occurring classes start to separate into their own clusters in both spaces (Figure 5B & C, dashed lines). Constrained by the information bottleneck condition, the system resorts to reusing the feature space eigenvectors and more information is selectively encoded in the dominant direction (note σ_1 grows faster compared to σ_2 after 200 steps in Figure 5D). This phase of anisotropic information encoding leads to a reduction in RankMe, akin to the "compression-seeking" phase. Taken together, these results suggest that gradient-based cross-entropy optimization dynamics under specific training conditions may result in non-monotonic changes in representation geometry we observed in LLMs.

Controls (Appendix): Removing skewed labels or information bottleneck eliminates "compression-seeking" (Figure 12); replacing cross-entropy with MSE yields monotonic, saturating expansion (Figure 13).

Key takeaway. Gradient descent on cross-entropy with (i) skewed token frequencies and (ii) a representation bottleneck ($d \ll |\mathcal{V}|$) suffices to produce expansion → compression via eigenvector alignment and singular-value growth proportional to magnitude. Negative controls (uniform labels / no bottleneck / MSE loss) remove "compression-seeking" , isolating necessary conditions. The eigenvector ablations (Table 2) show that downstream performance depends on the *full* eigenspectrum, justifying full-spectrum metrics over top-k proxies.

These mechanistic insights from simplified models establish fundamental principles governing representation geometry evolution. We now turn to examining how these geometric transformations manifest during post-training stages, where different optimization objectives and data distributions further sculpt the learned representations.

3.4. Representation geometric changes during Post-Training stages

While pretraining establishes the initial structure of LLM representations, subsequent post-training is instrumental for refining model capabilities and aligning them with downstream objectives. Here, we investigate the geometric changes that occur during each post-training stage. Our analysis centers on the Tülu-3.1 models (Wang et al., 2024), which utilize a sequential three-stage post-training recipe — Supervised Fine-tuning (SFT), Direct Preference Optimization (DPO), and Reinforcement Learning with Verifiable Rewards (RLVR) applied to the LLaMA-3.1-8B (Grattafiori et al., 2024) base model.

SFT exhibits "entropy-seeking" : We find that SFT is associated with a monotonic increase in the RankMe, indicating an increase in the underlying representation manifold complexity. See also detailed ID/OOD loss and win-rate behavior in Figure 14 (Appendix). We hypothesize that the manifold expansion is related to instruction memorization on in-distribution (ID) examples, while reducing robustness to out-of-distribution (OOD) samples. To test this, we perform SFT with Anthropic-HH dataset on OLMo2-1B intermediate checkpoints. As shown in Figure 6 B, we find that with more pretraining the ID loss on Anthropic-HH improves monotonically, while the OOD loss (on Alpaca farm data) increases. To understand the role of base-model geometry on the generalization gap, we perform SFT on Anthropic-HH (AH) and Alpaca farm (AF) datasets across checkpoints of OLMo2-1B, and measure chat winrates for AH using AF as reference on the AlpacaEval dataset. Strikingly, we find (Figure 6B bottom) that while more pretraining coincides with an increase in RankMe, the winrates decrease for AH. Notably, a drop in winrate from 14% to 9% suggests that the LLM judge is better able to distinguish between the outputs of the two instruction-tuned models. This reinforces that "overtrained" base models are more sensitive to distribution shifts under SFT.

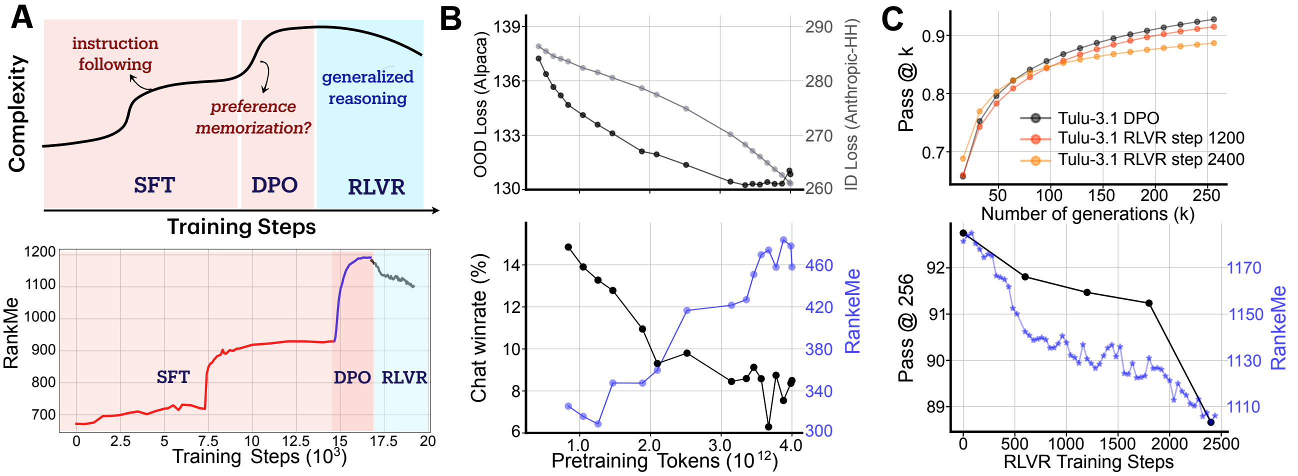

Figure 6: Post-training induces distinct geometric transformations in model representations, aligned with specific behavioral changes. (A) Conceptual overview of post-training (SFT, DPO and RLVR) **(top)**, corresponding $RankMe$ metrics from intermediate checkpoints of Llama-3.1-Tülu-3.1-8B **(bottom)** highlighting distinct progression for each stage. **(B)** Impact of pretraining on OLMo-2-1B SFT (Anthropic-HH): **(top)** longer pretraining improves in-distribution (ID) performance, while out-of-distribution (OOD) generalization (Alpaca farm) saturates **(bottom)** *Overtrained* models with higher RankMe exhibit markedly distinct outputs on AlpacaEval after undergoing SFT on two different datasets (Anthropic-HH and Alpaca farm). **(C)** RLVR post-training narrows base model's (Llama-3.1-8B-Tülu-3-DPO) exploratory behavior on AMC-23 (particularly at higher sampling counts e.g. $k = 256$), suggesting higher effective-rank facilitates better search.

DPO exhibits "entropy-seeking" : Prior works in self-supervised vision pretraining (Zhai et al., 2024; Ghosh et al., 2024) have established that contrastive learning objectives, e.g. SimCLR, are associated with an increase in representation complexity, as the network progressively learns the relevant eigenfunctions (Simon et al., 2023) to separate the positive and negative examples. We observe a similar trend in the DPO stage, notably a monotonic increase (decrease) in the RankMe (α_{ReQ}), c.f. fig. 6A. This parallel between the two settings can be attributed to the analogous formulations in the objective function. Note below that eq. (3) can be written as the Noise Contrastive Estimation (NCE) loss (Gutmann and Hyvärinen, 2010), often used in contrastive vision and multimodal pretraining (Oord et al., 2018; Chen et al., 2020; Radford et al., 2021), with *one* negative example.

$$\mathcal{L}_{DPO} = -\mathbb{E}_{x,y_w,y_l}\left[\log(\sigma(\hat{r}_\theta(x,y_w) - \hat{r}_\theta(x,y_l)))\right] = -\mathbb{E}_{x,y_w,y_l}\left[\log\frac{e^{\hat{r}_\theta(x,y_w)}}{e^{\hat{r}_\theta(x,y_w)} + e^{\hat{r}_\theta(x,y_l)}}\right] \tag{5}$$

RLVR exhibits "compression-seeking" : In sharp contrast to SFT and DPO, we observe that RLVR is associated with a monotonic decrease in RankMe (cf. Figure 6A). To probe the implications of this "compression-seeking" stage, we evaluate the unbiased `pass@k` performance on AMC-23 math benchmark. Figure 6C shows that while RLVR-training for 2400 steps outperforms the base (post-DPO) model at `pass@16`, the base model as well as an intermediate checkpoints outperform the RLVR-trained model at `pass@256`. This decline in `pass@256` performance as training progresses, reinforces prior work (Yue et al., 2025) suggesting that RLVR constraints the exploratory behavior of base models while amplifying some pre-existing behaviors of the base model (Zhao et al., 2025).

Key takeaway. Post-training induces mirrored spectral transformations with practical trade-offs: SFT/DPO (RankMe ↑, α_{ReQ} ↓) enhance in-distribution fit but increase sensitivity to dataset idiosyncrasies; RLVR (RankMe ↓, α_{ReQ} ↑) consolidates reward-aligned behaviors and narrows high-k exploration (pass@k), consistent with reduced solution diversity.

4. Related Work

Dynamics of Knowledge Acquisition and Representation Learning A central theme in understanding neural networks is that learning is a dynamic, often phased process rather than a monolithic one. Recent work by (Zucchet et al., 2025) identified distinct stages in how LLMs learn factual information, highlighting the formation of critical internal structures like attention circuits during performance plateaus. This notion of staged learning is further supported by the "Distributional Simplicity Bias" (DSB) established by (Refinetti et al., 2023; Belrose et al., 2024), which posits that networks learn simpler statistical properties of data (e.g., lower-order moments) before more complex ones. Our work provides a geometric lens on these phenomena, using spectral measures to track how the effective dimensionality of representations evolve non-monotonically. Furthermore, (Michaud et al., 2023) proposed that scaling laws and emergent abilities arise from learning discrete "quanta" of knowledge. (DeMoss et al., 2024) explained grokking (Power et al., 2022) as a transition from high-complexity memorization to low-complexity generalization, measured via algorithmic information theory. Our spectral geometric phases offer a complementary perspective that could underpin these observed emergent jumps in performance and the dynamics of grokking.

Post-Training Alignment and Reasoning The adaptation of pretrained LLMs through fine-tuning is critical for aligning them with specific tasks and user preferences. (Ren and Sutherland, 2024) provided an empirical-NTK based framework to decompose the influence of fine-tuning updates, explaining complex behaviors such as hallucination amplification in SFT and the "squeezing effect" in DPO, where confidence in desired outputs can paradoxically decrease. Concurrently, (Springer et al., 2025) identified "catastrophic overtraining," showing that excessive pretraining can make models overly sensitive to parameter changes, thereby degrading performance after SFT. Our work contributes to this area by demonstrating that different post-training strategies (SFT, DPO, RLVR) induce distinct transformations in the geometry and it's influence model capabilities.

5. Discussions

Geometry of Pretraining: Memorization vs Generalization. We show that LLM pretraining is multiphasic rather than monotonic, characterized mainly by "entropy-seeking" and "compression-seeking" phases. The observed geometric phases provide a quantitative framework for examining the relationship between memorizing short-context statistics and generalizing long-context information. The "entropy-seeking" phase expands the representational space to capture various short-context patterns, including n-gram memorization. Conversely, the "compression-seeking" phase promotes a more structured manifold and is likely to incentivize generalizable long-range language understanding. This geometric refinement process is consistent with and may offer an explanation for phenomena like *grokking*, where generalization capabilities can emerge after an initial period of fitting.

Our preliminary analysis further reveals the importance of full-spectrum information for model per-