

C.2. Computing spectral metrics, RankMe and α_{ReQ}

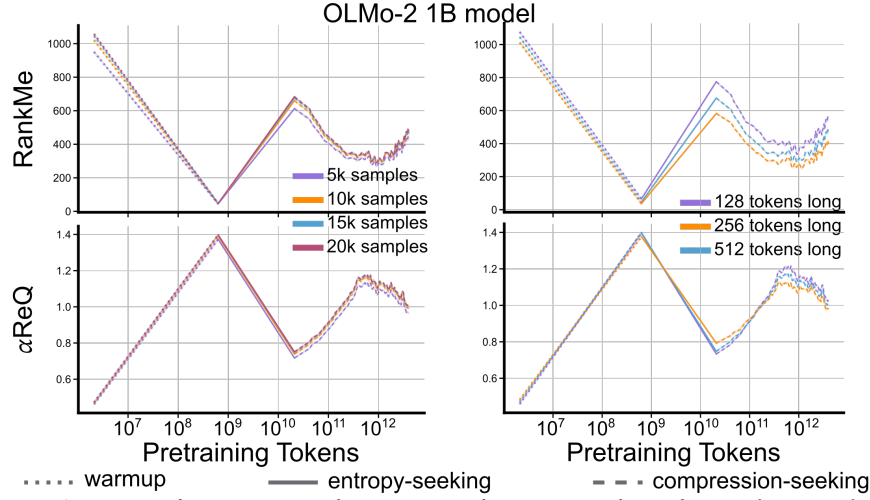


Figure 8: Spectral metrics are robust to sample count and sequence length. RankMe and α_{ReQ} computed for intermediate checkpoints of the OLMo-2-1B model using (Left) different number of samples, and (Right) sequence length. The three-phase pattern remains consistent across these methodological choices. Shaded error bars indicate standard error about mean.

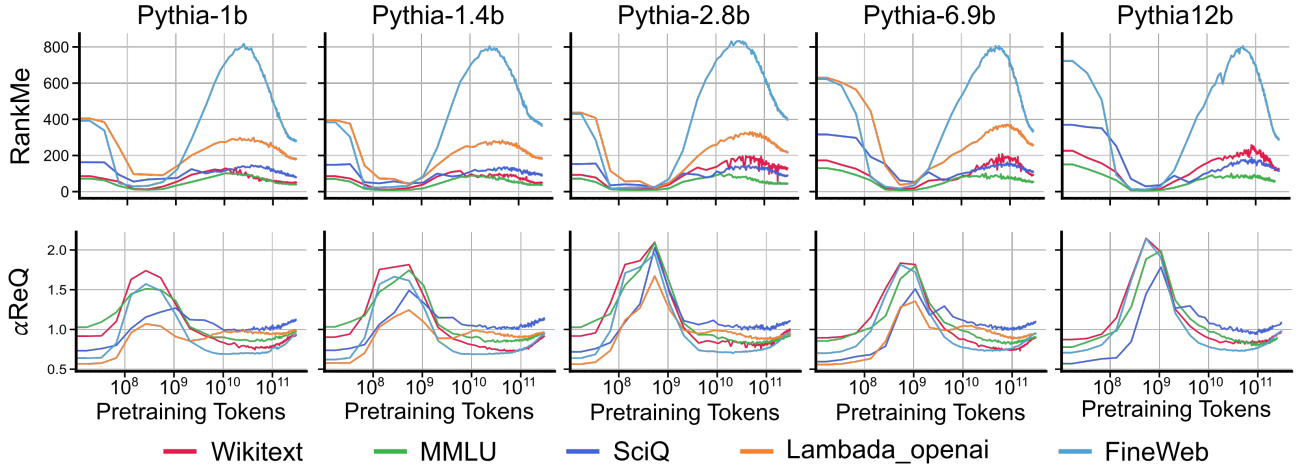


Figure 9: Pythia spectral metrics are robust across datasets. RankMe and α_{ReQ} computed for intermediate checkpoints of models from the Pythia family (1B-12B) on different datasets, showing consistent phase patterns across evaluation data.

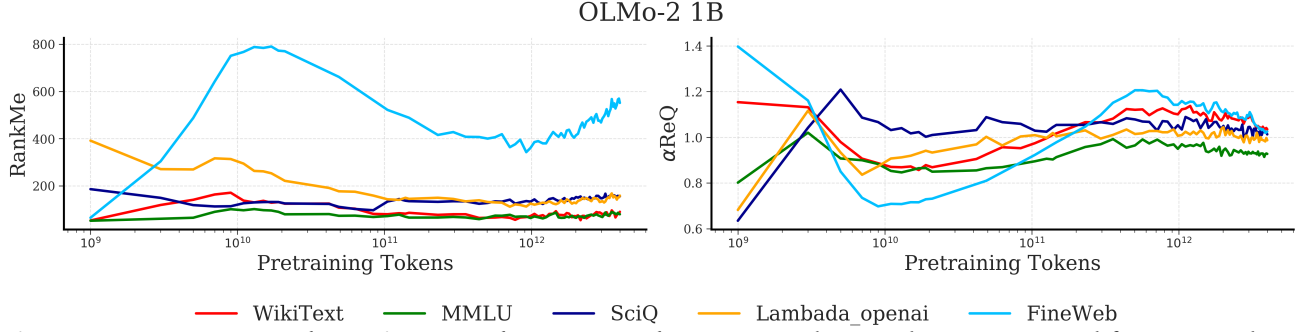


Figure 10: OLMo spectral metrics are robust across datasets. RankMe and α_{ReQ} computed for intermediate checkpoints of OLMo-2 1B model on different datasets, showing consistent phase patterns across evaluation data.

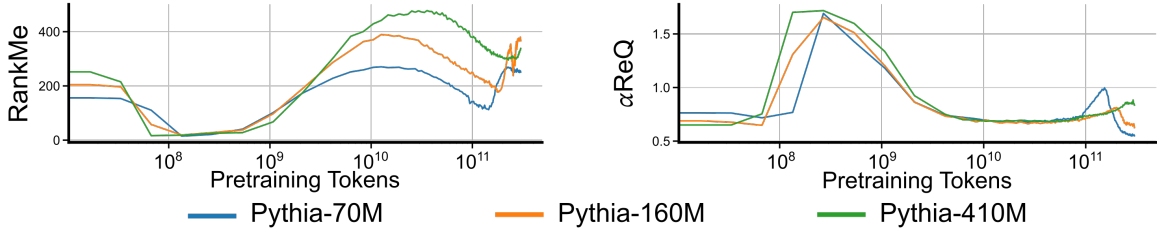


Figure 11: RankMe and α_{ReQ} computed for intermediate checkpoints of smaller models ($< 1\text{B}$) from the Pythia family on the FineWeb dataset.

Model	α_{ReQ} (p-value)	RankMe (p-value)
Pythia-1B	0.810 (1.50e-5)	-0.759 (1.04e-4)
Pythia-1.4B	0.668 (1.29e-3)	-0.713 (4.18e-4)
Pythia-2.8B	0.694 (6.88e-4)	-0.635 (2.63e-3)
Pythia-6.9B	0.837 (4.20e-6)	-0.885 (2.19e-7)
Pythia-12B	0.836 (4.42e-6)	-0.839 (3.79e-6)
OLMo2-1B	0.540 (4.920e-7)	-0.616 (3.201e-9)

Table 9: SciQ accuracy correlates with spectral geometry. Positive correlation with α_{ReQ} (compactness) and negative correlation with RankMe (effective dimensionality) across Pythia (1–12B) and OLMo2-1B models. The p-values in parentheses indicate high statistical significance for all correlations.