DO REASONING MODELS ENHANCE EMBEDDING MODELS?

Anonymous authors
Paper under double-blind review

ABSTRACT

Reinforcement Learning with Verifiable Rewards (RLVR) makes large language models better problem solvers, but does it also make their representations better foundations for text embeddings? We compare embedding models initialized from matched base backbones and backbones tuned with RLVR, using identical recipes for contrastive learning. Across MTEB Multilingual, MTEB Code, and BRIGHT, RLVR initialization provides no consistent advantage: the absolute paired mean differences reach at most 0.67 points. To explain this result, we introduce Hierarchical Representation Similarity Analysis (HRSA), which separates three properties often conflated by a single similarity measure: feature correspondence at the coordinate level, global and local manifold geometry, and functional readouts relevant to downstream tasks. HRSA shows that RLVR can reorganize local manifold geometry and, under prolonged training, shift the coordinate basis, while largely preserving global manifold geometry and linear readout structure. When the base and reasoning backbones are subsequently trained for embeddings, their shared contrastive objective rapidly pulls them toward the same global and functional organization, although differences in local manifold geometry persist. We call this phenomenon Manifold Realignment. In contrast, supervised fine tuning more strongly disrupts coordinate and geometric alignment. These findings suggest that RLVR primarily changes how a model traverses and locally organizes an existing semantic landscape rather than constructing a globally superior one. Consequently, stronger reasoning behavior does not necessarily produce better embedding models.

1 INTRODUCTION

Reinforcement Learning with Verifiable Rewards (RLVR) has made large language models (LLMs) markedly better problem solvers (DeepSeek-AI, 2025). Text embeddings, meanwhile, remain a core abstraction in natural language processing (NLP), giving retrieval, clustering, and matching a shared semantic space (Mikolov et al., 2013). Recent embedding models increasingly build on decoder-only LLM backbones, training them with contrastive objectives so that they inherit the semantic knowledge acquired during pre-training (Lee et al., 2025a; Shanbhogue et al., 2026). Since both lines of work rest on the same backbone, a natural question follows: *does RLVR also make these backbones better foundations for text embeddings?* Intuitively, stronger reasoning ought to yield representations that organize semantic relationships more effectively.

Counter-intuitively, our results reveal a **null effect**. Across comprehensive benchmarks including MTEB Multilingual (Enevoldsen et al., 2025), MTEB Code (Muennighoff et al., 2023), and BRIGHT (Su et al., 2025), embedding models initialized from RLVR-tuned reasoning models perform nearly identical to base-initialized models after contrastive learning (van den Oord et al., 2018; Gao et al., 2021). This observation presents a scientific puzzle: *Why do reasoning and non-reasoning backbones yield indistinguishable results following contrastive learning?*

Understanding this puzzle requires looking beyond aggregate performance. Two models may behave similarly while using different features, or they may use different coordinate systems while preserving the same underlying semantic relationships. Conversely, two representations may appear globally similar while differing sharply in their local neighborhoods. Treating all of these cases as a single notion of "representation similarity" obscures where post-training changes a model. To

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

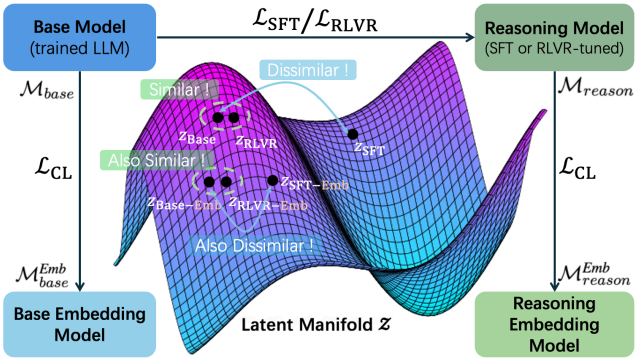


Figure 1: **Latent manifold and model relationships.** **z** indicates the representations of the corresponding models. Suffix "-Emb" is added to the model name to indicate the embedding model. We demonstrate the ideas of similar and dissimilar representations of RLVR-tuned pairs and SFT-tuned pairs, respectively.

Table 1: **Mean embedding benchmark performance (3 seeds).** We compare the base backbone $\mathcal{M}_{base}$ versus its RLVR-tuned reasoning model backbone $\mathcal{M}_{reason}$; we also include an SFT-tuned backbone for reference. The Δ (**Std**) column (gray) shows the mean performance gap $\pm$ standard deviation. The near-zero deltas for RLVR indicate that RLVR largely preserves the base models semantic effectiveness, contrasting with larger shifts under SFT.

	MTEB(Multilingual, v2)			MTEB(Code, v1)			BRIGHT		
Model Pair Backbone	$\mathcal{M}_{base}^{Emb}$	$\mathcal{M}_{reason}^{Emb}$	Δ (Std)	$\mathcal{M}_{base}^{Emb}$	$\mathcal{M}_{reason}^{Emb}$	Δ (Std)	$\mathcal{M}_{base}^{Emb}$	$\mathcal{M}_{reason}^{Emb}$	Δ (Std)
SFT									
Qwen3-0.6B-Base vs Qwen3-0.6B	53.50	41.47	-12.03 $\pm$ 0.14	55.29	56.28	+0.99 $\pm$ 0.05	13.06	13.71	+0.65 $\pm$ 0.06
Qwen2.5-Math-1.5B vs DS-Distill-1.5B	44.12	46.19	+2.07 $\pm$ 0.10	44.89	46.47	+1.58 $\pm$ 0.19	7.47	9.02	+1.55 $\pm$ 0.12
RLVR									
Qwen2.5-1.5B vs Qwen2.5-1.5B-SRL-Zoo	54.73	54.54	-0.19 $\pm$ 0.07	58.98	58.72	-0.26 $\pm$ 0.08	17.71	17.89	+0.18 $\pm$ 0.04
Qwen2.5-0.5B vs Qwen2.5-0.5B-SRL-Zoo	51.25	51.27	+0.02 $\pm$ 0.09	57.41	57.53	+0.12 $\pm$ 0.07	14.05	13.99	-0.06 $\pm$ 0.05
DS-Distill-1.5B vs NV-ProRL	46.19	46.25	+0.06 $\pm$ 0.06	45.47	45.87	+0.40 $\pm$ 0.07	9.02	9.47	+0.45 $\pm$ 0.02
Qwen3-4B vs Qwen3-4B-PSR	59.85	59.79	-0.06 $\pm$ 0.05	63.90	64.57	+0.67 $\pm$ 0.03	18.10	18.17	+0.07 $\pm$ 0.03

address this problem, we introduce **H**ierarchical **R**epresentation **S**imilarity **A**nalysis (**HRSA**), a conceptual framework that separates model similarity into three nested levels:

- **Representation level:** whether two models express information through the same coordinate basis and corresponding features.
- **Geometry level:** whether their latent spaces preserve the same global shape and local neighborhood relationships, independent of the particular coordinate system.
- **Function level:** whether their representations support the same input–output mappings and task-relevant readouts.

HRSA motivates our central account, which we call *Manifold Realignment*. Under this account, RLVR does not need to construct an entirely new latent manifold in order to improve reasoning. Instead, it can change how the model traverses and locally organizes an existing manifold while leaving much of its global geometry and task-relevant functional structure intact. Prolonged optimization may also move the coordinate basis, but such movement does not necessarily imply that the underlying latent manifold has been replaced.

When a base backbone and its reasoning-tuned descendant are subsequently adapted with the same contrastive learning, that shared objective acts as a strong organizing constraint. Because the two backbones retain a common global semantic structure, contrastive learning can draw their latent manifold back toward a common organization. This global realignment can coexist with persistent differences in local geometry: the models may agree on the broad semantic landscape and on the functions it supports without agreeing on every local neighborhood. Manifold Realignment therefore offers an account of how reasoning post-training can produce meaningful internal change without yielding a fundamentally superior embedding space.

Overall, these results suggest that, unlike SFT, RLVR primarily optimizes trajectories through an existing semantic landscape rather than fundamentally redrawing that landscape. Our contributions:

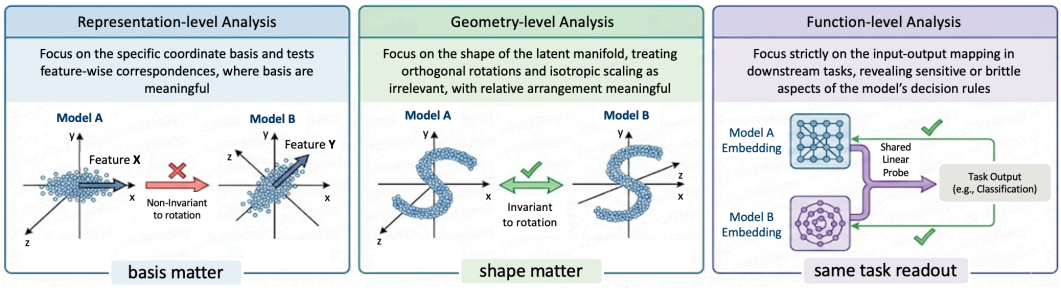


Figure 2: The overview of HRSA.

1. We formulate the reasoning–embedding puzzle: stronger reasoning behavior does not necessarily imply a better semantic embedding space.
2. We introduce **HRSA**, which organizes representation similarity into coordinate, geometry, and function levels according to what structure each level preserves.
3. We propose **Manifold Realignment** as a high-level account of how RLVR-induced representational changes interact with subsequent contrastive learning.

2 Embedding Model Performances

We first unify our terminology by separating a *starting checkpoint* from the *fine-tuning stage*.

Backbone LLMs. Given a base model $\mathcal{M}_{base}$ as a trained LLM, we consider a reasoning model $\mathcal{M}_{reason}$ as an LLM that undergoes either Supervised Fine-Tuning (SFT) or RLVR directly on top of the base model. We focus on *zero-RL* (DeepSeek-AI, 2025) where RLVR starts directly from the base model without performing a warm-start SFT stage. Concretely, we evaluate and compare a matched pair of $\mathcal{M}_{base}$ and $\mathcal{M}_{reason}$, where $\mathcal{M}_{reason}$ must be fine-tuned on $\mathcal{M}_{base}$. The SFT-tuned pairs are used as an explicit control to highlight the very close similarity observed in the RLVR-tuned comparisons.

Embedding models. We term the base embedding model $\mathcal{M}^{Emb}_{base}$ and reasoning embedding model $\mathcal{M}^{Emb}_{reason}$ (where suffix "-*Emb*" is added to the model name) with the backbone $\mathcal{M}_{base}$ and $\mathcal{M}_{reason}$ respectively. The embedding models are formed by removing the language modeling head and applying a pooling operator to the final-layer hidden states to produce a fixed-dimensional vector. We train embedding models with an InfoNCE objective (van den Oord et al., 2018). Within a pair the two embedding models share identical architectures and training recipes, differing only in their backbone initialization (i.e., $\mathcal{M}_{base}$ vs. $\mathcal{M}_{reason}$). Appendix A provides training details.

To rigorously assess the impact of RLVR across retrieval and clustering tasks, we evaluated matched pairs of $\mathcal{M}^{Emb}_{base}$ and $\mathcal{M}^{Emb}_{reason}$ on the MTEB and BRIGHT benchmarks. As presented in Table 1, $\mathcal{M}^{Emb}_{reason}$ with SFT-tuned backbone induces a 12% performance drop, demonstrating that it fundamentally restructures and destroys the pretrained semantic space. In contrast, $\mathcal{M}^{Emb}_{reason}$ with RLVR-tuned backbone consistently achieve performance parity with a near 0% change. Rather than interpreting this as a limitation, we view this parity as a profound indicator of representational robustness, demonstrating that RLVR safely refines trajectory generation policies without overwriting the rich world knowledge established during pre-training.

3 The HRSA Framework: Dissecting Model Similarity

In Section 2, we established that $\mathcal{M}^{Emb}_{reason}$ equipped with an RLVR tuned backbone maintains performance parity with $\mathcal{M}^{Emb}_{base}$. This macroscopic observation suggests that RLVR preserves the intrinsic geometry of the pre-trained Latent Manifold $\mathcal{Z}$, altering only the traversal policy rather than the landscape itself. Because aggregate benchmark scores often mask internal representational shifts, we introduce HRSA to rigorously dissect the relationship between $\mathcal{M}_{base}$ and $\mathcal{M}_{reason}$ across three nested levels of abstraction. As a diagnostic framework, HRSA is defined by strict invariance properties rather than specific metrics, allowing researchers to flexibly substitute alternative metrics that respect these invariance rules (Table 7).

3.1 Common Setup and Notation

To analyze the structural differences between models, we compare their representations on a shared sequence of N token positions. Let $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{N \times D}$ denote the D-dimensional token-level representation matrices produced by $\mathcal{M}_{base}$ (or $\mathcal{M}_{base}^{Emb}$) and $\mathcal{M}_{reason}$ (or $\mathcal{M}_{reason}^{Emb}$) respectively. The i-th row of each matrix, denoted as $\mathbf{x}_i$ or $\mathbf{y}_i$, represents a single token embedding such that $\mathbf{x}_i, \mathbf{y}_i \in \mathcal{Z}$, where $\mathcal{Z} \subset \mathbb{R}^D$ is the latent manifold induced by the distribution of the mapped inputs within the high-dimensional ambient space.

3.2 Representation-Level Analysis

Representation-level analysis focuses on the *specific coordinate basis* of the latent manifold and tests feature-wise correspondences between models. At this level, we treat the coordinate basis themselves as meaningful objects, where rotating or permuting features can change the outcome of the analysis. In other words, representation-level metrics are *not* invariant to orthogonal transformations or neuron permutations. See Appendix B.1 for a formal discussion. Intuitively, this level investigates *whether two models implement similar features along similar coordinate basis or realize similar solutions in very different coordinate systems*. We operationalize this question with two complementary tools:

1. **Dimension-Wise Correlation**: probe direct, axis-aligned feature correspondence (Beyer et al., 2020).
2. **Orthogonal Procrustes Analysis**: probe global linear equivalence (Schönemann, 1966).

3.2.1 Dimension-Wise Correlation

Dimension-wise correlation tests whether each coordinate in one model can be matched to the *same* coordinate in another model. Let $\mathbf{X}_{:j}$ and $\mathbf{Y}_{:j}$ denote the j-th column vectors (features) of the matrices $\mathbf{X}$ and $\mathbf{Y}$, respectively, corresponding to feature j across all token positions. After centering each column over tokens, we define the correlation for dimension j as

$$\rho_j(\mathbf{X}, \mathbf{Y}) = \frac{(\mathbf{X}_{:j})^\top \mathbf{Y}_{:j}}{\|\mathbf{X}_{:j}\| \|\mathbf{Y}_{:j}\|}, \quad j = 1, \ldots, D. \tag{1}$$

We summarize $\{\rho_j\}_{j=1}^D$ via mean. High per-dimension correlations indicate that many features are already aligned one-to-one without any transformation, while low per-dimension correlations suggest that, even if the models encode similar information overall, information carried by a single feature in one model may be distributed across multiple features in the other.

3.2.2 Orthogonal Procrustes Analysis

Dimension-wise correlation is strict: it does not allow any mixing between feature dimensions. Orthogonal Procrustes analysis relaxes this by asking whether one representation space can be mapped to the other via a single orthogonal transformation. Formally, we solve

$$O^* = \arg\min_{O^\top O = I} \|\mathbf{X}O - \mathbf{Y}\|_F^2, \tag{2}$$

where $O^* \in \mathbb{R}^{D \times D}$ is orthogonal and $\|\cdot\|_F$ denotes the Frobenius norm. This objective allows a global orthogonal mixing of features: each feature of $\mathbf{Y}$ can be an orthogonal combination of features in $\mathbf{X}$. If O^* is a near-diagonal or near-permutation, then features can be matched almost one-to-one after a rotation or permutation. This corresponds to relatively localized, interpretable feature correspondences. In contrast, if O^* is dense, then each feature in one model is a distributed combination of many features in the other. The same information may be present, but in a non-localized form. We consider the inverse row entropy of O^* as the quantified metric. See Appendix B.1.2.

3.3 Geometry-Level Analysis

Geometry-level analysis moves one step up in abstraction. Instead of caring about the specific coordinate system, we focus on the *shape* of the latent manifold. At this level, orthogonal rotations and neuron permutations are treated as irrelevant; the relative arrangement of points is paramount.

Geometry-level metrics are therefore invariant to changes of coordinate basis, but sensitive to deformations that alter distances or local neighborhoods. See Appendix B.2. Conceptually, this level investigates *whether two models organize embeddings into similar manifold shapes even using different axes*. We study geometry-level similarity using two metrics:

1. **Linear Centered Kernel Alignment (Linear CKA)**: measures *global geometry* of manifold, via their Gram matrices, up to orthogonal transforms and isotropic scaling (Kornblith et al., 2019).
2. **k-Nearest Neighbors (k-NN) Overlap**: measures *local geometry* of manifold by quantifying the preservation of nearest-neighbor relationships (Lin & Smith, 2019).

3.3.1 LINEAR CKA

CKA compares representations via their kernel matrices rather than their raw coordinates. In Linear CKA, we consider the linear kernel $K_X = \mathbf{X}\mathbf{X}^\top$ and $K_Y = \mathbf{Y}\mathbf{Y}^\top$, where $K_X, K_Y \in \mathbb{R}^{N\times N}$. The Hilbert–Schmidt Independence Criterion (HSIC) (Gretton et al., 2005) between the kernels K_X and K_Y is $\mathrm{HSIC}(K_X, K_Y) = \frac{1}{(N-1)^2}\,\mathrm{tr}(K_X H K_Y H)$, where $H = I - \frac{1}{N}\mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{N\times N}$ is the centering matrix, I is the identity matrix and $\mathbf{1}$ is a vector of 1. Linear CKA is then

$$\mathrm{CKA}(\mathbf{X}, \mathbf{Y}) = \frac{\mathrm{HSIC}(K_X, K_Y)}{\sqrt{\mathrm{HSIC}(K_X, K_X)\,\mathrm{HSIC}(K_Y, K_Y)}}. \tag{3}$$

Linear CKA quantifies how similarly two models organize *global geometry* of manifolds, capturing features like cluster and anisotropy, by assessing whether their manifolds can be aligned through orthogonal transformations and uniform scaling, without the need for nonlinear deformations.

3.3.2 k-NN OVERLAP

While CKA captures global manifold geometry, k-NN overlap focuses on local manifold geometry. Intuitively, it investigates whether each embedding's local neighborhood is preserved between two models. Let the cosine similarity be $s_{\mathbf{Z}}(i,j) = \frac{z_i^\top z_j}{\|z_i\|\,\|z_j\|}$, where $z_i \in \mathbb{R}^D$ is the i-th embedding (row) of the representation matrix $\mathbf{Z}$. We define the k-nearest neighbor sets under models with representations $\mathbf{X}$ and $\mathbf{Y}$ as $N_k^X(i) = \mathrm{TopK}_j\ s_{\mathbf{X}}(i,j)$ and $N_k^Y(i) = \mathrm{TopK}_j\ s_{\mathbf{Y}}(i,j)$, respectively. The k-NN overlap score $\tilde{s}_k$ is

$$\tilde{s}_k = \frac{1}{N}\sum_{i=1}^{N} \frac{\left|N_k^X(i) \cap N_k^Y(i)\right|}{\left|N_k^X(i) \cup N_k^Y(i)\right|} \tag{4}$$

where we use the Jaccard index to quantify agreement of neighbor sets. Because neighborhood relations are preserved under orthogonal transforms and permutations, but disrupted by non-isometric distortions (e.g., anisotropic scaling), k-NN overlap directly reflects how similarly two models instantiate the *local geometry* of the manifold.

3.4 FUNCTION-LEVEL ANALYSIS

Function-level analysis abstracts away from the internal representation manifold to focus strictly on *the input–output transformations exhibited during downstream tasks*. Two models may have different representation- or geometry-level metrics, yet still be functionally similar if the same tasks are solvable with comparable readouts. Conversely, even modest changes in embeddings can yield different behaviors under specific decoders. See Appendix B.3.

We instantiate function-level similarity with the **Cross-Model Linear Probes** (Nikooroo & Engel, 2025) to test whether the *same* linear readout generalizes across models.

3.4.1 CROSS-MODEL LINEAR PROBES

Cross-model linear probes provide a task-conditioned measure of function-level similarity between embedding spaces. Let $y \in \mathbb{R}^N$ be the labels. We first fit a linear probe on $\mathbf{X}$:

$$\hat{y} = \mathbf{X}W_X + b_X \tag{5}$$

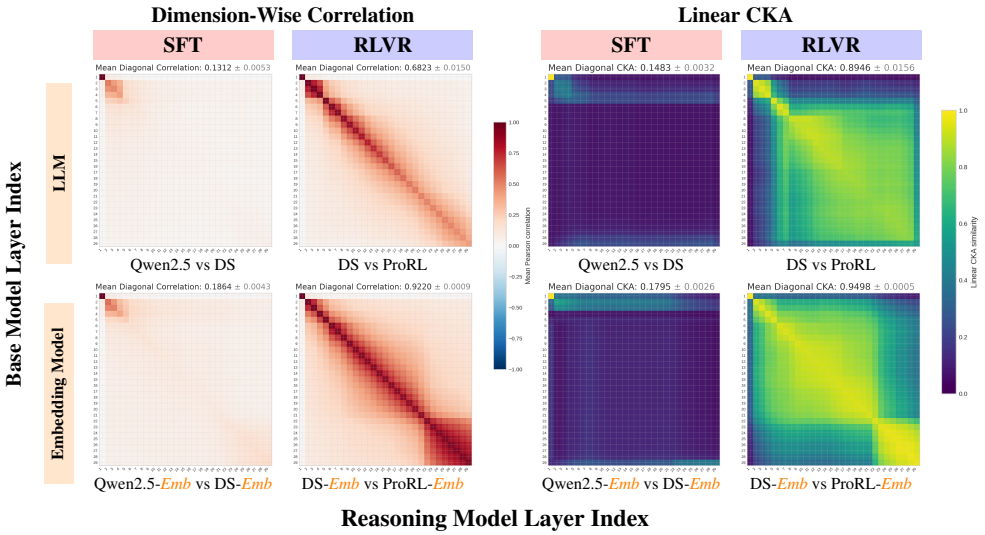


Figure 3: **Heatmaps of Dimension-Wise Correlation (left) and Linear CKA (right). Columns**: SFT versus RLVR. **Rows**: $\mathcal{M}_{base}$ vs. $\mathcal{M}_{reason}$ and $\mathcal{M}^{Emb}_{base}$ vs. $\mathcal{M}^{Emb}_{reason}$. Entries report the mean and gray values denote standard deviation across 3 seeds. SFT disrupts the intrinsic coordinate basis and global geometry while RLVR preserves them. Contrastive learning realigns the similarities.

where (W_X, b_X) are learned via logistic or ridge regression, depending on the task. We then *freeze* (W_X, b_X) and apply the same linear map to representations from the other model. High cross-model performance (probe trained on $\mathbf{X}$, evaluated on $\mathbf{Y}$) relative to self-performance (trained and evaluated on $\mathbf{X}$) indicates that the same linear decision boundary is useful in both spaces. This implies strong function-level similarity: two models support essentially the same set of linearly decodable functions for that task.

3.5 From Representations to Functions: How the Levels Fit Together

HRSA forms a hierarchy of abstraction over the same underlying representations. **Representation level** asks whether the latent manifolds of two models share the same coordinate basis. **Geometry level** discards the choice of coordinate basis and asks whether the latent manifold has a similar shape globally and locally. **Function level** discards most geometric detail and asks which input–output mappings are supported and realized.

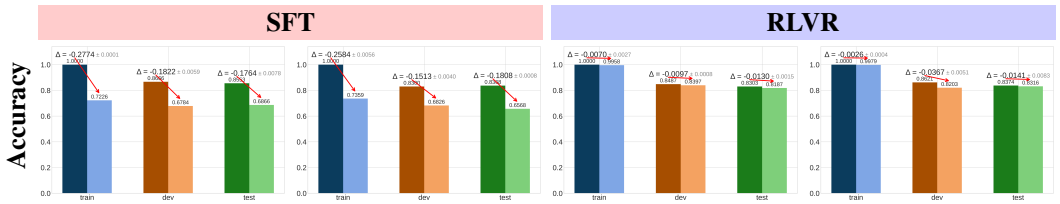
By separating these levels, we can distinguish:

- Cases where $\mathcal{M}_{base}$ and $\mathcal{M}_{reason}$ differ mainly by a reparameterization (e.g., rotation) but preserve geometry and function.
- Cases where global or local geometry changes but downstream behavior remains similar, suggesting redundant internal solutions.
- Cases where modest representational changes induce large functional differences in readout directions, revealing sensitive or brittle aspects of the model's decision rules.

Table 2: Inverse row entropy of O^* for backbone and embedding-model pairs. Gray values denote standard deviation across 3 seeds. Higher values indicate a more one-to-one feature mapping. RLVR preserves this mapping, which contrastive learning further solidifies.

Model Pair	Inverse Row Entropy ↑
SFT: Qwen2.5 vs DS	0.1082 ±0.0006
→ Qwen2.5-*Emb* vs DS-*Emb*	0.1450 ±0.0021
RLVR: DS vs ProRL	**0.1758** ±0.0145
→ DS-*Emb* vs ProRL-*Emb*	**0.8892** ±0.0255

This decomposition allows us to turn the initial puzzle—why $\mathcal{M}^{Emb}_{base}$ and $\mathcal{M}^{Emb}_{reason}$ look so similar on benchmarks—into an investigation of *where* any differences live. Does the root cause of the negligible deviation lie within the underlying backbone models?



(a) Qwen2.5 vs DS (b) Qwen2.5-*Emb* vs DS-*Emb* (c) DS vs ProRL (d) DS-*Emb* vs ProRL-*Emb*

Figure 4: **Cross-Model Linear Probe Results.** For each dataset split (train, dev, test), the left bar corresponds to $\mathcal{M}_{base}$ (or $\mathcal{M}_{base}^{Emb}$) and the right bar corresponds to $\mathcal{M}_{reason}$ (or $\mathcal{M}_{reason}^{Emb}$). Gray values denote standard deviation across 3 seeds. The smaller the Δ, the stronger the cross-model linear probe transfer. RLVR pairs (both LLMs and embedding models) preserve the linear readout direction of the base model.

4 Evaluation Setups

To empirically validate the hypothesis, we apply the HRSA framework across two dimensions: LLMs ($\mathcal{M}_{base}$ vs. $\mathcal{M}_{reason}$) and downstream adaptation ($\mathcal{M}_{base}^{Emb}$ vs. $\mathcal{M}_{reason}^{Emb}$). Furthermore, we extend this analysis by also comparing the SFT-tuned reasoning models, showing the clear differences in SFT-tuned and RLVR-tuned reasoning models. See Appendix D for more experiment results.

To verify if the latent manifold is preserved even within reasoning trajectories, we construct a Chain-of-Thought (CoT) dataset and use the hard-level subset for evaluation. Further generation details are provided in Appendix C. In function-level analysis, we use the AG's News Topic Classification Dataset (Zhang et al., 2015) to evaluate the linear readout directions. For SFT comparison, we use *Qwen2.5-Math-1.5B* (base) (Yang et al., 2024) vs. *DeepSeek-R1-Distill-Qwen-1.5B* (reasoning) (DeepSeek-AI, 2025). For RLVR comparison, we use *DeepSeek-R1-Distill-Qwen-1.5B* (base) vs. *Nemotron-Research-Reasoning-Qwen-1.5B* (Liu et al., 2025b) (reasoning, trained with prolonged RLVR). We abbreviate these as **Qwen2.5**, **DS**, and **ProRL** respectively. Downstream embedding models add "-*Emb*" suffix.[1] We also show the training dynamic of manifold realignment. We use seeds 42, 1016, and 2048 for all the experiments. Our HRSA analysis examines model activations at every layer, before any pooling. Specifically, for each model in a matched pair with L layers, we collect all the hidden states: $\{\mathbf{X}_l\}_{l=1}^{L}$ and $\{\mathbf{Y}_l\}_{l=1}^{L}$, where each $\mathbf{X}_l, \mathbf{Y}_l \in \mathbb{R}^{N\times D}$. This per-layer, per-token perspective preserves the full representational structure for a more comprehensive analysis, avoiding any information loss due to pooling.

5 Results

5.1 Representation-Level Results

Dimension-Wise Correlation Figure 3 (left) shows that SFT yields weak axis-aligned feature correspondence, while RLVR retains substantially higher per-dimension correlations. Notably, the clearest deviation from diagonal structure appears only under *prolonged* RLVR (ProRL), whereas contrastive learning largely restores axis alignment between the resulting embedding models, consistent with *Manifold Realignment*.

Orthogonal Procrustes Analysis Table 2 supports this global alignment perspective. While SFT results in a dense orthogonal map O^* (implying high feature mixing), prolonged RLVR yields an O^* that is nearly a permutation matrix, becoming strongly permutative after contrastive learning. Table 11 shows that O^* is already near-permutation for $\mathcal{M}_{base}$ vs. RLVR-tuned $\mathcal{M}_{reason}$ comparisons. This suggests coordinate basis drift is limited to prolonged training scenarios (as in ProRL).

[1]Additional results for RLVR models with different training algorithms (GRPO (Shao et al., 2024), DAPO (Yu et al., 2025)), model sizes (0.6B-7B), and training datasets are in Appendix D, showing consistent results across variations.

5.2 Geometry-Level Results

Linear CKA Figure 3 (right) shows a sharp contrast in *global manifold geometry*. Linear CKA drops under SFT but remains high under RLVR. After contrastive learning, the $\mathcal{M}_{base}^{Emb}$ and $\mathcal{M}_{reason}^{Emb}$ move even closer in CKA, highlighting *Manifold Realignment* at the geometry level. RLVR functions as a near-isometric transformation, rigidly preserving the shape of the latent manifold. Consequently, the semantic distances established during pre-training remain invariant, which explains why downstream embedding performance does not improve.

k-NN Overlap Table 3 shows that RLVR preserves substantially more *local* geometry (higher mean overlap) than SFT, yet overlap remains substantially below 1, indicating local geometry reorganization. This gap persists even when the embedding model manifolds are pulled closer by contrastive learning, showing the idea that RLVR introduces irreversible local geometry reorganization, which is different from the global geometry.

Table 3: Mean k-NN overlap across layers for backbone and embedding-model pairs. Gray values denote standard deviation across 3 seeds. Higher values indicate greater local-geometry preservation.

Model Pairs	Mean Overlap ↑		
	k=5	k=10	k=50
SFT: Qwen2.5 vs DS	0.052 ±0.019	0.068 ±0.027	0.132 ±0.062
→ Qwen2.5-*Emb* vs DS-*Emb*	0.069 ±0.026	0.068 ±0.027	0.091 ±0.055
RLVR: DS vs ProRL	**0.455** ±0.073	**0.484** ±0.065	**0.577** ±0.042
→ DS-*Emb* vs ProRL-*Emb*	**0.451** ±0.070	**0.474** ±0.069	**0.531** ±0.063

5.3 Function-Level Results

Cross-Model Linear Probes The linear probe is trained on $\mathcal{M}_{base}$ (or $\mathcal{M}_{base}^{Emb}$ in embedding model analysis) representations and evaluated on both models. Figure 4 shows stronger cross-model probe transfer under RLVR than SFT, implying that task-relevant linear readout directions of the latent manifold are more stable. For the embedding model pairs, transfer remains consistently high, reflecting *Manifold Realignment*: contrastive learning maintains strong functional alignment even when local geometry do not fully coincide.

5.4 Manifold Realignment in Training Dynamics

Figure 5 illustrates the dynamics of adapting LLMs to embedding models over training steps in a single training trajectory. By applying HRSA to intermediate checkpoints, we observe that manifold realignment occurs rapidly in the early training stages (Steps 0–200), after which representational similarity stabilizes. This trajectory demonstrates the manifold realignment, where contrastive learning effectively drives strong alignment between base- and reasoning-initialized embedding models. In contrast, the k-NN mean overlap across layers decreases during this process, confirming that the RLVR-induced reorganization of local geometry is irreversible.

6 Discussion

Training objectives dictate latent manifold. We trace the destructive, anisotropic distortion observed during SFT directly to its token-level cross-entropy objective. Essentially, this objective forces a global overhaul of the latent manifold simply to accommodate new offline data. In contrast, RLVR optimizes sequential decisions on-policy. Because it relies on the model's inherent semantic representations to discover valid reasoning paths, this explicit reward structure confines geometric shifts to localized areas. As a result, it effectively clusters advantageous states together while preserving the broader global semantic structure.

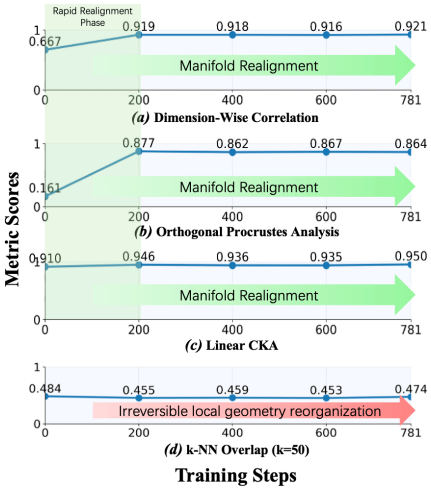


Figure 5: **Training dynamics for DS-*Emb* vs ProRL-*Emb*.** Step 0 and 781 denote the backbones and final embedding models.

Does functional convergence require identical neighborhoods? Gröger et al. (2026) find that cross-modal neighborhood agreement survives null calibration, whereas apparent global spectral convergence largely disappears. This agreement concerns neighbor identities rather than exact distances. Our matched base–RLVR comparisons show a complementary pattern: high global CKA coexists with incomplete k-NN overlap. After contrastive learning, neighborhood differences persist despite strong probe transfer and similar embedding performance. This raises a question: *which local relations matter for function, and which can vary without affecting measured performance?* Our results suggest that local reorganization alone does not reliably improve embedding quality.

Geometry-aware regularization offers a novel training frontier. Our investigation provides an actionable hypothesis for future architecture design. If the distinctive footprint of RLVR is local geometry reorganization beneath global geometric stability, then researchers might achieve similar cognitive leaps through SFT augmented with geometry aware regularization. Explicitly constraining global manifold distances or penalizing excessive coordinate basis drift, while simultaneously encouraging controlled local geometry reorganization, presents a highly promising direction for representation learning.

7 Related Works

Reinforcement Learning with Verifiable Rewards (RLVR) RLVR optimizes models using verifiable rewards rather than heuristic preference signals (DeepSeek-AI, 2025). Recent analyses suggest RLVR stays close to the pretrained solution (e.g., KL-anchored/on-policy behavior) (Shenfeld et al., 2025) and improves via weight updates that avoid large principal-subspace changes (Zhu et al., 2025a), without introducing fundamentally novel reasoning beyond the base model (Yue et al., 2025). However, these works do not directly characterize the *representational* changes induced by RLVR; we show (via HRSA) that RLVR largely preserves global manifold geometry while reorganizing local geometry and, with prolonged training, exhibiting some coordinate basis drift.

Embedding Models Many state-of-the-art text embedding models now leverage decoder-only LLM backbones with bidirectional attention and contrastive learning to produce strong encoders (Lee et al., 2025b; Shanbhogue et al., 2026). While reward-driven or RL-based embedding learning has been explored (Tennenholtz et al., 2024; Gui & Cheng, 2025), it remains unclear whether RLVR-tuned reasoning models improve embedding geometry or retrieval. We introduce the first study that directly tests this connection and finds that RLVR-tuned reasoning models do not reliably enhance embedding quality.

Representational Similarity Analysis Representational similarity analysis and related metrics (e.g., CKA) are widely used to compare layer representations across models and tasks (Kriegeskorte et al., 2008; Kornblith et al., 2019; Yousefi et al., 2023; Gröger et al., 2026). Nevertheless, prior work typically reports single-level alignment and does not organize how changes manifest across abstraction levels, nor does it connect RLVR update properties to representation geometry (Shenfeld et al., 2025; Balashov, 2025). We address this with HRSA, which disentangles coordinate basis, manifold geometry, and readout-direction changes, showing substantial preservation in RLVR-tuned reasoning models.

8 Conclusion

We introduced the HRSA framework to evaluate latent manifold transformations and benchmarked embedding models initialized from reasoning versus base LLMs. Our analysis isolates the precise manifold components that mutate during training, uncovering an empirical pattern of manifold realignment. Standard SFT destructively redraws the latent manifold. RLVR, however, operates as a trajectory planner. It irreversibly alters local manifold geometry yet preserves global manifold geometry and linear readouts. Subsequent contrastive learning rectifies temporary coordinate basis drift and realigns the latent manifold. Ultimately, current reasoning optimizations navigate existing semantic landscapes rather than designing entirely new ones. This has important implications for latent space-centric paradigms like World Models (Ha & Schmidhuber, 2018) and JEPA (Maes et al., 2026), where the quality of the underlying representational structure matters greatly.

AI USE STATEMENT

In this work, we used generative AI tools for generating synthetic data (where the details are disclosed in Appendix C), implementing methods, supporting qualitative and thematic data analysis, interpreting results. We have not used generative AI tools for helping develop theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, and assisting with translation are not applicable to this work. We have reviewed all AI-assisted work, including checking LLM-generated writing for potential plagiarism through a manual literature survey by 3 authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work is an offline study of publicly released language models and datasets made available for academic research and evaluation. It involves no direct interaction with human participants and no collection of new user data. The CoT-Activations dataset consists of model-generated reasoning traces derived from public mathematical benchmarks. We use existing artifacts in accordance with their licenses and intended uses, and will distribute only those research artifacts for which redistribution is permitted, subject to the source datasets' terms and restrictions.

REPRODUCIBILITY STATEMENT

We provide anonymous source code for embedding-model training, HRSA analyses, and evaluation scripts at `https://anonymous.4open.science/r/reasoning-embedding-anonymize-AAE7/`. The contrastive training recipe, hyperparameters, hardware, and training-data composition are specified in Appendix A. Matched backbone pairs, the activation-collection protocol, and evaluation datasets are described in Section 4, with additional model pairs listed in Tables 8 and 9. Definitions of the three HRSA levels and the associated metrics appear in Section 3; invariance assumptions and complete proofs are given in Appendix B. Construction, prompting, and quality-control procedures for the CoT-Activations dataset are documented in Appendix C. Embedding-benchmark results are reported in Section 2, and extended HRSA analyses are provided in Appendix D.

REFERENCES

Chenxin An, Zhihui Xie, Xiaonan Li, Lei Li, Jun Zhang, Shansan Gong, Ming Zhong, Jingjing Xu, Xipeng Qiu, Mingxuan Wang, and Lingpeng Kong. Polaris: A post-training recipe for scaling reinforcement learning on advanced reasoning models, 2025. URL `https://hkunlp.github.io/blog/2025/Polaris`.

Andrii Balashov. Reinforcement learning fine-tunes a sparse subnetwork in large language models. *arXiv preprint arXiv:2507.17107*, 2025.

Anne Beyer, Göran Kauermann, and Hinrich Schütze. Embedding space correlation as a measure of domain similarity. In *LREC*, pp. 2431–2439. European Language Resources Association, 2020.

Adithya Bhaskar, Xi Ye, and Danqi Chen. Language models that think, chat better. *CoRR*, abs/2509.20357, 2025.

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

MohammadReza Davari, Stefan Horoi, Amine Natik, Guillaume Lajoie, Guy Wolf, and Eugene Belilovsky. Reliability of CKA as a similarity measure in deep learning. In *ICLR*. OpenReview.net, 2023.

Gabriel de Souza Pereira Moreira, Radek Osmulski, Mengyao Xu, Ronay Ak, Benedikt Schifferer, and Even Oldridge. Nv-retriever: Improving text embedding models with effective hard-negative mining. *CoRR*, abs/2407.15831, 2024.

DeepSeek-AI. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *CoRR*, abs/2501.12948, 2025.

DeepSeek-AI. Deepseek-v3.2-exp: Boosting long-context efficiency with deepseek sparse attention, 2025.

Kenneth C. Enevoldsen, Isaac Chung, Imene Kerboua, Márton Kardos, Ashwin Mathur, David Stap, Jay Gala, Wissam Siblini, Dominik Krzeminski, Genta Indra Winata, Saba Sturua, Saiteja Utpala, Mathieu Ciancone, Marion Schaeffer, Diganta Misra, Shreeya Dhakal, Jonathan Rystrøm, Roman Solomatin, Ömer Veysel Çagatan, Akash Kundu, and et al. MMTEB: massive multilingual text embedding benchmark. In *ICLR*. OpenReview.net, 2025.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. Simcse: Simple contrastive learning of sentence embeddings. *CoRR*, abs/2104.08821, 2021.

Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *Algorithmic Learning Theory (ALT 2005), 16th International Conference, Proceedings*, volume 3734 of *Lecture Notes in Artificial Intelligence*, pp. 63–77. Springer, 2005.

Fabian Gröger, Shuo Wen, and Maria Brbic. Revisiting the platonic representation hypothesis: An aristotelian view. *CoRR*, abs/2602.14486, 2026.

Yuntao Gui and James Cheng. Search-r3: Unifying reasoning and embedding generation in large language models. *CoRR*, abs/2510.07048, 2025.

David Ha and Jürgen Schmidhuber. World models. *CoRR*, abs/1803.10122, 2018.

Sarah E. Harvey, David Lipshutz, and Alex H. Williams. What representational similarity measures imply about decodable information. In *UniReps*, volume 285 of *Proceedings of Machine Learning Research*, pp. 140–151. PMLR, 2025.

Lucas Hayne, Heejung Jung, and R. McKell Carter. Does representation similarity capture function similarity? *Trans. Mach. Learn. Res.*, 2024, 2024.

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

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey E. Hinton. Similarity of neural network representations revisited. In *ICML*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3519–3529. PMLR, 2019.

Nikolaus Kriegeskorte, Marieke Mur, and Peter A. Bandettini. Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, Volume 2 - 2008, 2008. ISSN 1662-5137. doi: 10.3389/neuro.06.004.2008. URL https://www.frontiersin.org/journals/systems-neuroscience/articles/10.3389/neuro.06.004.2008.

Chankyu Lee, Rajarshi Roy, Mengyao Xu, Jonathan Raiman, Mohammad Shoeybi, Bryan Catanzaro, and Wei Ping. Nv-embed: Improved techniques for training llms as generalist embedding models. In *ICLR*. OpenReview.net, 2025a.

Jinhyuk Lee, Feiyang Chen, Sahil Dua, Daniel Cer, Madhuri Shanbhogue, Iftekhar Naim, Gustavo Hernández Ábrego, Zhe Li, Kaifeng Chen, Henrique Schechter Vera, Xiaoqi Ren, Shanfeng Zhang, Daniel Salz, Michael Boratko, Jay Han, Blair Chen, Shuo Huang, Vikram Rao, Paul Suganthan, Feng Han, Andreas Doumanoglou, Nithi Gupta, Fedor Moiseev, Cathy Yip, Aashi Jain, Simon Baumgartner, Shahrokh Shahi, Frank Palma Gomez, Sandeep Mariserla, Min Choi, Parashar Shah, Sonam Goenka, Ke Chen, Ye Xia, Koert Chen, Sai Meher Karthik Duddu, Yichang Chen, Trevor Walker, Wenlei Zhou, Rakesh Ghiya, Zach Gleicher, Karan Gill, Zhe Dong, Mojtaba Seyedhosseini, Yun-Hsuan Sung, Raphael Hoffmann, and Tom Duerig. Gemini embedding: Generalizable embeddings from gemini. *CoRR*, abs/2503.07891, 2025b.

Lucy H. Lin and Noah A. Smith. Situating sentence embedders with nearest neighbor overlap. *CoRR*, abs/1909.10724, 2019.

Junnan Liu, Hongwei Liu, Linchen Xiao, Ziyi Wang, Kuikun Liu, Songyang Gao, Wenwei Zhang, Songyang Zhang, and Kai Chen. Are your llms capable of stable reasoning? In *ACL (Findings)*, pp. 17594–17632. Association for Computational Linguistics, 2025a.

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. *CoRR*, abs/2505.24864, 2025b.

Lucas Maes, Quentin Le Lidec, Damien Scieur, Yann LeCun, and Randall Balestriero. Leworld-model: Stable end-to-end joint-embedding predictive architecture from pixels. *CoRR*, abs/2603.19312, 2026.

Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In *ICLR (Workshop Poster)*, 2013.

Niklas Muennighoff, Nouamane Tazi, Loïc Magne, and Nils Reimers. MTEB: massive text embedding benchmark. In *EACL*, pp. 2006–2029. Association for Computational Linguistics, 2023.

Niklas Muennighoff, Hongjin Su, Liang Wang, Nan Yang, Furu Wei, Tao Yu, Amanpreet Singh, and Douwe Kiela. Generative representational instruction tuning. *CoRR*, abs/2402.09906, 2024.

Yixin Nie, Adina Williams, Emily Dinan, Mohit Bansal, Jason Weston, and Douwe Kiela. Adversarial nli: A new benchmark for natural language understanding. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2020.

Saleh Nikooroo and Thomas Engel. Cross-model semantics in representation learning. *CoRR*, abs/2508.03649, 2025.

Peter H Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *CoRR*, abs/1707.06347, 2017.

Madhuri Shanbhogue, Zhe Li, Shanfeng Zhang, Gustavo Hernández Ábrego, Shih-Cheng Huang, Aashi Jain, Daniel Salz, Sonam Goenka, Chaitra Hegde, Ji Ma, Feiyang Chen, Jiaxing Wu, Tanmaya Dabral, Babak Samari, Kevin Poulet, Daniel Cer, Kaifeng Chen, Paul Suganathan, Hui Hui, Jovan Andonov, Philippe Schlattner, Jay Han, Iftekhar Naim, Wing Lowe, Vladimir Pchelin, Albert Chih-Chieh Yang, Yi-Ting Chen, Zhongli Ding, Grace Zhang, Georg Heigold, Yichang Chen, Antoine Reveillon, Brendan Mccloskey, Wenlei Zhou, Dahun Kim, Rui Meng, Emma Wang, Jack Zheng, Halley Fede, Zhen Yang, Keegan Mosley, Brian Potetz, Sahil Dua, Henrique Schechter Vera, Shen Gao, Hesen Zhang, Andreas Hess, Hengxuan Ying, Alberto Montes, Karan Gill, Min Choi, Sebastian Russo, Anja Hauth, Jinhyuk Lee, Michael Boratko, Megan Barnes, Vikram Rao, Claudiu Musat, Cyril Allauzen, Ehsan Variani, Shankar Kumar, Tom Bagby, Junyi Jiao, Yang Gu, Tengxin Li, Ayush Agrawal, Roberto Santana, Dev Nath, Stephen Karukas, Shuoxuan Han, Lucia Loher, Alice Twu, Nidhi Vyas, Siddharth Bhai, Frank Palma Gomez, Wangyuan Zhang, Chaoren Liu, Jizheng Yang, Steve Qiu, Shijie Zhang, Sujay Kulkarni, Sascha Rothe, Sean Nakamoto, Raphael Hoffmann, Zach Gleicher, Yun-Hsuan Sung, Qin Yin, Tom Duerig, and Mojtaba Seyedhosseini. Gemini embedding 2: A native multimodal embedding model from gemini. *CoRR*, abs/2605.27295, 2026.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *CoRR*, abs/2402.03300, 2024.

Idan Shenfeld, Jyothish Pari, and Pulkit Agrawal. Rl's razor: Why online reinforcement learning forgets less. *CoRR*, abs/2509.04259, 2025.

Hongjin Su, Howard Yen, Mengzhou Xia, Weijia Shi, Niklas Muennighoff, Han-yu Wang, Haisu Liu, Quan Shi, Zachary S. Siegel, Michael Tang, Ruoxi Sun, Jinsung Yoon, Sercan Ö. Arik, Danqi Chen, and Tao Yu. BRIGHT: A realistic and challenging benchmark for reasoning-intensive retrieval. In *ICLR*. OpenReview.net, 2025.

Guy Tennenholtz, Yinlam Chow, Chih-Wei Hsu, Lior Shani, Ethan Liang, and Craig Boutilier. Embedding-aligned language models. *CoRR*, abs/2406.00024, 2024.

Aäron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *CoRR*, abs/1807.03748, 2018.

Yubo Wang, Xueguang Ma, Ge Zhang, Yuansheng Ni, Abhranil Chandra, Shiguang Guo, Weiming Ren, Aaran Arulraj, Xuan He, Ziyan Jiang, Tianle Li, Max Ku, Kai Wang, Alex Zhuang, Rongqi Fan, Xiang Yue, and Wenhu Chen. Mmlu-pro: A more robust and challenging multi-task language understanding benchmark. In *NeurIPS*, 2024.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 technical report. *CoRR*, abs/2412.15115, 2024.

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, Jian Yang, Jianhong Tu, Jianwei Zhang, Jian Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report. *CoRR*, abs/2505.09388, 2025.

Safoora Yousefi, Leo Betthauser, Hosein Hasanbeig, Raphaël Millière, and Ida Momennejad. Decoding in-context learning: Neuroscience-inspired analysis of representations in large language models. *arXiv preprint arXiv:2310.00313*, 2023.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Tiantian Fan, Gaohong Liu, Lingjun Liu, Xin Liu, Haibin Lin, Zhiqi Lin, Bole Ma, Guangming Sheng, Yuxuan Tong, Chi Zhang, Mofan Zhang, Wang Zhang, Hang Zhu, Jinhua Zhu, Jiaze Chen, Jiangjie Chen, Chengyi Wang, Hongli Yu, Weinan Dai, Yuxuan Song, Xiangpeng Wei, Hao Zhou, Jingjing Liu, Wei-Ying Ma, Ya-Qin Zhang, Lin Yan, Mu Qiao, Yonghui Wu, and Mingxuan Wang. DAPO: an open-source LLM reinforcement learning system at scale. *CoRR*, abs/2503.14476, 2025.

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.

Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification, 2015.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, Fei Huang, and Jingren Zhou. Qwen3 embedding: Advancing text embedding and reranking through foundation models. *CoRR*, abs/2506.05176, 2025.

Hanqing Zhu, Zhenyu Zhang, Hanxian Huang, DiJia Su, Zechun Liu, Jiawei Zhao, Igor Fedorov, Hamed Pirsiavash, Zhizhou Sha, Jinwon Lee, et al. The path not taken: Rlvr provably learns off the principals. *arXiv preprint arXiv:2511.08567*, 2025a.

Xinyu Zhu, Mengzhou Xia, Zhepei Wei, Wei-Lin Chen, Danqi Chen, and Yu Meng. The surprising effectiveness of negative reinforcement in LLM reasoning. *CoRR*, abs/2506.01347, 2025b.

Appendix Contents

A Embedding Model Training **14**

A.1 Training Details 14

A.2 Training Data Statistics 15

B HRSA Proof **15**

B.1 Representation-Level Proof 16

B.1.1 Dimension-Wise Correlation 16

B.1.2 Orthogonal Procrustes Analysis 17

B.2 Geometry-Level Proof 17

B.2.1 Linear CKA 18

B.2.2 k-NN Overlap 19

B.3 Function-Level Proof 19

B.3.1 Cross-Model Linear Probes 20

C CoT Datasets **20**

C.1 Dataset Composition and Hierarchy 21

C.2 Generation Protocol 21

C.3 Quality Control and Evaluation 21

D Additional Results **22**

A Embedding Model Training

In this section, we reveal all the training details of the embedding models.

A.1 Training Details

We optimize the InfoNCE loss (van den Oord et al., 2018) defined in Equation 6. This objective aims to maximize the similarity between the query q and the positive passage p, while simultaneously minimizing the similarity between q and the negative passages. Let B denote the set of in-batch passages (which includes p and negatives from other instances), $\mathcal{N}$ be the set of hard negatives, and sim be the cosine similarity. The loss is calculated as:

$$\mathcal{L}(q, p, B, \mathcal{N}) = -\log \frac{\exp(\text{sim}(q, p)/\tau)}{\sum_{d \in B \cup \mathcal{N}} \exp(\text{sim}(q, d)/\tau)} \tag{6}$$

We select decoder-only LLMs as the embedding model backbone, take the last layer's activation as the final output, and perform mean pooling to obtain a fixed-dimension embedding vector. We also enable bi-directional attention in the backbone by discarding the causal attention mask to capture more semantic details and relationships between tokens. We use mixed precision with `bfloat16` and gradient checkpointing to reduce the memory pressure on the hardware. We use Flash Attention 2 as the attention backend algorithm. For more details on the settings, the reader can refer to Table 4. We employ the instruction-tuning technique. In particular, we use the instruction template `Instruction: {instruction}\nQuery: {query}`, where `{instruction}` and `{query}` are the placeholders for the instruction and query, respectively. All of our training is conducted on 4x Nvidia L20 GPUs, with VRAM 44GB per GPU.

Table 4: Training Hyperparameters

Variables	Values
Batch Size	2048
Learning Rate (LR)	2×10^{-5}
LR Warm-up Ratio	0.03
LR Scheduler	Cosine
Weight Decay	0.05
Optimizer	AdamW
Padding Side	Right
Number of data	1,603,172
Number of training steps	782
Number of hard negatives	3
Temperature	0.02
Pooling	Mean

Although many prior works (Zhang et al., 2025; Lee et al., 2025b;a) use LoRA to train the embedding models, in our work, we discard it, since we find that training without LoRA yields better performance, and full parameters can better record the training dynamics. See Table 5.

Table 5: LoRA comparison on performance in MTEB (Multilingual, v2).

Model	Performance
With LoRA	
DS-Distill-Qwen-1.5B-*Emb*	42.450
NV-ProRL-*Emb*	42.064
Without LoRA	
DS-Distill-Qwen-1.5B-*Emb*	**46.185**
NV-ProRL-*Emb*	**46.247**

A.2 Training Data Statistics

We consider a wide range of datasets, forming the training dataset by composing 11 separate datasets. We used *Qwen3-Embedding-0.6B* (Zhang et al., 2025) to mine 3 hard negatives per query, and employ the positive-aware hard negative mining technique introduced in de Souza Pereira Moreira et al. (2024), with 95% margin to the positive score.

We utilized publicly available datasets which are explicitly licensed for public academic research and evaluation. Our use of these benchmarks strictly adheres to their original terms of use and intended distribution constraints.

B HRSA Proof

In this section, we provide more details on HRSA, including all the invariance properties of each level analysis and the proof of their invariance properties.

We emphasize again that HRSA is not dependent on the specific metrics selected for this study, such as Dimension-Wise Correlation or Linear CKA. Rather, it is grounded in the hierarchy of invariance properties established earlier. Consequently, any metric that satisfies the invariance requirements of a specific level can be employed to analyze that level's focus. Refer to Table 7 for a summary of these properties and a catalog of alternative valid metrics.

Table 6: Training Datasets Details

Datasets	Number of Samples
FEVER	105,893
NaturalQuestions	97,912
NLI	277,217
MSMARCO	499,184
Quora	94,443
Mr.Tydi	102,796
DUReader	17,493
TriviaQA	65,465
HotpotQA	167,808
SQuAD	84,494
T2Ranking	90,467
Total	1,603,172

Table 7: **HRSA Framework Extensibility.** The HRSA framework is defined by invariance properties, not specific metrics. Researchers can select alternative metrics (right column) for different modalities or theoretical needs, provided they respect the invariance constraints of the target analysis level.

Level	Invariance Constraints	Default Metric	Alternative Valid Metrics
Representation	**Non-Invariant to:** Orthogonal Transformation	Dimension-Wise Correlation	**Optimal Transport (Wasserstein)** Measures cost to move mass from basis X to Y without rotation.
	Goal: Assess alignment of specific axes.	Orthogonal Procrustes (O^*)	**Manifold Alignment Loss** Direct penalization of feature mismatch.
Geometry	**Invariant to:** Similarity Transforms	Linear CKA	**RBF Kernel CKA** Captures non-linear similarity.
	Non-Invariant to: Invertible Linear Transforms (Scaling/Shear)	k-NN Overlap (Jaccard)	**Riemannian Metrics** Geodesic distance comparison.
Function	**Invariant to:** Any transform preserving the decision boundary.	Linear Probing Transfer	**Mutual Information** $I(X;Y)$ Information theoretic upper bound.
		Zero-Shot Accuracy	**Behavioral Consistency** Exact match on downstream tasks.

B.1 Representation-Level Proof

Definition 1. The representation-level analysis examines the explicit coordinate basis of the latent manifold. A metric at this level must demonstrate sensitivity to coordinate basis rotations. Specifically:

- **Non-invariant to:** Orthogonal transformations (Rotation/Permutation) and General Linear transformations.

B.1.1 Dimension-Wise Correlation

Recall the definition of Dimension-Wise Correlation from equation 1.

Proposition 1. *Dimension-Wise Correlation is non-invariant to orthogonal transformations.*

Proof. Let $Q \in \mathbb{R}^{D\times D}$ be an orthogonal matrix ($Q^\top Q = I$) such that $X' = XQ$. The j-th column becomes $x'_{:j} = \sum_{k=1}^{D} X_{:k} Q_{kj}$. The correlation of the j-th column becomes:

$$\rho_j(XQ, Y) = \frac{(\sum_k X_{:k} Q_{kj})^\top Y_{:j}}{\|\sum_k X_{:k} Q_{kj}\|_2 \|Y_{:j}\|_2}. \tag{7}$$

Since Q mixes information from multiple columns $X_{:k}$ into the new column $x'_{:j}$, the correlation with the fixed target $Y_{:j}$ changes arbitrarily depending on Q. Thus, $\rho_j(XQ, Y) \neq \rho_j(X, Y)$, satisfying the requirement for coordinate basis sensitivity. □

B.1.2 Orthogonal Procrustes Analysis

Recall the Orthogonal Procrustes solution O^* defined in Equation 2. To quantify the extent of coordinate alignment, we introduce the *inverse row entropy*, denoted as H_{inv}. We interpret the squared elements of each row in O^* as a probability distribution. This is mathematically valid because O^* is orthogonal, meaning its rows have unit Euclidean norm (i.e., $\sum_j (O^*_{ij})^2 = 1$).

We compute H_{inv} by calculating the mean row entropy, normalizing it by the maximum possible entropy ($\log D$), and taking the complement:

$$H = -\frac{1}{D \log D} \sum_{i=1}^{D} \sum_{j=1}^{D} (O^*_{ij})^2 \log (O^*_{ij})^2$$
$$H_{\text{inv}} = 1 - H$$

where O^*_{ij} denotes the element of O^* at row i and column j, and D represents the dimensionality. The intermediate term H is normalized to the range $[0, 1]$. Consequently, a higher H_{inv} indicates that the coordinate basis is preserved (i.e., O^* is sparse and approximates a permutation matrix), whereas a lower H_{inv} indicates that features are "smeared" or rotated across multiple dimensions.

Proposition 2. *The structure of the optimal mapping in Orthogonal Procrustes Analysis is non-invariant to orthogonal transformations.*

Proof. For any orthogonal matrices $Q, R \in \mathbb{R}^{D\times D}$, if we transform X, Y to $X' = XQ$, $Y' = YR$, then an optimal map for the new problem is

$$O^*(X', Y') = Q^\top O^*(X, Y) R. \tag{8}$$

This is a conjugation of O^* by orthogonal matrices, which in general destroys diagonality or one-hot structure. □

Remark. One may claim that Orthogonal Procrustes Analysis should be classified as a geometry-level measurement because the residual is invariant to orthogonal transformation. In our work, we only focus on the structure of O^*, specifically by considering its inverse row entropy. As shown in Proposition 2, O^* remains dependent on the chosen coordinate system.

B.2 Geometry-Level Proof

Definition 2. The geometry-level analysis examines the intrinsic shape and topology of the latent manifold $\mathcal{Z}$. Metrics at this level must quantify the arrangement of points relative to one another, independent of the specific coordinate system used to describe them.

- **Invariant to:** Similarity transformations, defined as the composition of orthogonal rotation/reflection ($Q \in \mathbb{R}^{D\times D}, Q^\top Q = I$) and isotropic scaling ($c \in \mathbb{R}, c > 0$).
- **Non-invariant to:** Anisotropic linear transformations (e.g., non-uniform scaling, shearing) where the transformation matrix A satisfies $A^\top A \neq cI$.

In the following, we provide proofs for the invariance properties of Linear CKA and k-NN Overlap.

B.2.1 LINEAR CKA

Recall that Linear CKA is defined via the Hilbert–Schmidt Independence Criterion (HSIC) of centered Gram matrices. Let $K_X = XX^\top$ and $H = I - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$.

Proposition 3. *Linear CKA is invariant to similarity transformations $X \mapsto cXQ$ where $c > 0$ and Q is orthogonal.*

Proof. Let $X' = cXQ$. We first derive the Gram matrix for the transformed representation:

$$K_{X'} = (cXQ)(cXQ)^\top = c^2 XQQ^\top X^\top. \tag{9}$$

Since Q is orthogonal ($QQ^\top = I$), this simplifies to:

$$K_{X'} = c^2 XX^\top = c^2 K_X. \tag{10}$$

Now we examine the HSIC term in the numerator. Using the property $\mathrm{tr}(cA) = c\,\mathrm{tr}(A)$:

$$\begin{aligned} \mathrm{HSIC}(K_{X'}, K_Y) &= \frac{1}{(N-1)^2}\mathrm{tr}(K_{X'} H K_Y H) \\ &= \frac{1}{(N-1)^2}\mathrm{tr}(c^2 K_X H K_Y H) \\ &= c^2 \cdot \mathrm{HSIC}(K_X, K_Y). \end{aligned} \tag{11}$$

Similarly, for the normalization term in the denominator:

$$\begin{aligned} \mathrm{HSIC}(K_{X'}, K_{X'}) &= \frac{1}{(N-1)^2}\mathrm{tr}(c^2 K_X H c^2 K_X H) \\ &= c^4 \cdot \mathrm{HSIC}(K_X, K_X). \end{aligned} \tag{12}$$

Substituting these into the full Linear CKA equation:

$$\begin{aligned} \mathrm{CKA}(X', Y) &= \frac{c^2 \mathrm{HSIC}(K_X, K_Y)}{\sqrt{c^4 \mathrm{HSIC}(K_X, K_X) \cdot \mathrm{HSIC}(K_Y, K_Y)}} \\ &= \frac{c^2 \mathrm{HSIC}(K_X, K_Y)}{c^2 \sqrt{\mathrm{HSIC}(K_X, K_X) \cdot \mathrm{HSIC}(K_Y, K_Y)}} \\ &= \mathrm{CKA}(X, Y). \end{aligned} \tag{13}$$

The scalar factors cancel perfectly, proving invariance. □

Proposition 4. *Linear CKA is generally non-invariant to anisotropic linear transformations.*

Proof. Let $X' = XA$, where $A \in \mathbb{R}^{D \times D}$ is invertible and anisotropic ($AA^\top \neq cI$). The Gram matrix becomes:

$$K_{X'} = XAA^\top X^\top. \tag{14}$$

Let $M = AA^\top$. The numerator HSIC term becomes proportional to $\mathrm{tr}(XMX^\top H K_Y H)$. Unlike the isotropic case, the matrix M is "trapped" between X and $X^\top$ inside the trace. Unless M is a scalar multiple of the identity, it reweights the singular values of X, effectively altering the principal components of the representation space. Since Linear CKA measures the alignment of these principal components, $\mathrm{CKA}(XA, Y) \neq \mathrm{CKA}(X, Y)$. □

Remark. Harvey et al. (2025) has shown that Linear CKA quantifies the average alignment between optimal linear readouts across a distribution of decoding tasks. However, it can be manipulated without large changes in functional behavior under high dimensions (Davari et al., 2023; Hayne et al., 2024), so it should not be interpreted as a direct proxy for linear separability or task equivalence. Here, we use Linear CKA as a global geometry descriptor.

B.2.2 k-NN OVERLAP

As defined in the Section 3.3.2, k-NN overlap relies on the ranking of cosine similarities $s(u, v) = \frac{u^\top v}{\|u\|\|v\|}$.

Proposition 5. *Cosine-based k-NN Overlap is invariant to similarity transformations.*

Proof. Let x and y be any two embedding vectors (rows of X and Y, respectively). We apply the transformation $x' = cQx$ and $y' = cQy$, with $c > 0$ and $Q^\top Q = I$. The cosine similarity between the transformed vectors is:

$$\begin{aligned} s(x', y') &= \frac{(cQx)^\top (cQy)}{\|cQx\|\|cQy\|} \\ &= \frac{c^2 x^\top Q^\top Q y}{\sqrt{(cQx)^\top (cQx)}\sqrt{(cQy)^\top (cQy)}}. \end{aligned} \tag{15}$$

Using the orthogonality property $Q^\top Q = I$:

$$\begin{aligned} s(x', y') &= \frac{c^2 x^\top y}{\sqrt{c^2 x^\top x}\sqrt{c^2 y^\top y}} \\ &= \frac{c^2 (x^\top y)}{c\|x\| \cdot c\|y\|} \\ &= \frac{x^\top y}{\|x\|\|y\|} = s(x, y). \end{aligned} \tag{16}$$

Since the pairwise similarity scores remain exactly the same, the ranking of neighbors is preserved. Thus, the set of top-k nearest neighbors is identical: $N_k^{X'}(i) = N_k^X(i)$, and the overlap score is invariant. □

Proposition 6. *Cosine-based k-NN Overlap is generally non-invariant to anisotropic linear transformations.*

Proof. Let $x' = Ax$ and $y' = Ay$ with anisotropic A. The transformed similarity is:

$$s(x', y') = \frac{x^\top A^\top A y}{\sqrt{x^\top A^\top A x}\sqrt{y^\top A^\top A y}}. \tag{17}$$

Let $M = A^\top A$. This expression represents the cosine of the angle between x and y in a space equipped with the inner product $\langle u, v \rangle_M = u^\top M v$.

Because A is anisotropic, M has distinct eigenvalues. This transformation distorts angles: vectors aligned with the large eigenvectors of M are "pulled" closer together in angular space, while vectors aligned with small eigenvectors are pushed apart.

Consequently, if we have $s(x, y) > s(x, z)$ (meaning y is a closer neighbor to x than z), an anisotropic A can reverse this relationship such that $s(x', z') > s(x', y')$. This alters the composition of the k-nearest neighbor sets, changing the overlap score. □

B.3 FUNCTION-LEVEL PROOF

Definition 3. The function-level analysis examines the usable information accessible via linear readouts (probes) or the final behavioral output. This level specifically tests whether two models share the same "readout directions" for solving a task.

- **Invariant to:** Isomorphic transformations *if and only if* the readout mechanism is transformed correspondingly.
- **Non-invariant to:** Linear Reparameterization under a *fixed* readout hypothesis.

Table 8: LLM pairs used in additional HRSA analyses, separated by the training algorithms (SFT, RLVR).

Base Model $\mathcal{M}_{base}$	Reasoning Model $\mathcal{M}_{reason}$	Algorithm	Data
SFT			
Qwen2.5-Math-1.5B	DeepSeek-R1-Distill-Qwen-1.5B	SFT	Mixed
RLVR			
Qwen3-4B	Polaris-4B-Preview	DAPO	Math
DeepSeek-R1-Distill-Qwen-7B	Polaris-7B-Preview	DAPO	Math
Qwen2.5-7B	zero__ppo__think__Qwen2.5-7B	PPO	Chat
Qwen2.5-1.5B	Qwen-2.5-1.5B-SimpleRL-Zoo	GRPO	Math
Qwen2.5-0.5B	Qwen-2.5-0.5B-SimpleRL-Zoo	GRPO	Math
DeepSeek-R1-Distill-Qwen-1.5B	Nemotron-Research-Reasoning-Qwen-1.5B	GRPO	Math
Qwen3-4B	Qwen3-4B-PSR	PSR	Math

Table 9: Embedding model pairs used in additional HRSA analyses. All of the embedding models are trained on the same dataset with InfoNCE loss. They are separated by the training algorithms used to train their reasoning model backbone.

Base Embedding Model $\mathcal{M}_{base}^{Emb}$	Reasoning Embedding Model $\mathcal{M}_{reason}^{Emb}$
SFT	
Qwen2.5-Math-1.5B-*Emb*	DeepSeek-R1-Distill-Qwen-1.5B-*Emb*
Qwen3-0.6B-Base-*Emb*	Qwen3-0.6B-*Emb*
RLVR	
Qwen2.5-1.5B-*Emb*	Qwen-2.5-1.5B-SimpleRL-Zoo-*Emb*
Qwen2.5-0.5B-*Emb*	Qwen-2.5-0.5B-SimpleRL-Zoo-*Emb*
DeepSeek-R1-Distill-Qwen-1.5B-*Emb*	Nemotron-Research-Reasoning-Qwen-1.5B-*Emb*
Qwen3-4B-*Emb*	Qwen3-4B-PSR-*Emb*

B.3.1 Cross-Model Linear Probes

Let w_X^* be the optimal probe weights for task Z on representations X, i.e., $w_X^* = \text{argmin}_w \|Xw - Z\|^2$. We evaluate these weights on Y: Error $= \|Yw_X^* - Z\|^2$.

Proposition 7. *Cross-Model Linear Probes are non-invariant to linear reparameterization under a fixed readout.*

Proof. Assume Y contains the exact same information as X but is linearly transformed: $Y = XA$ (where A is invertible). The prediction using transferred weights is:

$$\hat{Z}_Y = Yw_X^* = (XA)w_X^*. \tag{18}$$

The original prediction was $\hat{Z}_X = Xw_X^*$. For the predictions to be identical ($\hat{Z}_Y = \hat{Z}_X$) for all X, we require $XAw_X^* = Xw_X^*$, implying $Aw_X^* = w_X^*$. This equality only holds if w_X^* is an eigenvector of A with eigenvalue 1. For a general transformation A, $Aw_X^* \neq w_X^*$. Therefore, even if Y is geometrically isomorphic to X, the cross-model probe will fail if the *direction* of the solution has shifted. This proves the metric satisfies the requirement set in Definition 3. □

C CoT Datasets

To rigorously verify if the latent manifold is preserved within reasoning trajectories (as discussed in Section 4), we constructed a specialized **CoT-Activations dataset**. Unlike standard semantic datasets, this corpus focuses on long-range, multi-step reasoning traces generated by state-of-the-art reasoning models.

C.1 Dataset Composition and Hierarchy

We curated a diverse suite of mathematical reasoning benchmarks to ensure our analysis covers varying degrees of reasoning complexity, ranging from elementary arithmetic to competition-level problem solving. The dataset is stratified into three difficulty levels:

- **Easy:** Sourced from **GSM8K** (Cobbe et al., 2021), focusing on grade-school math word problems that require multi-step arithmetic but limited abstract reasoning.
- **Moderate:** Sourced from **MATH-500** (Hendrycks et al., 2021) (a curated subset of the MATH benchmark including AMC/AIME problems) and **NuminaMath** (CN K-12 curriculum). These datasets introduce higher-dimensional algebraic and geometric reasoning.
- **Hard:** Sourced from **LiveMathBench** (Liu et al., 2025a) (2025 Hard Subset). These are recent competition-level problems requiring extremely long context windows and complex logical deductions.

Table 10 summarizes the statistics of the generated CoT dataset. We have verified that the raw data source allows for the collection and derivative redistribution of this data for non-commercial research purposes in accordance with their Terms of Service.

C.2 Generation Protocol

To extract high-quality reasoning traces, we utilized *Qwen3-32B* (Yang et al., 2025) as the generator backbone. The generation process was designed to maximize the explicitness of the internal reasoning process (the "chain of thought").

Inference Configuration. We enabled the internal "thinking" mode (`enable_thinking: true`) to expose the raw reasoning tokens before the final answer. The generation parameters were set to temperature $T = 0.6$ and nucleus sampling probability $p = 0.95$ to balance creativity with logical coherence.

Token Limits. To accommodate deep reasoning, we set a high context limit. For standard datasets, we allowed up to 8,000 reasoning tokens. For the **LiveMathBench** subset, we removed the CoT token limit entirely to allow for exhaustive search trajectories in hard problems.

Prompting. We employed a standardized two-message chat format to enforce rigorous step-by-step reasoning. The **System Prompt** was defined as:

System Prompt of the CoT dataset generation

You are a helpful and rigorous math reasoning assistant.

The **User Prompt** wrapped the specific dataset problem with instructions to act as a competition solver:

User Prompt of the CoT dataset generation

You are an expert competition math solver. Read the problem carefully and solve it step by step.

Problem:
{Problem}

C.3 Quality Control and Evaluation

To ensure that our latent manifold analysis is based on valid reasoning trajectories rather than hallucinations, we implemented a strict verification pipeline using an LLM-as-a-Judge approach.

We employed **DeepSeek-V3.2-exp** (DeepSeek-AI, 2025) as the external evaluator. To ensure deterministic and strictly formatted outputs, we used greedy decoding ($T = 0$) and explicitly disabled the model's internal chain-of-thought feature (`enable_thinking: False`).

The interaction was structured as follows:

System Prompt of the external evaluator

You are a precise math answer evaluator. Respond only with 0 or 1.

User Prompt of the external evaluator

You are an expert math problem evaluator. Your task is to determine if the provided answer correctly solves the given problem.

Problem: {Problem}
Answer: {Answer}

Evaluate whether the answer is correct. Respond with ONLY "1" if the answer is correct, or "0" if it is incorrect. Do not provide any explanation.

The evaluator's output was parsed using a simple inclusion check: if the token "1" appeared in the response, the reasoning trace was marked as valid (`correctness_label` $= 1$); otherwise, it was discarded.

Table 10: **Statistics of the Generated CoT Reasoning Dataset.** We report the yield of our generation pipeline across difficulty tiers. **Total**: Number of initial prompts; **Valid**: Traces that passed the correctness verification (Correctness $= 1$). **Acc.**: The effective yield rate (Valid / Total).

	Generation Statistics		
Dataset (Difficulty)	**Total**	**Valid**	**Acc. (%)**
GSM8K *(Easy)*	500	471	94.20
NuminaMath *(Moderate)*	161	154	95.65
MATH-500 *(Moderate)*	479	364	75.99
LiveMathBench *(Hard)*	57	32	56.14
Total / Average	**1,197**	**1,021**	**85.30**

D Additional Results

In this section, we demonstrate additional results of HRSA applying on more model pairs, including $\mathcal{M}_{base}$ vs. $\mathcal{M}_{reason}$ and $\mathcal{M}_{base}^{Emb}$ vs. $\mathcal{M}_{reason}^{Emb}$ (suffix $-Emb$). See Table 8 and Table 9 for the detailed model pairs.

Instead of only considering the CoT dataset, we also apply HRSA with the MMLU-Pro (Wang et al., 2024) dataset to study the difference (if any) between the models in a general field rather than only the maths domain. For Cross-Model Linear Probe, we additionally evaluate it on the Adversarial Natural Language Inference (ANLI) (Nie et al., 2020), where we particularly focus on Round 1 (ANLI R1) to consider the reasoning of natural language inference (NLI) tasks.

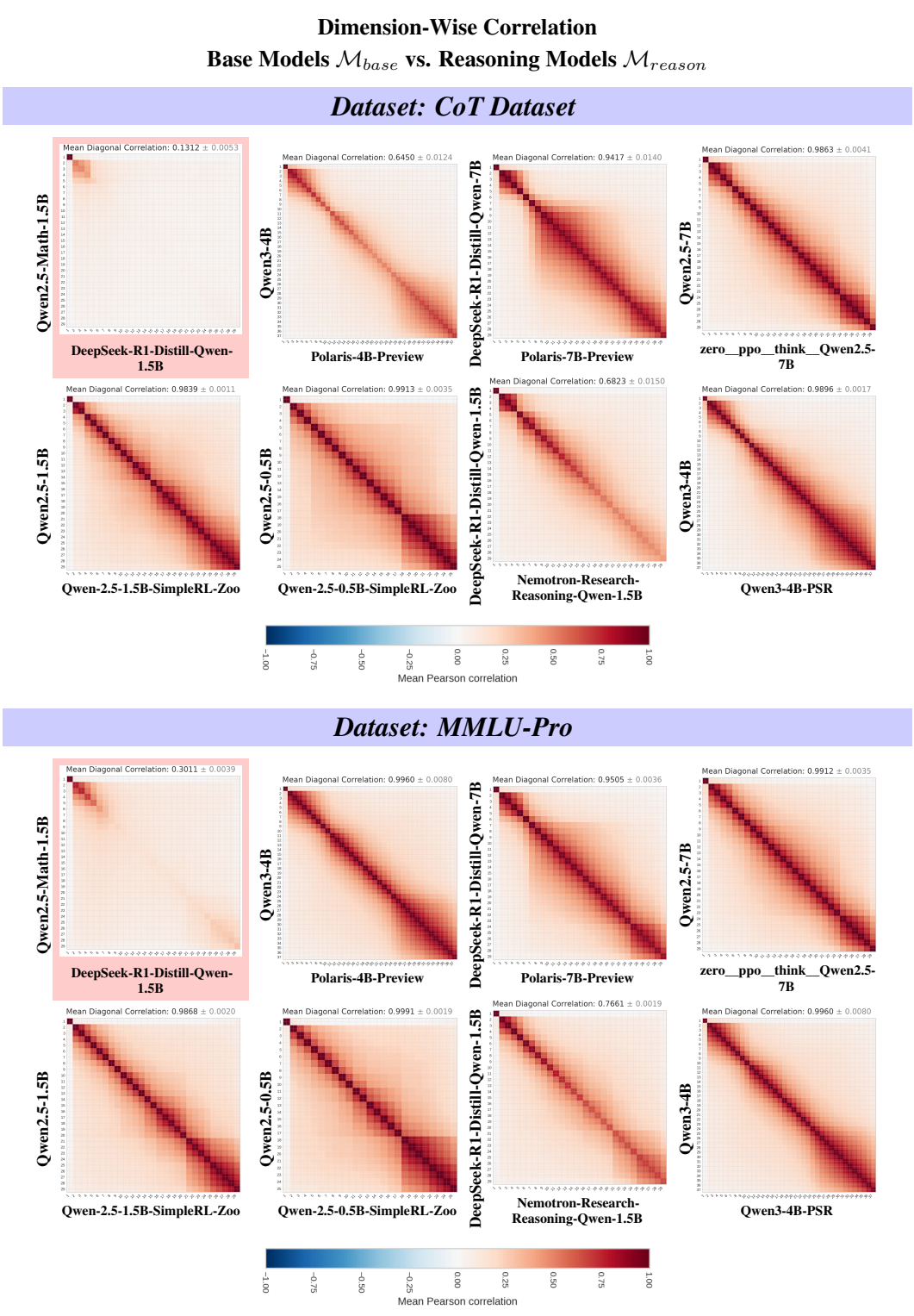
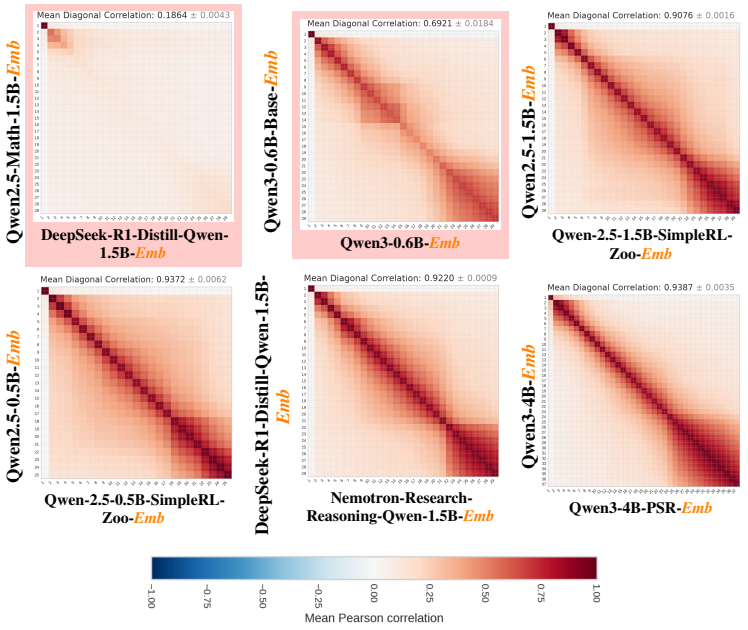


Figure 6: Additional Results on Dimension-Wise Correlation separated by dataset. The vertical axis and horizontal axis are Base Model Layer Index and Reasoning Model Layer Index, respectively. Gray values denote standard deviation across 3 seeds. The **red** background indicates SFT-tuned pairs. Across both datasets, RLVR pairs generally retain substantially stronger axis-aligned layer correspondence than SFT pairs, although the degree of alignment varies across model pairs and datasets.

Dimension-Wise Correlation

Base Embedding Models $\mathcal{M}_{base}^{Emb}$ vs. Reasoning Embedding Models $\mathcal{M}_{reason}^{Emb}$

Dataset: CoT Dataset



Dataset: MMLU-Pro

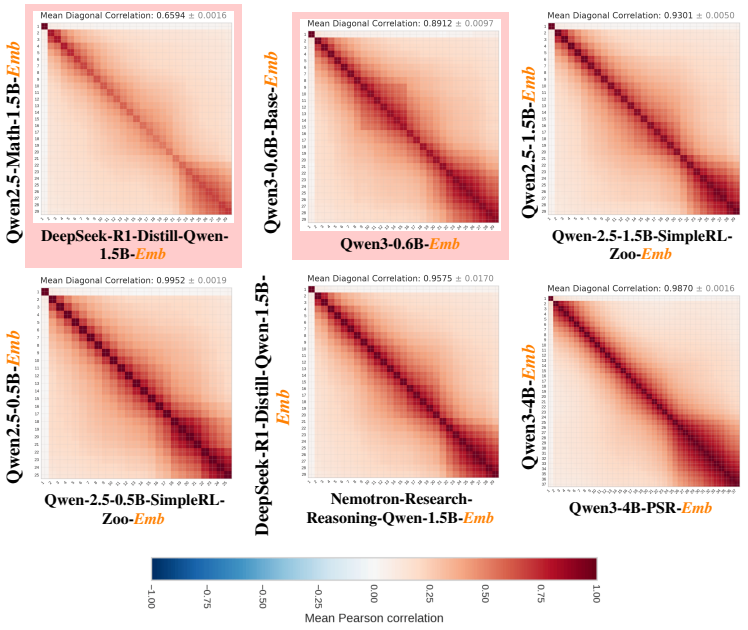


Figure 7: Additional Results on Dimension-Wise Correlation separated by dataset. The vertical axis and horizontal axis are Base Model Layer Index and Reasoning Model Layer Index, respectively. Gray values denote standard deviation across 3 seeds. The **red** background indicates their backbone LLMs are SFT-tuned pairs. The resulting embedding-model pairs show consistently strong axis alignment under RLVR-tuned backbones, whereas SFT pairs remain less uniformly aligned across datasets.

Table 11: Inverse row entropy of the orthogonal matrix O^* for $\mathcal{M}_{base}$ vs. $\mathcal{M}_{reason}$ across different datasets. Gray values denote standard deviation across 3 seeds. Higher inverse row entropy indicates more axis-aligned correspondence, while lower inverse row entropy indicates more globally mixed features. The model pairs are separated by the algorithm used to train their reasoning model $\mathcal{M}_{reason}$.

Orthogonal Procrustes Analysis

Base Models $\mathcal{M}_{base}$ vs. Reasoning Models $\mathcal{M}_{reason}$

Dataset: CoT Dataset

Model Pair	Inverse Row Entropy H_{inv} ↑
SFT	
Qwen2.5-Math-1.5B vs DeepSeek-R1-Distill-Qwen-1.5B	0.1076 ±0.0012
RLVR	
Qwen3-4B vs Polaris-4B-Preview	0.4365 ±0.0184
DeepSeek-R1-Distill-Qwen-7B vs Polaris-7B-Preview	0.6576 ±0.0127
Qwen2.5-7B vs zero__ppo__think__Qwen2.5-7B	0.6338 ±0.0153
Qwen2.5-1.5B vs Qwen-2.5-1.5B-SimpleRL-Zoo	0.9481 ±0.0068
Qwen2.5-0.5B vs Qwen-2.5-0.5B-SimpleRL-Zoo	0.9923 ±0.0019
DeepSeek-R1-Distill-Qwen-1.5B vs Nemotron-Research-Reasoning-Qwen-1.5B	0.1613 ±0.0094
Qwen3-4B vs Qwen3-4B-PSR	0.8122 ±0.0116

Dataset: MMLU-Pro

Model Pair	Inverse Row Entropy H_{inv} ↑
SFT	
Qwen2.5-Math-1.5B vs DeepSeek-R1-Distill-Qwen-1.5B	0.2229 ±0.0087
RLVR	
Qwen3-4B vs Polaris-4B-Preview	0.9711 ±0.0042
DeepSeek-R1-Distill-Qwen-7B vs Polaris-7B-Preview	0.9623 ±0.0051
Qwen2.5-7B vs zero__ppo__think__Qwen2.5-7B	0.9922 ±0.0014
Qwen2.5-1.5B vs Qwen-2.5-1.5B-SimpleRL-Zoo	0.9963 ±0.0009
Qwen2.5-0.5B vs Qwen-2.5-0.5B-SimpleRL-Zoo	0.9981 ±0.0005
DeepSeek-R1-Distill-Qwen-1.5B vs Nemotron-Research-Reasoning-Qwen-1.5B	0.8336 ±0.0138
Qwen3-4B vs Qwen3-4B-PSR	0.9843 ±0.0026

Table 12: Inverse row entropy of the orthogonal matrix O^* for $\mathcal{M}_{base}^{Emb}$ vs. $\mathcal{M}_{reason}^{Emb}$ across different datasets. Gray values denote standard deviation across 3 seeds. Higher inverse row entropy indicates more axis-aligned correspondence, while lower inverse row entropy indicates more globally mixed features. The model pairs are separated by the algorithm used to train their reasoning model backbone $\mathcal{M}_{reason}$.

Orthogonal Procrustes Analysis

Base Embedding Models $\mathcal{M}_{base}^{Emb}$ vs. Reasoning Embedding Models $\mathcal{M}_{reason}^{Emb}$

Dataset: CoT Dataset

Model Pair	Inverse Row Entropy H_{inv} ↑
SFT	
Qwen2.5-Math-1.5B-*Emb* vs DeepSeek-R1-Distill-Qwen-1.5B-*Emb*	0.1429 ±0.0034
Qwen3-0.6B-Base-*Emb* vs Qwen3-0.6B-*Emb*	0.4915 ±0.0162
RLVR	
Qwen2.5-1.5B-*Emb* vs Qwen-2.5-1.5B-SimpleRL-Zoo-*Emb*	0.8826 ±0.0091
Qwen2.5-0.5B-*Emb* vs Qwen-2.5-0.5B-SimpleRL-Zoo-*Emb*	0.9835 ±0.0023
DeepSeek-R1-Distill-Qwen-*Emb* vs Nemotron-Research-Reasoning-Qwen-*Emb*	0.8637 ±0.0108
Qwen3-4B-*Emb* vs Qwen3-4B-PSR-*Emb*	0.5863 ±0.0175

Dataset: MMLU-Pro

Model Pair	Inverse Row Entropy H_{inv} ↑
SFT	
Qwen2.5-Math-1.5B-*Emb* vs DeepSeek-R1-Distill-Qwen-1.5B-*Emb*	0.7105 ±0.0146
Qwen3-0.6B-Base-*Emb* vs Qwen3-0.6B-*Emb*	0.9164 ±0.0073
RLVR	
Qwen2.5-1.5B-*Emb* vs Qwen-2.5-1.5B-SimpleRL-Zoo-*Emb*	0.9794 ±0.0031
Qwen2.5-0.5B-*Emb* vs Qwen-2.5-0.5B-SimpleRL-Zoo-*Emb*	0.9978 ±0.0006
DeepSeek-R1-Distill-Qwen-*Emb* vs Nemotron-Research-Reasoning-Qwen-*Emb*	0.9814 ±0.0028
Qwen3-4B-*Emb* vs Qwen3-4B-PSR-*Emb*	0.9933 ±0.0011

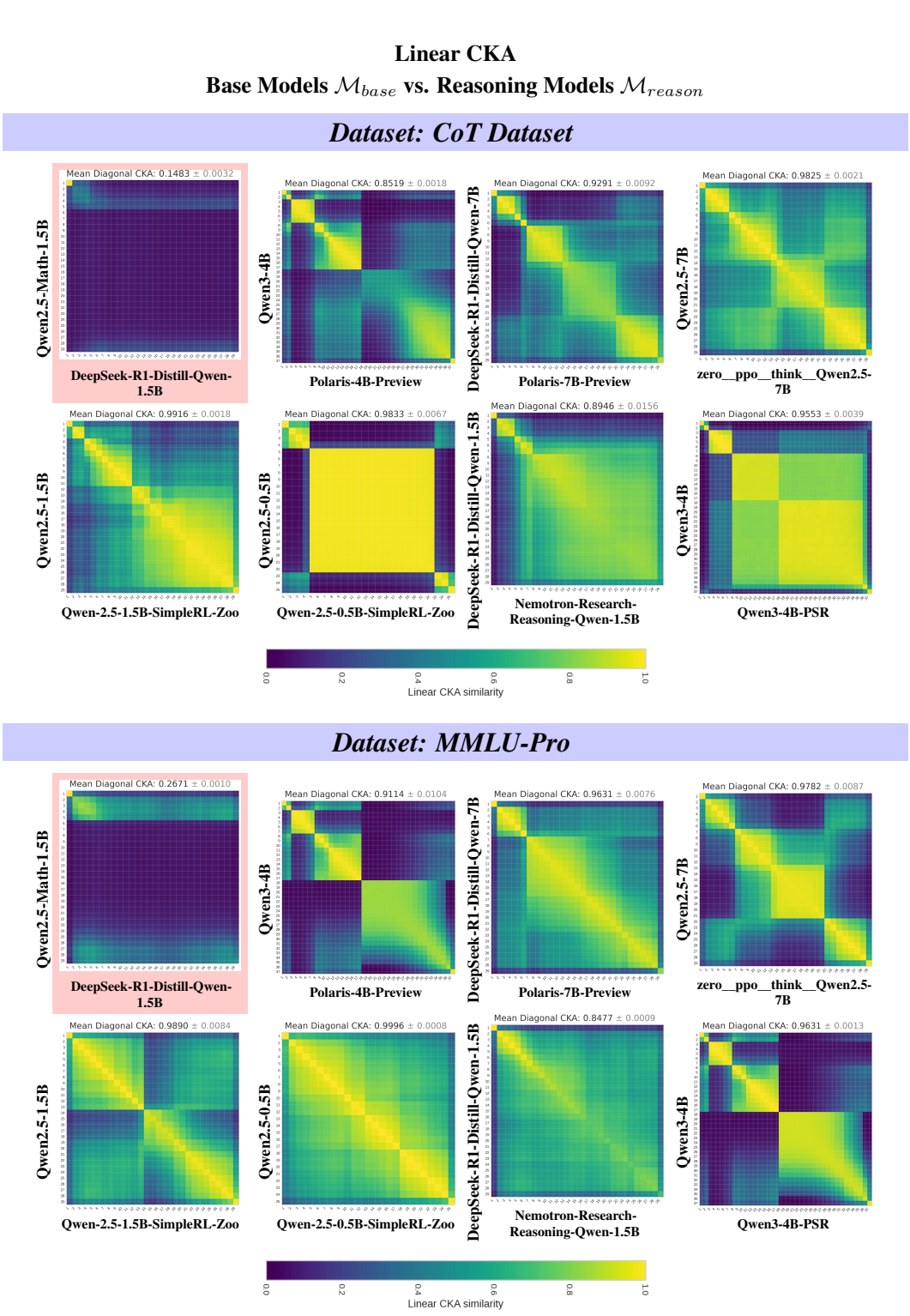


Figure 8: Additional Results on Linear CKA separated by dataset. The vertical axis and horizontal axis are Base Model Layer Index and Reasoning Model Layer Index, respectively. Gray values denote standard deviation across 3 seeds. The **red** background indicates SFT-tuned pairs. Across both datasets, RLVR pairs preserve substantially higher global manifold similarity than SFT pairs, a pattern that remains consistent across training algorithms and model sizes.

Linear CKA

Base Embedding Models $\mathcal{M}_{base}^{Emb}$ vs. Reasoning Embedding Models $\mathcal{M}_{reason}^{Emb}$

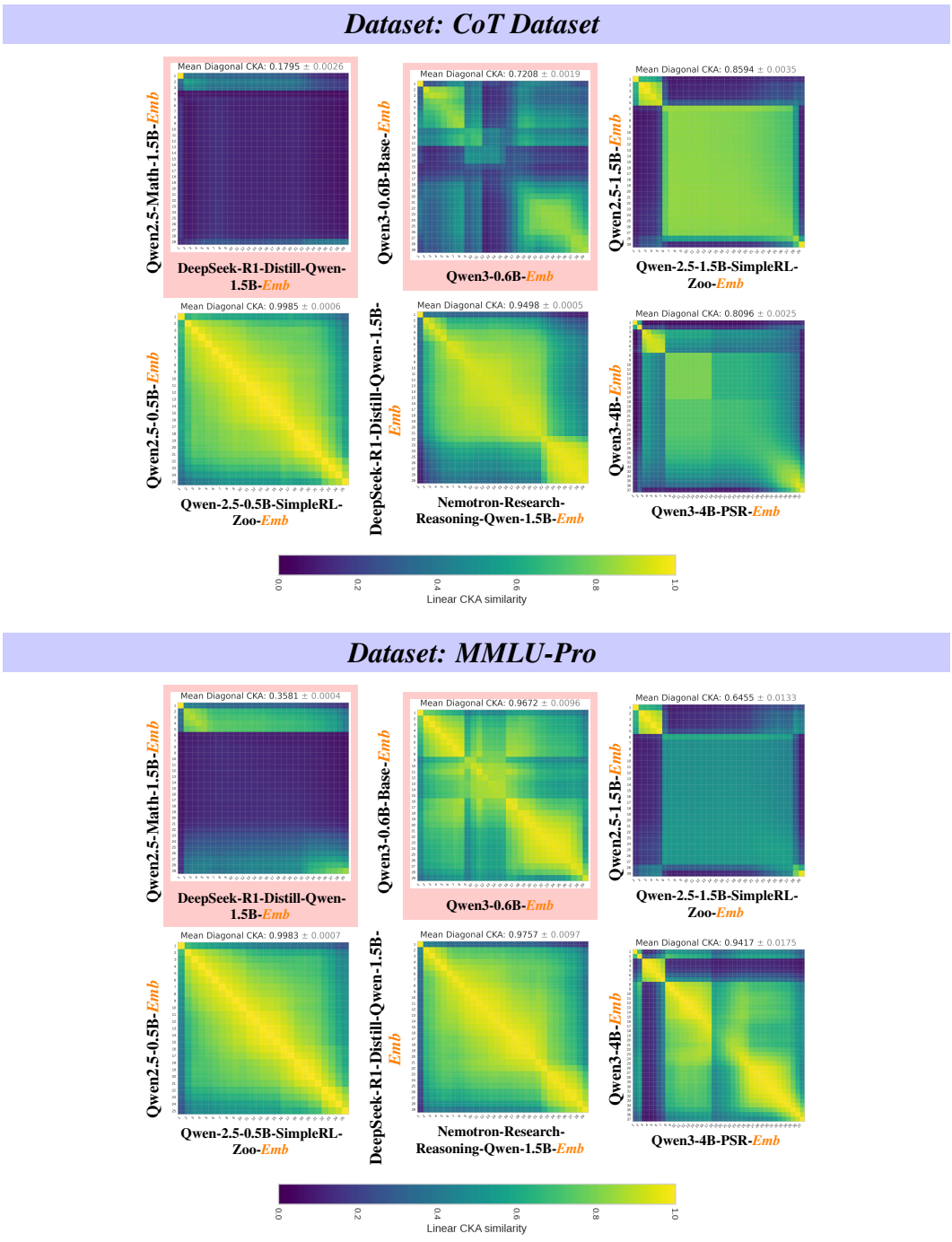


Figure 9: Additional Results on Linear CKA separated by dataset. The vertical axis and horizontal axis are Base Model Layer Index and Reasoning Model Layer Index, respectively. Gray values denote standard deviation across 3 seeds. The **red** background indicates their backbone LLMs are SFT-tuned pairs. The embedding-model comparisons preserve the same overall distinction: RLVR pairs generally maintain high global similarity, whereas SFT pairs exhibit less consistent alignment.

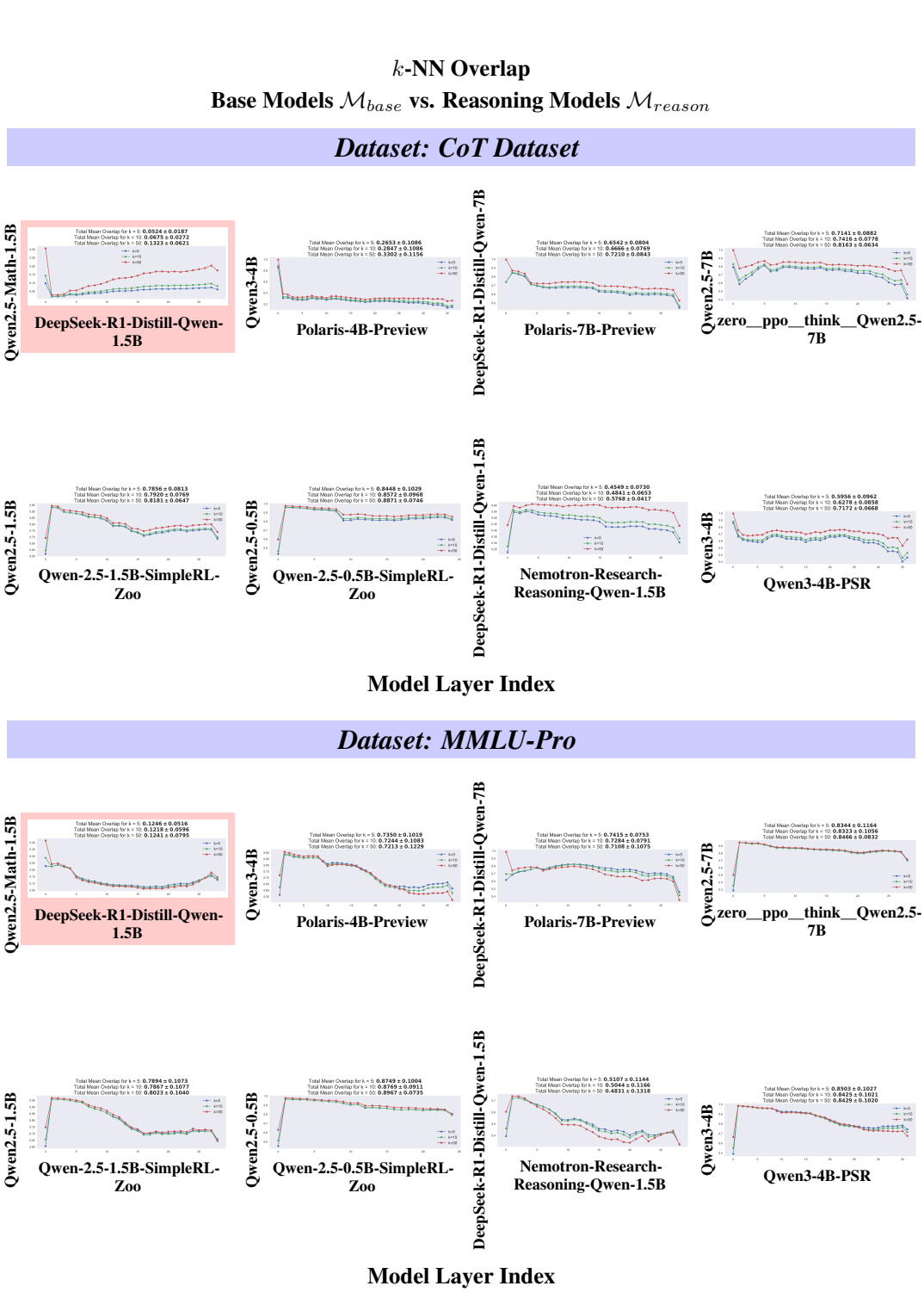


Figure 10: Additional Results on k-NN Overlap analysis separated by dataset. The vertical axis and horizontal axis are Mean Overlap and Model Layer Index, respectively. Gray values denote standard deviation across 3 seeds. The **red** background indicates SFT-tuned pairs. RLVR preserves substantially more local neighborhood structure than SFT across both datasets, although the sub-unity overlap indicates heterogeneous local reorganization.

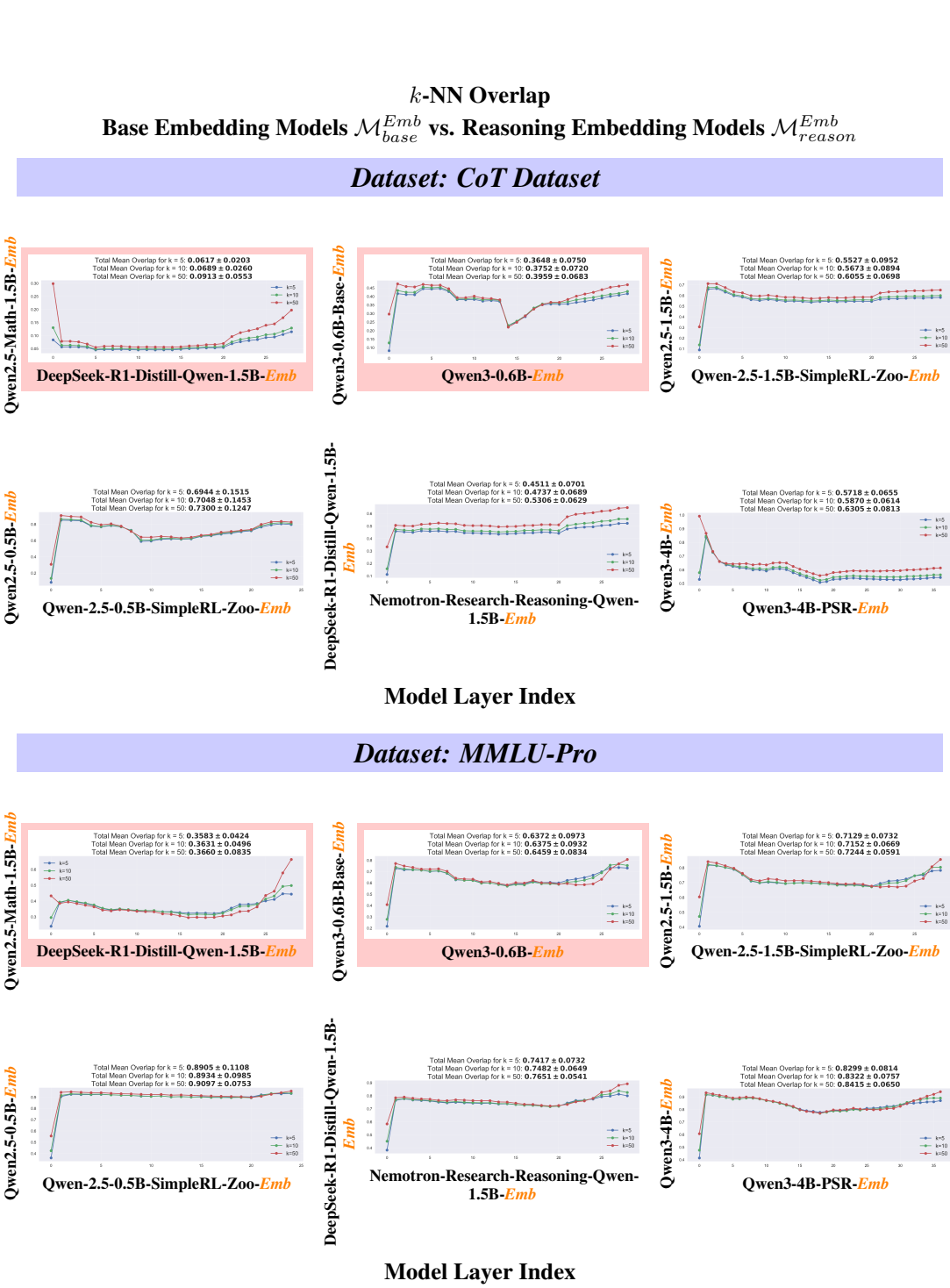


Figure 11: Additional Results on k-NN Overlap analysis separated by dataset. The vertical axis and horizontal axis are Mean overlap and Model Layer Index, respectively. Gray values denote standard deviation across 3 seeds. The **red** background indicates their backbone LLMs are SFT-tuned pairs. Embedding-model pairs retain only partial neighborhood overlap: RLVR generally remains more locally aligned than SFT, but contrastive learning does not restore local geometry.

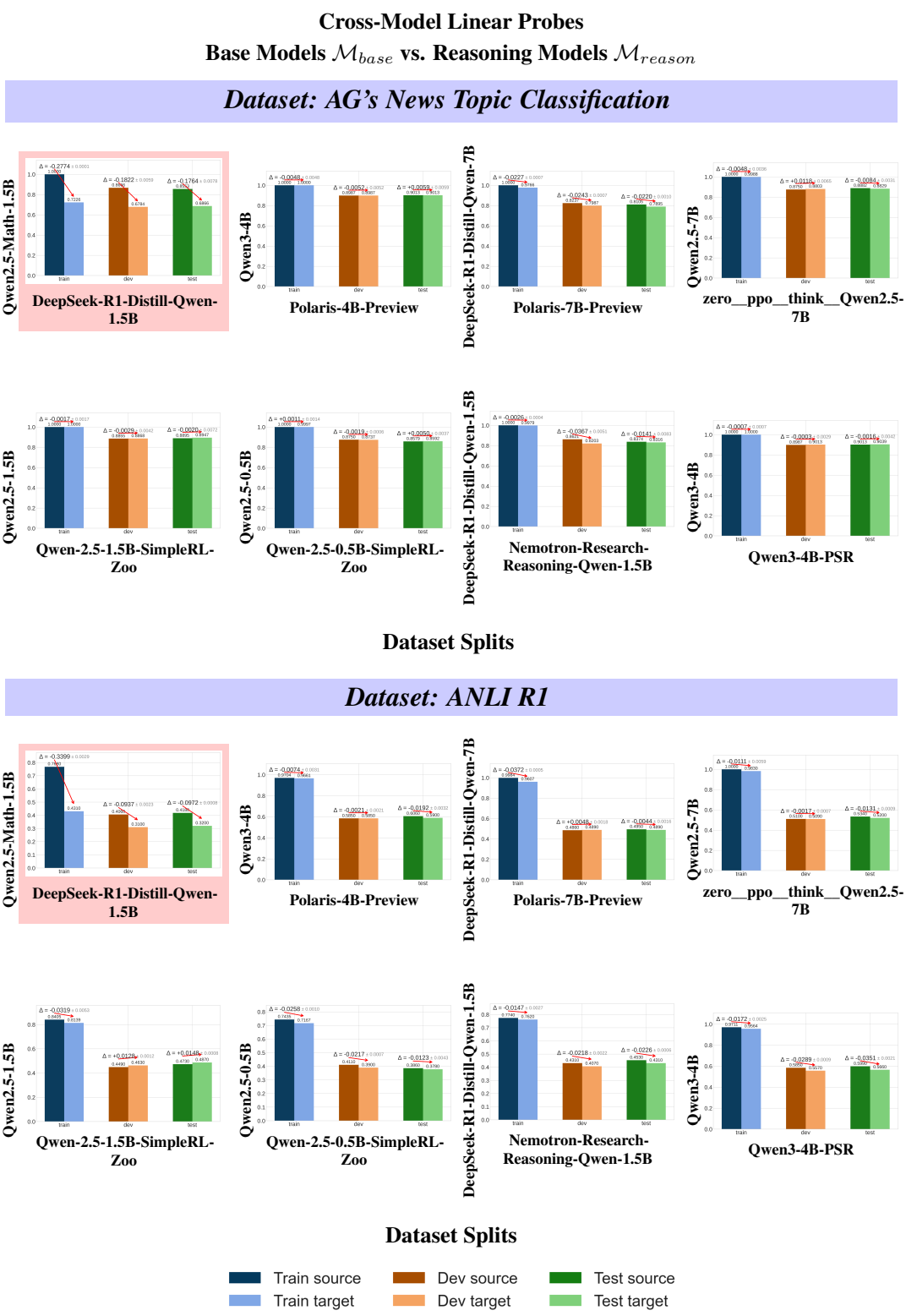


Figure 12: Additional Results on Cross-Model Linear Probe analysis. The vertical axis and horizontal axis are the Accuracy of the linear probe and Dataset types (train, dev, test), respectively. Gray values denote standard deviation across 3 seeds. The **red** background indicates SFT-tuned pairs. Across AG News and ANLI R1, RLVR pairs exhibit consistently smaller source-to-target accuracy gaps than SFT pairs, indicating stronger preservation of task-relevant linear readout directions.

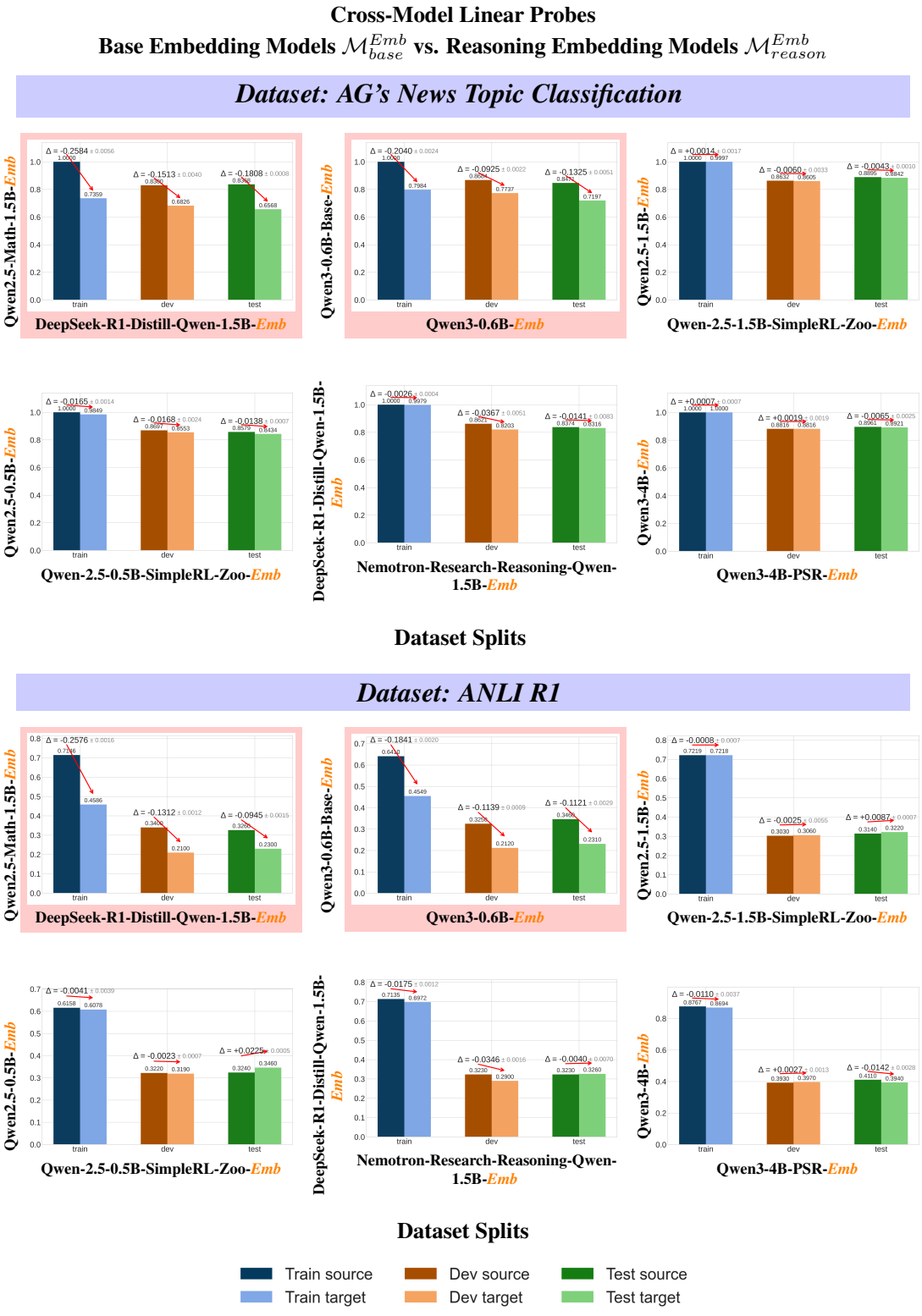


Figure 13: Additional Results on Cross-Model Linear Probe Analysis. The vertical axis and horizontal axis are the Accuracy of the linear probe and Dataset types (train, dev, test), respectively. Gray values denote standard deviation across 3 seeds. The **red** background indicates their backbone LLMs are SFT-tuned pairs. The embedding-model results reinforce this function-level pattern: RLVR-initialized pairs retain near-zero transfer gaps across both tasks, whereas SFT-initialized pairs exhibit substantially larger degradation.