Localizing Reasoning Training-Induced Changes in Large Language Models

Max Klabunde
University of Passau
max.klabunde@uni-passau.de

Florian Lemmerich
University of Passau

Abstract

Reasoning language models demonstrate excellent performance, but how reasoning training transforms models internally remains poorly understood. We systematically analyze where reasoning training induces changes by studying 19 publicly released reasoning models based on Qwen, trained using supervised fine-tuning (SFT) or reinforcement learning (RL). Using direct weight comparison and representation similarity via centered kernel alignment (CKA), we reveal a striking discrepancy: weight differences are broadly distributed, while CKA highlights middle layers as most strongly changed. We trace the CKA effect to "dominant datapoints" as proposed by Nguyen et al.–tokens with large activations that might affect attention mechanisms. Additionally, SFT induces substantially larger changes than RL. While our work makes steps towards better understanding of reasoning training, more work is needed to conclusively understand its effects.

1 Introduction

Reasoning models, i.e, large language models (LLMs) with long chain of thought, have become widely popular and have shown better performance compared to standard LLMs [12, 33].

Although reasoning models are widely adopted, our understanding of them is lacking. For example, it is not well understood what kind of changes different reasoning training approaches like reinforcement learning (RL) or supervised finetuning (SFT) create in LLMs. And even for a fixed training method like GRPO [26], it is not clear whether reasoning training only elicits existing capabilities of the base models [35, 25] or can install new capabilities [18].

In this paper, as a first step towards better understanding the effects of reasoning training, we aim to localize the changes in models from reasoning training. This will help our understanding of reasoning models and may accelerate future interpretability work by giving parts of models to focus on.

We analyze publicly released reasoning models that use Qwen-Math-Base-7B [34] or DeepSeek-R1-Distill-Qwen-7B/1.5B [12] as base models. The models differ among multiple dimensions including training data and training method (RL/SFT).

In our exploratory analysis with two different techniques–direct *weight comparison* and *representation comparison* with centered kernel alignment (CKA) [16]–we find a big discrepancy between the methods. While comparing weights does not highlight any particular layer as being especially affected by reasoning training, CKA puts surprisingly strong emphasis on middle layers as being different (Figure 1). We trace the CKA effects back to so-called dominant datapoints, as proposed by [24]. These might be further related to massive activations in LLMs, which are input-independent activations with huge magnitude that are used as attention sinks [30].

Overall, our work makes the following contributions: (1) We localize potentially interesting effects of reasoning training in middle layers. (2) We make progress towards a large-scale empirical study of

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Mechanistic Interpretability Workshop.

public reasoning models. (3) By comparing two methods to localize model changes, we highlight the value of a multi-method approach in comparing models.

Code is available at https://github.com/mklabunde/localizing-reasoning.

2 Related Work

Reasoning models have been studied in several recent works. For example, [1] use crosscoders to discover reasoning features in LLMs. While sparse autoencoders and variants have been widely used for interpretability [e.g., 23, 5], they are computationally expensive compared to other approaches.

Localizing important model components for reasoning was explored by [31]. They find that steering in middle layers is most effective for affecting reasoning behavior. [25] compare weights similar to this work on self-trained models, which eliminates confounders, but limits scope. Our work follows this simpler group of approaches, but uses a wider range of models.

3 Experimental Setup

Models We consider 19 reasoning models in our experiments, see Appendix A for the full list. These form three groups of models based on the model that they are based on: 3 descend from DeepSeek-R1-Distill-Qwen-7B [12], 11 from DeepSeek-R1-Distill-Qwen-1.5B, and 5 from Qwen-2.5-Math-Base-7B[1] [34]. All groups of models have at least one model that is trained with SFT, the others with RL, so we can compare effects across training methods.

Data for representations We use the first 50 problems from MATH-500 [17, 15] to generate reasoning traces for comparing representations. For this, we need alignment between tokens, such that the representations per trace can be compared across models. Thus, all models will generate representations from the same reasoning trace per problem (even if the trace was not generated by them). Each trace is hundreds to thousands of tokens long, so even a small number of traces provides robust similarity estimates.

Centered Kernel Alignment Given two collections of representations $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{N \times D}$, where N is the number of samples and D the dimensionality, CKA [16] returns a similarity score between 0 and 1. The comparison is based on the Hilbert-Schmidt Independence Criterion (HSIC) [11] between the two gram matrices $\boldsymbol{K}, \boldsymbol{L}$ of $\boldsymbol{X}, \boldsymbol{Y}$:

$$\mathrm{CKA}(\boldsymbol{K}, \boldsymbol{L}) = \frac{\mathrm{HSIC}(\boldsymbol{K}, \boldsymbol{L})}{\sqrt{\mathrm{HSIC}(\boldsymbol{K}, \boldsymbol{K})\,\mathrm{HSIC}(\boldsymbol{L}, \boldsymbol{L})}} \tag{1}$$

We compute these matrices of pairwise similarities with a linear and RBF kernel (scaled by 0.8 times the median distance, as suggested by [16]). We use the unbiased estimator of HSIC [27]. Linear CKA can also be expressed in terms of principal components of $\boldsymbol{X}$ and $\boldsymbol{Y}$ and their explained variance (see Appendix B.1).

Weight Difference Since the reasoning model has the same architecture as the base model, we directly compare weights of corresponding components. Given the weight matrices $\boldsymbol{W}_{base}$ and $\boldsymbol{W}_{new}$, we compute the normalized weight difference as

$$\frac{\|\boldsymbol{W}_{new} - \boldsymbol{W}_{base}\|_F}{0.5\|\boldsymbol{W}_{base}\|_F + 0.5\|\boldsymbol{W}_{new}\|_F}. \tag{2}$$

The normalization helps identify changes when the weight norm was small initially.

4 Results

4.1 Representation Comparison with CKA

We compare representations between the base model and the reasoning model. We always compare the representations per layer of the response for a specific problem in MATH-500, and report the

[1] One model descends of a variant with extended context length. However we found no practical difference in our experiments.

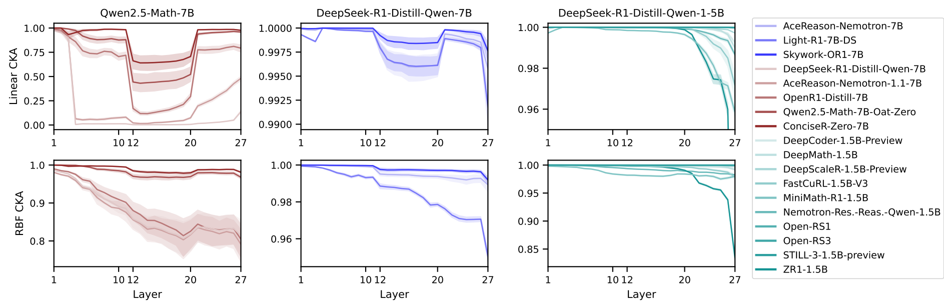


Figure 1: Average representational similarity distribution as measured by linear CKA (top) and RBF CKA (bottom) over 50 reasoning traces of DeepSeek-R1-Distill-Qwen-7B for MATH-500. The similarity of the final layer is always much lower and thus excluded from the plot. The 7B models all show a dip in linear representational similarity in the middle layers. Although it is most obvious for linear CKA, it can also be seen to a lesser extent with RBF CKA.

mean similarity over 50 samples as measured by linear and RBF CKA in Figure 1. For additional results, see Appendix B.2.

Suprising similarity pattern Linear CKA shows a surprising "bathtub" pattern for all 7B models: similarity starts out high, then drops in the middle layers, and finally recovers at the end. For most models, the drop starts at layer 12 and the recovery at layer 21. However, two models descending from Qwen-Math have the drop earlier at layer 4 and (imperfect) recovery near the final layers. The final layer representations are always much less similar than the rest. The 1.5B models do not show the pattern in the middle of the models.

RBF CKA has a larger focus on the local neighborhood of representation and thus provides a different perspective on representational similarity. Here, similarity goes down more consistently, but still shows a dip near layer 12.

Where does the pattern come from? Prior work [24, 7] points out potential issues with linear CKA. We find that the "dominant datapoints" approach from [24] mostly explains the bathtub. Dominant datapoints in our contexts are activations of tokens that have much larger projection on the first principal component of the collected activations than other tokens. When the first principal component explains much of the variance of the activations, the dominant data points have large influence on the CKA score.

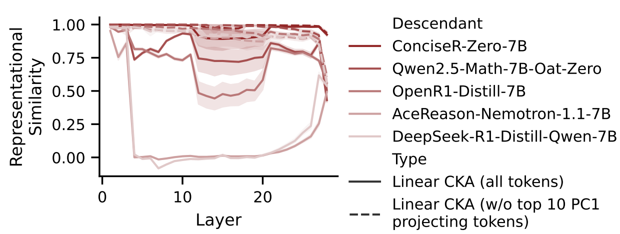


Figure 2: Regular linear CKA (solid lines) and linear CKA without dominant data points (dashed). The bathtub pattern disappears when the 10 tokens that get maximally projected to the first principal component are excluded from the comparison. Instead representational similarity decreases slowly consistently from early to late layers.

In Figure 2, we show how the bathtub disappears when we remove the top 10 dominant datapoints. Since they might differ between base and reasoning model, we remove the union of the top 10 tokens for each model (i.e., up to 20 tokens if they are completely distinct).

But why do dominant datapoints exist? Massive activations [30] appear commonly in LLMs: usually data-independent tokens that receive representations with magnitude much larger than the median token. They are likely used as attention sinks [30]. We observe that activation norm and projection on the first principal component are highly correlated in the bathtub layers and that there is large overlap between dominant data points and the tokens with largest activation norm (average ≈ 0.4 in middle layers). Hence, reasoning training might affect which tokens receive massive activations and thus act

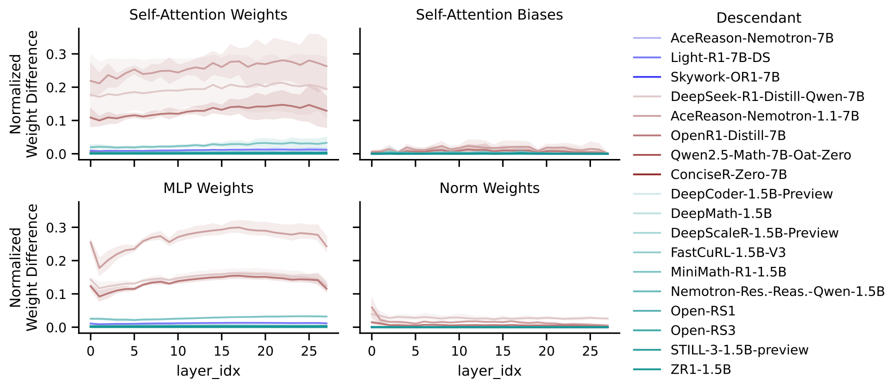


Figure 3: Weight difference between base model and descendants from Qwen-Math-7B, DeepSeek-R1-Distill-Qwen-7B, and DeepSeek-R1-Distill-Qwen-1.5B. Initial reasoning training leads to substantial changes in weights of attention layers and MLPs, whereas further reasoning training of the distilled DeepSeek-R1 models makes only minor changes in weights. In contrast to CKA, weight changes are broadly distributed across layers.

as attention sinks. Thus, this is an indicator that the attention mechanism is significantly affected by reasoning training. However, a detailed and conclusive investigation is left for future work.

4.2 Weight Difference

To verify the previous results, we use a simple approach of comparing weights directly. Figure 3 shows the results color-coded by the three groups of models. The weight matrices in the self-attention layers and the MLPs change relatively consistently throughout the whole network. Thus, weight difference does not clearly indicate localized changes for reasoning, similar to CKA (after taking dominant datapoints into account). This is in contrast to the recent work by [25], who find localized changes in a GRPO-trained OLMo model. However, we provide additional evidence that SFT leads to larger changes in models than RL, which could also be seen with RBF CKA. In general, only the reasoning training of the Qwen-Math-based models seems to induce large changes. The other models are only slightly affected by additional training.

5 Discussion and Conclusion

Conflict between diffing methods We used three lenses to analyze reasoning models: weight comparison, linear CKA, and RBF CKA. As they all take slightly different perspectives on what makes two models similar, it perhaps unsurprising that they give conflicting results. Weight comparison does not highlight any model component as especially changed by reasoning training. In contrast, CKA highlights middle layers in some of the models–however, some of the effect could be explained by unintuitive reliance on few datapoints, which needs to be explored in future work. Overall, this emphasizes the value of using multiple unsupervised methods for initial exploration.

Implications for analysis of reasoning models Should the difference of representations in middle layers be resolved as an artifact of the similarity measure, reasoning could be a distributed capability that relies on many different model components and might emerge from the combination of other capabilities. Otherwise, these middle layers may be useful objects of study.

Further, we see that SFT models change much more than RL-trained reasoning models, which points towards possibly distinct mechanisms that these training methods create in reasoning models. These differences are interesting directions for future work.

Limitations We only study models that descend from Qwen models. Other models may show different behavior [9]. Still, Qwen models are popular as a base for reasoning models. Further, we only use a single small dataset from the math domain. Finally, we do not establish any causal links between the model changes and reasoning behavior, which is an exciting direction of future work.

References

[1] David D Baek and Max Tegmark. Towards understanding distilled reasoning models: A representational approach. *arXiv preprint arXiv:2503.03730*, 2025.

[2] Yang Chen, Zhuolin Yang, Zihan Liu, Chankyu Lee, Peng Xu, Mohammad Shoeybi, Bryan Catanzaro, and Wei Ping. Acereason-nemotron: Advancing math and code reasoning through reinforcement learning. *arXiv preprint arXiv:2505.16400*, 2025.

[3] Zhipeng Chen, Yingqian Min, Beichen Zhang, Jie Chen, Jinhao Jiang, Daixuan Cheng, Wayne Xin Zhao, Zheng Liu, Xu Miao, Yang Lu, et al. An empirical study on eliciting and improving r1-like reasoning models. *arXiv preprint arXiv:2503.04548*, 2025.

[4] Xiangxiang Chu, Hailang Huang, Xiao Zhang, Fei Wei, and Yong Wang. Gpg: A simple and strong reinforcement learning baseline for model reasoning. *arXiv preprint arXiv:2504.02546*, 2025.

[5] Hoagy Cunningham, Aidan Ewart, Logan Riggs, Robert Huben, and Lee Sharkey. Sparse autoencoders find highly interpretable features in language models. *arXiv preprint arXiv:2309.08600*, 2023.

[6] Quy-Anh Dang and Chris Ngo. Reinforcement learning for reasoning in small llms: What works and what doesn't. *arXiv preprint:2503.16219*, 2025.

[7] MohammadReza Davari, Stefan Horoi, Amine Natik, Guillaume Lajoie, Guy Wolf, and Eugene Belilovsky. Reliability of CKA as a similarity measure in deep learning. In *The Eleventh International Conference on Learning Representations*, 2023.

[8] Hugging Face. Open r1: A fully open reproduction of deepseek-r1, January 2025.

[9] Kanishk Gandhi, Ayush Chakravarthy, Anikait Singh, Nathan Lile, and Noah D Goodman. Cognitive behaviors that enable self-improving reasoners, or, four habits of highly effective stars. *arXiv preprint arXiv:2503.01307*, 2025.

[10] Jeremiah Greer. Minimath-r1-1.5b, February 2025.

[11] Arthur Gretton, Kenji Fukumizu, Choon Teo, Le Song, Bernhard Schölkopf, and Alex Smola. A kernel statistical test of independence. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007.

[12] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

[13] Jujie He, Jiacai Liu, Chris Yuhao Liu, Rui Yan, Chaojie Wang, Peng Cheng, Xiaoyu Zhang, Fuxiang Zhang, Jiacheng Xu, Wei Shen, Siyuan Li, Liang Zeng, Tianwen Wei, Cheng Cheng, Bo An, Yang Liu, and Yahui Zhou. Skywork open reasoner 1 technical report. *arXiv preprint arXiv:2505.22312*, 2025.

[14] Zhiwei He, Tian Liang, Jiahao Xu, Qiuzhi Liu, Xingyu Chen, Yue Wang, Linfeng Song, Dian Yu, Zhenwen Liang, Wenxuan Wang, Zhuosheng Zhang, Rui Wang, Zhaopeng Tu, Haitao Mi, and Dong Yu. Deepmath-103k: A large-scale, challenging, decontaminated, and verifiable mathematical dataset for advancing reasoning. *arXiv preprint arXiv:2504.11456*, 2025.

[15] Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the math dataset. *NeurIPS*, 2021.

[16] Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *International conference on machine learning*, pages 3519–3529. PMlR, 2019.

[17] Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harrison Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. In *The Twelfth International Conference on Learning Representations*, 2023.

[18] Mingjie Liu, Shizhe Diao, Ximing Lu, Jian Hu, Xin Dong, Yejin Choi, Jan Kautz, and Yi Dong. Prorl: Prolonged reinforcement learning expands reasoning boundaries in large language models. *arXiv preprint arXiv:2505.24864*, 2025.

[19] Zichen Liu, Changyu Chen, Wenjun Li, Penghui Qi, Tianyu Pang, Chao Du, Wee Sun Lee, and Min Lin. Understanding r1-zero-like training: A critical perspective. *arXiv preprint arXiv:2503.20783*, 2025.

[20] Zihan Liu, Zhuolin Yang, Yang Chen, Chankyu Lee, Mohammad Shoeybi, Bryan Catanzaro, and Wei Ping. Acereason-nemotron 1.1: Advancing math and code reasoning through sft and rl synergy. *arXiv preprint arXiv:2506.13284*, 2025.

[21] Michael Luo, Sijun Tan, Roy Huang, Ameen Patel, Alpay Ariyak, Qingyang Wu, Xiaoxiang Shi, Rachel Xin, Colin Cai, Maurice Weber, Ce Zhang, Li Erran Li, Raluca Ada Popa, and Ion Stoica. Deepcoder: A fully open-source 14b coder at o3-mini level. `https://pretty-radio-b75.notion.site/DeepCoder-A-Fully-Open-Source-14B-Coder-at-O3-mini-Level-1cf81902c14680b3bee5eb349a512a51`, 2025. Notion Blog.

[22] Michael Luo, Sijun Tan, Justin Wong, Xiaoxiang Shi, William Y. Tang, Manan Roongta, Colin Cai, Jeffrey Luo, Li Erran Li, Raluca Ada Popa, and Ion Stoica. Deepscaler: Surpassing o1-preview with a 1.5b model by scaling rl. `https://pretty-radio-b75.notion.site/DeepScaleR-Surpassing-O1-Preview-with-a-1-5B-Model-by-Scaling-RL-19681902c1468005bed8ca3030`
2025. Notion Blog.

[23] Julian Minder, Clément Dumas, Caden Juang, Bilal Chugtai, and Neel Nanda. Robustly identifying concepts introduced during chat fine-tuning using crosscoders. *arXiv preprint arXiv:2504.02922*, 2025.

[24] Thao Nguyen, Maithra Raghu, and Simon Kornblith. On the origins of the block structure phenomenon in neural network representations. *Trans. Mach. Learn. Res.*, 2022, 2022.

[25] Neel Rajani, Aryo Pradipta Gema, Seraphina Goldfarb-Tarrant, and Ivan Titov. Scalpel vs. hammer: Grpo amplifies existing capabilities, sft replaces them. *Actionable Interpretability Workshop ICML 2025*, 2025.

[26] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

[27] Le Song, Alex Smola, Arthur Gretton, Justin Bedo, and Karsten Borgwardt. Feature selection via dependence maximization. *Journal of Machine Learning Research*, 13(47):1393–1434, 2012.

[28] Mingyang Song and Mao Zheng. Walk before you run! concise llm reasoning via reinforcement learning, 2025.

[29] Mingyang Song, Mao Zheng, Zheng Li, Wenjie Yang, Xuan Luo, Yue Pan, and Feng Zhang. Fastcurl: Curriculum reinforcement learning with stage-wise context scaling for efficient training r1-like reasoning models. *arXiv preprint arXiv:2503.17287*, 2025.

[30] Mingjie Sun, Xinlei Chen, J Zico Kolter, and Zhuang Liu. Massive activations in large language models. In *First Conference on Language Modeling*, 2024.

[31] Constantin Venhoff, Iván Arcuschin, Philip Torr, Arthur Conmy, and Neel Nanda. Understanding reasoning in thinking language models via steering vectors. *arXiv preprint arXiv:2506.18167*, 2025.

[32] Liang Wen, Yunke Cai, Fenrui Xiao, Xin He, Qi An, Zhenyu Duan, Yimin Du, Junchen Liu, Lifu Tang, Xiaowei Lv, Haosheng Zou, Yongchao Deng, Shousheng Jia, and Xiangzheng Zhang. Light-r1: Curriculum sft, dpo and rl for long cot from scratch and beyond, 2025.

[33] An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

[34] An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, et al. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*, 2024.

[35] Yang Yue, Zhiqi Chen, Rui Lu, Andrew Zhao, Zhaokai Wang, Shiji Song, and Gao Huang. Does reinforcement learning really incentivize reasoning capacity in llms beyond the base model? *arXiv preprint arXiv:2504.13837*, 2025.

[36] Zyphra. Zr1-1.5b: A small but powerful reasoning model for math and code, 2025.

Table 1: All models used in the experiments.

Basemodel	Model	Model Training	Huggingface Repository
Qwen2.5-Math-7B-Base [34]	OpenR1-Distill-7B* [8]	SFT	https://huggingface.co/open-r1/OpenR1-Distill-7B
	DeepSeek-R1-Qwen-7B [12]	SFT	https://huggingface.co/deepseek-ai/DeepSeek-R1-Distill-Qwen-7B
	AceReason-Nemotron 1.1-7B [20]	SFT + RL	https://huggingface.co/nvidia/AceReason-Nemotron-1.1-7B
	Qwen2.5-Math-7B-Oat-Zero [19]	RL	https://huggingface.co/sail/Qwen2.5-Math-7B-Oat-Zero
	ConciseR-7B [28]	RL	https://huggingface.co/Nickyang/ConciseR-Zero-7B
DeepSeek-R1-Distill-Qwen-7B [12]	AceReason-7B [2]	RL	https://huggingface.co/nvidia/AceReason-Nemotron-7B
	Skywork-OR1-7B [13]	RL	https://huggingface.co/Skywork/Skywork-OR1-7B
	Light-R1-7B-DS [32]	SFT	https://huggingface.co/qihoo360/Light-R1-7B-DS
DeepSeek-R1-Distill-Qwen-1.5B [12]	Nemotron-Research-Reasoning-Qwen-1.5B [18]	RL	https://huggingface.co/nvidia/Nemotron-Research-Reasoning-Qwen-1.5B
	ZR1-1.5 [36]	RL	https://huggingface.co/Zyphra/ZR1-1.5B
	DeepScaleR-1.5B [22]	RL	https://huggingface.co/agentica-org/DeepScaleR-1.5B-Preview
	DeepCoder-1.5B [21]	RL	https://huggingface.co/agentica-org/DeepCoder-1.5B-Preview
	Open-RS3 [6]	RL	https://huggingface.co/knoveleng/Open-RS3
	DeepMath-1.5B [14]	RL	https://huggingface.co/zwhe99/DeepMath-1.5B
	MiniMath-R1-1.5B [10]	SFT	https://huggingface.co/oumi-ai/MiniMath-R1-1.5B
	FastCuRL-1.5B-V3 [29]	RL	https://huggingface.co/Nickyang/FastCuRL-1.5B-V3
	GPG-RS1 [4]	RL	https://huggingface.co/GD-ML/Open-RS1
	Still-3-1.5B [3]	RL	https://huggingface.co/RUC-AIBOX/STILL-3-1.5B-preview

*: Descends from a context-extended Qwen2.5-Math model.

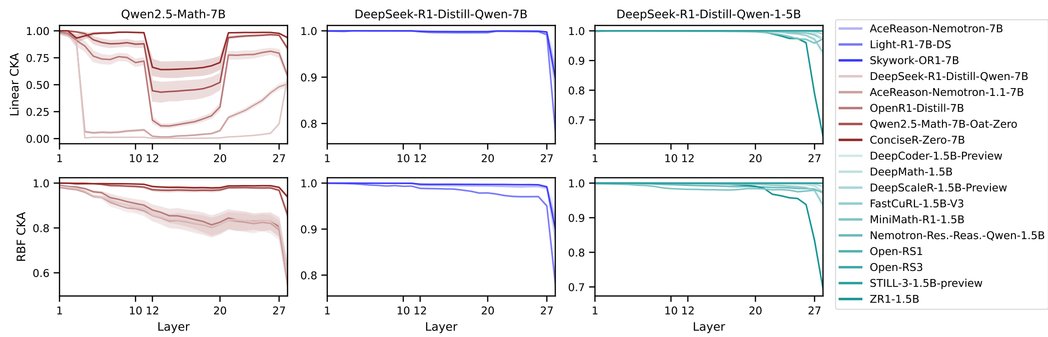


Figure 4: CKA comparison with final layer included, using DeepSeek-R1-Distilled-Qwen-7B reasoning traces.

A Models

See Table 1 for the comprehensive list of models used.

B Additional Information for CKA

B.1 Principal Component Formulation of Linear CKA

As noted by [16], CKA with a linear kernel can be formulated as follows for two matrices of representations $\boldsymbol{X} \in \mathbb{R}^{N \times D_1}, \boldsymbol{Y} \in \mathbb{R}^{N \times D_2}$:

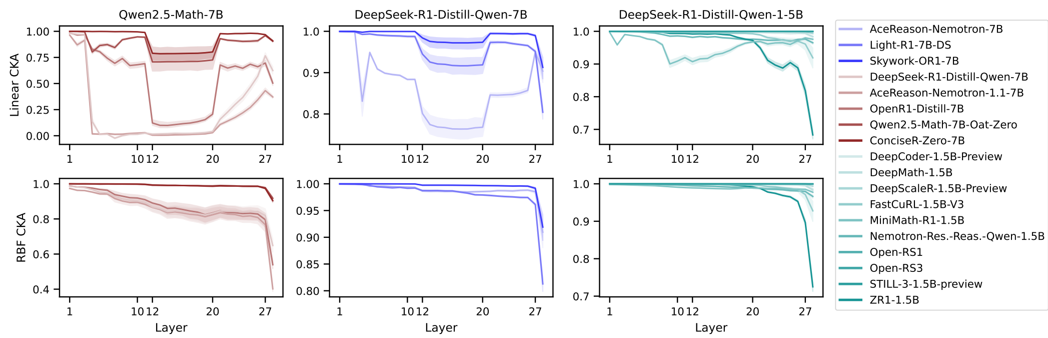
$$\mathrm{CKA}_{linear}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{\sum_{i=1}^{D_1} \sum_{j=1}^{D_2} \lambda_X^i \lambda_Y^j \langle u_X^i, u_Y^j \rangle^2}{\sqrt{\sum_{i=1}^{D_1} (\lambda_X^i)^2} \sqrt{\sum_{j=1}^{D_2} (\lambda_Y^j)^2}}, \tag{3}$$

where $\boldsymbol{u}_X^i \in \mathbb{R}^N$ and $\boldsymbol{u}_Y^i \in \mathbb{R}^N$ are the i-th normalized principal components of $\boldsymbol{X}, \boldsymbol{Y}$, respectively, and λ_X^i, λ_Y^i are the fraction of variance explained by the i-th principal component.

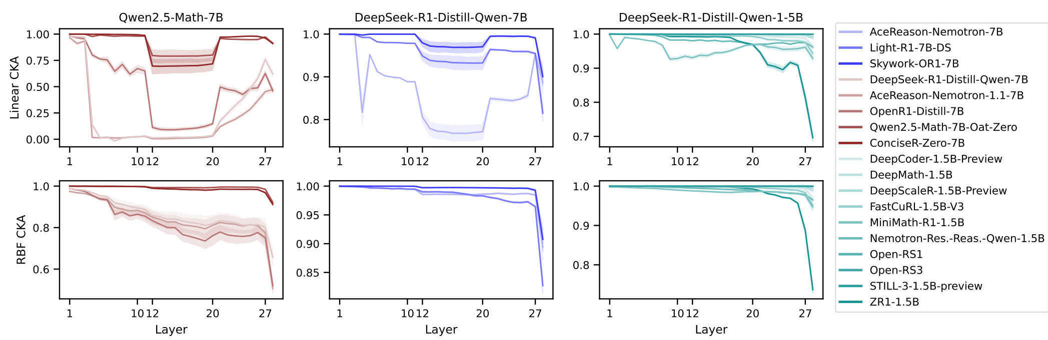
B.2 Additional Results

In Figure 4, we show the results from Figure 1 including the final layer.

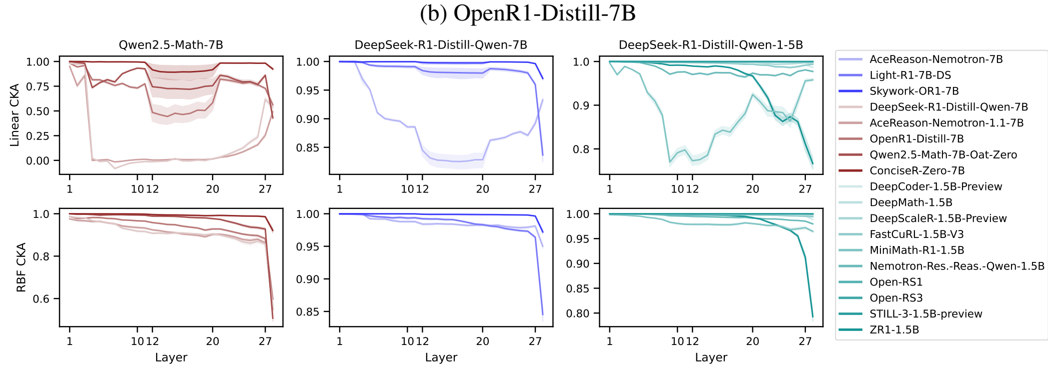
In Figure 5, we show the results using CKA with the representations from other reasoning traces. Despite different traces, the graphs are very similar.



(a) AceReason-Nemotron-1.1-7B



(b) OpenR1-Distill-7B



(c) Qwen2.5-Math-7B

Figure 5: CKA Comparison with reasoning traces generated by other models.