^t}{V\tau} + O(\tau^{-2}).$$

Applying the same argument to the student logits gives

$$q_i^\tau = \frac{1}{V} + \frac{z_i^s}{V\tau} + O(\tau^{-2}).$$

**For FKL part.** Substituting these into the FKL gradient in Eq. (2) gives

$$\nabla_{z_i^s}\mathcal{L}_{\mathrm{FKL}} = \frac{1}{V\tau^2}(z_i^s - z_i^t) + O(\tau^{-3}).$$

**For RKL part.** Since $q_j^\tau = 1/V + O(\tau^{-1})$ and the centered-logit assumption gives $\sum_j \Delta_j = 0$, we have

$$\begin{aligned}\mathbb{E}_{j\sim q^\tau}[\Delta_j] &= \sum_j q_j^\tau \Delta_j \\ &= \frac{1}{V}\sum_j \Delta_j + \sum_j O(\tau^{-1})\Delta_j \\ &= O(\tau^{-1}).\end{aligned}$$

Using this, Proposition 2.2, and $q_i^\tau = 1/V + O(\tau^{-1})$ as $\tau \to \infty$, we have

$$\nabla_{z_i^s}\mathcal{L}_{\mathrm{RKL}} = \frac{1}{V\tau^2}(z_i^s - z_i^t) + O(\tau^{-3}).$$

This completes the proof. □

## C Proof for Proposition 2.2

*Proof.* The log-ratio can be decomposed as

$$\log\frac{q_i^\tau}{p_i^\tau} = \frac{\Delta_i}{\tau} + C_\tau, \text{where } C_\tau = \log\frac{Z_t^\tau}{Z_s^\tau}, \quad (5)$$

$Z_t^\tau = \sum_j \exp(z_j^t/\tau)$ and $Z_s^\tau = \sum_j \exp(z_j^s/\tau)$. The term $C_\tau$ comes only from the softmax normalization constants and is independent of token index $i$. From the definition of $\mathcal{L}_{\mathrm{RKL}}$, we have

$$\mathcal{L}_{\mathrm{RKL}} = \mathbb{E}_{j\sim q^\tau}\left[\log\frac{q_j^\tau}{p_j^\tau}\right] = \frac{1}{\tau}\mathbb{E}_{j\sim q^\tau}[\Delta_j] + C_\tau.$$

Table 5: Effect of temperature scaling on distillation losses. We report ROUGE-L mean and standard deviation over five random seeds. Avg. is computed over Self-Inst. and Vicuna.

| Method | Self-Instruct | Vicuna | Avg. (↑) | Δ Avg. |
|---|---|---|---|---|
| ***GPT-2 XL (1.5B) → GPT-2 Base (0.1B)*** | | | | |
| GPT-2 XL | 14.07± 0.37 | 16.31± 0.32 | 15.19 | – |
| FKL | 9.68± 0.49 | 14.53± 0.31 | 12.11 | – |
| FKL + $\tau$ | 12.16± 0.42 | **16.71**± 0.17 | 14.44 | ↑2.33 |
| RKL | 11.25± 0.35 | 15.82± 0.43 | 13.54 | – |
| RKL + $\tau$ | 11.97± 0.36 | 16.46± 0.38 | 14.22 | ↑0.68 |
| Sym-KL | 11.07± 0.43 | 15.72± 0.75 | 13.40 | – |
| Sym-KL + $\tau$ | 11.88± 0.24 | 16.35± 0.37 | 14.12 | ↑0.72 |
| JS | 11.67± 0.33 | 15.47± 0.37 | 13.57 | – |
| JS + $\tau$ | **12.71**± 0.37 | 16.05± 0.29 | 14.38 | ↑0.81 |
| SFKL | 11.57± 0.15 | 15.35± 0.30 | 13.46 | – |
| SFKL + $\tau$ | 11.77± 0.26 | 16.68± 0.59 | 14.23 | ↑0.77 |
| SRKL | 10.69± 0.21 | 14.96± 0.34 | 12.83 | – |
| SRKL + $\tau$ | 11.17± 0.40 | 15.42± 0.22 | 13.30 | ↑0.47 |
| AKL | 9.57± 0.24 | 13.70± 0.21 | 11.64 | – |
| AKL + $\tau$ | 12.64± 0.13 | 16.52± 0.34 | **14.58** | **↑2.94** |
| AB | 11.05± 0.21 | 15.82± 0.58 | 13.44 | – |
| DRKL | 12.34± 0.58 | 16.33± 0.31 | 14.34 | – |
| ***OPT 6.7B → OPT 1.3B*** | | | | |
| OPT 6.7B | 16.42± 0.69 | 17.64± 0.27 | 17.03 | – |
| FKL | 12.84± 0.32 | 16.71± 0.33 | 14.78 | – |
| FKL + $\tau$ | 15.21± 0.49 | 17.95± 0.57 | 16.58 | ↑1.80 |
| RKL | 12.29± 0.74 | 17.54± 0.13 | 14.92 | – |
| RKL + $\tau$ | 15.03± 0.40 | 18.11± 0.65 | 16.57 | ↑1.66 |
| Sym-KL | 13.56± 0.68 | 16.95± 0.37 | 15.26 | – |
| Sym-KL + $\tau$ | 14.94± 1.01 | **18.14**± 0.69 | 16.54 | ↑1.29 |
| JS | 12.63± 0.51 | 17.15± 0.39 | 14.89 | – |
| JS + $\tau$ | 14.93± 0.73 | 18.05± 0.37 | 16.49 | ↑1.60 |
| SFKL | 11.95± 0.84 | 16.87± 0.45 | 14.41 | – |
| SFKL + $\tau$ | 14.50± 0.41 | 17.87± 0.46 | 16.19 | ↑1.78 |
| SRKL | 11.99± 0.28 | 17.34± 0.25 | 14.67 | – |
| SRKL + $\tau$ | **15.41**± 0.45 | 18.04± 0.49 | **16.73** | **↑2.06** |
| AKL | 13.11± 0.31 | 17.26± 0.66 | 15.19 | – |
| AKL + $\tau$ | 15.27± 0.79 | 18.10± 0.24 | 16.69 | ↑1.50 |
| AB | 12.93± 0.47 | 17.59± 0.41 | 15.26 | – |
| DRKL | 14.37± 0.31 | 18.04± 0.24 | 16.21 | – |

Taking Eq. (5) minus this $\mathcal{L}_{\mathrm{RKL}}$, we obtain

$$\log\frac{q_i^\tau}{p_i^\tau} - \mathcal{L}_{\mathrm{RKL}} = \frac{1}{\tau}\left(\Delta_i - \mathbb{E}_{j\sim q^\tau}[\Delta_j]\right).$$

Substituting this into Eq. (3) gives

$$\nabla_{z_i^s}\mathcal{L}_{\mathrm{RKL}} = \frac{q_i^\tau}{\tau^2}\left(\Delta_i - \mathbb{E}_{j\sim q^\tau}[\Delta_j]\right).$$

This completes the proof. □

## D More Results

**Additional benchmarks support the main finding.** Table 5 reports results on Self-Instruct and Vicuna. Temperature improves every objective across both GPT-2 Base and OPT-1.3B, confirming that its benefit generalizes beyond the main benchmarks. For GPT-2 Base, AKL and FKL obtain

Table 6: Rouge-L (↑) for distillation from Llama3.2-3B to Llama3.2-1B using FKL, RKL, AKL, SFKL, and SRKL. $\Delta$ denotes the change in average Rouge-L from applying temperature tuning. Bold: best value per column.

| Loss | Dolly | Self-Inst | Super-NI | UnNI | Vicuna | Avg. | $\Delta$ |
|---|---|---|---|---|---|---|---|
| FKL | 25.86±0.54 | 15.46±1.03 | 28.14±0.25 | 29.52±0.13 | 15.53±0.04 | 22.90 | – |
| FKL+$\tau$ | 27.34±0.23 | 18.73±0.24 | 33.04±0.29 | 34.07±0.14 | 17.11±0.66 | 26.06 | +3.16 |
| RKL | 28.85±0.23 | 18.90±0.10 | **34.94**±0.15 | **35.90**±0.09 | **17.44**±0.90 | **27.21** | – |
| RKL+$\tau$ | 27.61±0.89 | 17.44±0.93 | 33.57±0.42 | 34.37±0.11 | 17.10±0.53 | 26.02 | −1.19 |
| AKL | 27.88±0.11 | 17.37±0.63 | 31.24±0.20 | 32.27±0.07 | 16.82±0.49 | 25.12 | – |
| AKL+$\tau$ | 28.31±0.61 | 18.39±1.10 | 34.31±0.48 | 34.83±0.15 | 17.00±0.40 | 26.57 | +1.45 |
| SFKL | 26.98±0.18 | 17.60±1.36 | 31.62±0.20 | 33.06±0.22 | 16.53±0.47 | 25.16 | – |
| SFKL+$\tau$ | 28.08±0.17 | 18.14±0.79 | 34.41±0.24 | 34.68±0.10 | 17.11±0.32 | 26.48 | +1.32 |
| SRKL | **28.93**±0.17 | **19.31**±0.73 | 32.88±0.27 | 34.28±0.12 | 16.94±0.58 | 26.47 | – |
| SRKL+$\tau$ | 27.80±0.50 | 18.07±0.53 | 34.56±0.31 | 34.80±0.09 | 16.94±0.37 | 26.38 | −0.09 |

Table 7: Rouge-L (↑) for distillation from Qwen3-4B to Qwen3-1.7B using FKL, RKL, AKL, SFKL, and SRKL. $\Delta$ denotes the change in average Rouge-L from applying temperature tuning. Bold: best value per column.

| Loss | Dolly | Self-Inst | Super-NI | UnNI | Vicuna | Avg. | $\Delta$ |
|---|---|---|---|---|---|---|---|
| FKL | 26.56±0.46 | 19.17±0.55 | 34.83±0.73 | 30.56±0.13 | 16.69±0.49 | 25.57 | – |
| FKL+$\tau$ | 27.75±0.71 | 19.47±0.66 | 35.36±0.29 | 32.73±0.20 | 17.91±0.50 | 26.65 | +1.08 |
| RKL | **29.39**±0.50 | 20.66±1.16 | **39.39**±0.03 | **36.48**±0.20 | **19.12**±0.51 | **28.99** | – |
| RKL+$\tau$ | 29.02±0.30 | 19.11±0.43 | 35.97±0.51 | 34.44±0.09 | 18.88±0.74 | 27.49 | −1.50 |
| AKL | 26.94±0.29 | 18.75±0.59 | 36.86±0.43 | 33.03±0.08 | 17.74±0.35 | 26.67 | – |
| AKL+$\tau$ | 28.04±0.45 | 19.37±0.30 | 35.39±0.26 | 33.42±0.14 | 18.09±0.69 | 26.86 | +0.19 |
| SFKL | 27.87±0.52 | **20.70**±1.59 | 36.88±0.25 | 33.44±0.19 | 17.28±0.13 | 27.23 | – |
| SFKL+$\tau$ | 28.20±0.36 | 20.22±0.10 | 35.77±0.34 | 33.63±0.03 | 18.45±0.39 | 27.25 | +0.02 |
| SRKL | 28.91±0.29 | 20.46±1.12 | 39.12±0.12 | 36.11±0.37 | 18.34±0.86 | 28.59 | – |
| SRKL+$\tau$ | 29.01±0.04 | 20.47±1.09 | 36.64±0.12 | 34.44±0.09 | 18.57±0.49 | 27.83 | −0.76 |

the largest gains, improving by +2.94 and +2.33 average ROUGE-L, respectively, and both surpass DRKL. For OPT-1.3B, all objectives improve by +1.29 to +2.06, with SRKL+$\tau$ achieving the best average score of 16.73. These results show that KD temperature is a broadly useful factor for KL-based distillation and can alter the ranking among objectives.

**Modern model families**. To further strengthen the empirical evidence on modern model families, we additionally evaluate two teacher–student pairs: Llama 3.2-3B → Llama 3.2-1B and Qwen3-4B → Qwen3-1.7B. The results are reported in Tables 6 and 7. Both modern models exhibit the same qualitative pattern observed for GPT-2 and OPT: FKL-based objectives benefits more from temperature tuning than RKL-based objectives. On Llama3.2, FKL, AKL, and SFKL improve by +3.16, +1.45, and +1.32 average Rouge-L, whereas RKL and SRKL decrease by −1.19 and −0.09. On Qwen3, FKL, AKL, and SFKL gain by +1.08, +0.19, and +0.02, while RKL and SRKL drop by −1.50 and −0.76. These results indicate that the asymmetric effect of temperature is not specific to any particular LLM architecture.

## E Case Studies

Tables 8 and 9 show qualitative examples from the Unnatural Instructions benchmark (Honovich et al., 2023). The OPT-1.3B student is distilled from the OPT-6.7B teacher under different distillation objectives, with and without KD temperature. Across both cases, applying temperature consistently improves instruction following and output format compared with the corresponding $\tau = 1$ objectives.

In Case study #1, the task requires extracting proper nouns in their original order. Without temperature, several objectives omit entities, add incorrect nouns, or generate unrelated continuations: RKL includes "Class," SFKL omits "Mr. Krebs," and JS/SRKL produce long hallucinated passages. With temperature, all KL-based objectives produce the exact ground-truth output: "Toby, Mr. Krebs, Mrs. Dvorak," suggesting better preservation of fine-grained constraints such as output format and

Table 8: Qualitative examples from the Unnatural Instructions (Honovich et al., 2023) benchmark. The OPT-1.3B student is distilled from the OPT-6.7B teacher. With KD temperature, all distillation objectives generate responses that better satisfy the given instructions compared with their $\tau = 1$ counterparts.

| | ***Case study # 1*** |
|---|---|
| Instruction | Given a paragraph, write down all the proper nouns in the order they occur. A proper noun is a specific name for a person, place, or thing and is always capitalized. |
| Input | Passage: Toby thought that Mr. Krebs was being unfair when he gave out detentions for talking during class, but Mrs. Dvorak said that it was part of his job as a teacher. |
| Ground-truth | Toby, Mr. Krebs, Mrs. Dvorak |
| FKL | Toby, Ms. Dvorak, and Mr. Krebs are all proper nouns. |
| FKL + $\tau$ | Toby, Mr. Krebs, Mrs. Dvorak |
| RKL | Toby, Mrs. Dvorak, Mr. Krebs, Class |
| RKL + $\tau$ | Toby, Mr. Krebs, Mrs. Dvorak |
| Sym-KL | Toby, Mr. Krebs, Mrs. Dvorak |
| Sym-KL + $\tau$ | Toby, Mr. Krebs, Mrs. Dvorak |
| JS | Toby, Mrs. Dvorak, and Mr. Krebs Waited for the perfect opportunity to strike back at Mr. Krabs. When he was delivering the class talk, he stopped cold and asked them to exchange their places. Toby assumed that Mr. Krabs was being unfair because he was giving out detentions for speaking during class. Mrs. Dvorak explained that it was part of her job as a teacher to make sure that class was taught properly. |
| JS + $\tau$ | Toby, Mr. Krebs, Mrs. Dvorak |
| SFKL | Toby, Mrs. Dvorak |
| SFKL + $\tau$ | Toby, Mr. Krebs, Mrs. Dvorak |
| SRKL | Toby, Mrs. Dvorak, and Mr. Krebs all agreed that Mr. Krebs should have given out detentions for discussing classes during class, but Mrs., Ms., and Mr. Krebs disagreed. Regarding Mr. Krebs' position on detentions, Mrs. Dvorac stated that it was his job to teach, and therefore he should teach what is required of him. Regarding Mr. Bacon's position on speech, Mrs. Dvorack said that he was a teacher, and therefore he should impart wisdom and knowledge to his students. Regarding Mr. Schmidt's position on language, Mrs. Dvoraken stated that he was a professor, and therefore he should use his knowledge to teach what is required of himself. |
| SRKL + $\tau$ | Toby, Mr. Krebs, Mrs. Dvorak |
| AKL | Toby, Mrs. Dvorak, and Mr. Krebs are all proper nouns. |
| AKL + $\tau$ | Toby, Mr. Krebs, Mrs. Dvorak |
| AB | 1. Toby thought that Mr. Krebs was being unfair when He gave out detentions For talking during class, but Mr. Dvorak said it was part of his Job as a teacher. <n>2. Mrs. Dvorak added that it was his job to teach, and he should know what was expected of him. <n>3. Toby agreed, and added that he would remember this when he got home. |
| DRKL | Toby, Mr. Krebs, Mrs. Dvorak |

entity selection.

Case study #2 is more challenging, but shows a similar trend. At $\tau = 1$, several objectives misunderstand the blank-filling instruction: FKL and RKL copy or corrupt the incomplete sentence, Sym-KL repeats the word options, and SRKL critiques the prompt instead of answering. With temperature, the outputs become more aligned with the intended completion, often expressing that the person did not want coffee. Although some responses do not exactly match the ground-truth word "no," they are substantially closer than their $\tau = 1$ counterparts. Overall, these examples show that KD temperature improves not only ROUGE-L scores, but also the student's ability to follow instructions and produce task-appropriate responses.

## F Use of AI Assistants

In preparing this paper, we made limited use of AI Assistants such as ChatGPT only to help refine wording, correct grammar, and enhance clarity. All core research contributions, including conceptual development, theoretical analysis, experimental evaluations, and paper organization and composition, were conducted entirely by the authors.

Table 9: Qualitative examples from the Unnatural Instructions (Honovich et al., 2023) benchmark. The OPT-1.3B student is distilled from the OPT-6.7B teacher. With KD temperature, all distillation objectives generate responses that better satisfy the given instructions compared with their $\tau = 1$ counterparts.

| | ***Case study # 2*** |
|---|---|
| Instruction | You will be given a piece of text that is missing important information, and a list of words. Your job is to fill in the blank with one of the words from the list so that the sentence/paragraph makes grammatical sense. |
| Input | I asked if he wanted to go for coffee, but he said _____.<n>Word options: ['no', 'not', 'didnt'] |
| Ground-truth | I asked if he wanted to go for coffee, but he said no. |
| FKL | "I asked if he wanted to going for coffee, but he says ____."<n><n>Thus, "I asked if he wanted go for coffee, but [name] said ____." This constitutes fixing the grammatical error in the sentence "I asked if he wanting to go for coffee, ____. |
| FKL + $\tau$ | I asked if he wanted to went for coffee, but he says he doesn't want to. |
| RKL | I asked if he wanted to going for coffee, but he says _____. |
| RKL + $\tau$ | I asked if he wanted to take a coffee, but he said 'not' |
| Sym-KL | word options: 'no', 'not','didnt' |
| Sym-KL + $\tau$ | I asked if he wanted to Go for coffee, but he says he doesn't want to. |
| JS | He would rather go for a walk. |
| JS + $\tau$ | He wanted to go for coffee but he said he didn't want to. |
| SFKL | The sentence is grammatically incorrect because it should read as:<n>I asked if he wanted to take a coffee break, but he said ____.<n>The correct sentence is:<n>I asked if He wanted to take a coffee walk, but he said _____.<n>The grammatical problem with the statement is that it is missing the crucial phrase - hence it is incorrect grammatically. |
| SFKL + $\tau$ | He asked if he wanted to Go for coffee, but he told me that he didn't want to go for coffee. |
| SRKL | Because you provided no information in the sentence, and only provided a list of words, the context is deeply misunderstood, and your answer is incorrect. You should have provided more information in the text to clarify the context. |
| SRKL + $\tau$ | He wants to go for coffee, but he says he doesn't want to go. |
| AKL | Iamar did not want to go for coffee; he said ["not", "didnt", "not"]. |
| AKL + $\tau$ | He wanted to go for coffee but he said he didn't want to. |
| AB | These statements are correct, but I would also recommend using the word "to" consistently in this case. In the first statement, "The glass fell three stories" is a complete statement, while "It broke into pieces" is also a complete statement. In the second statement, "A piece hit John on the head" is a correct statement, whereas "A piece hit John onthe head" and "A piece hit Johnon thehead" contain spacing errors and should be corrected. |
| DRKL | The correct answer is 'no'. Unless he specifically asks for a coffee, he doesn't want to go for coffee. |