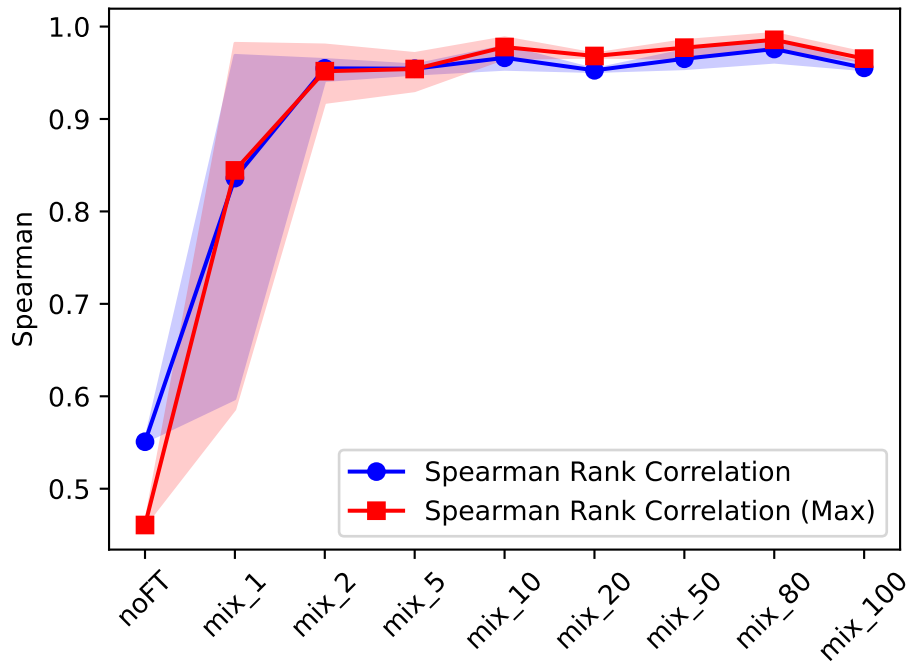
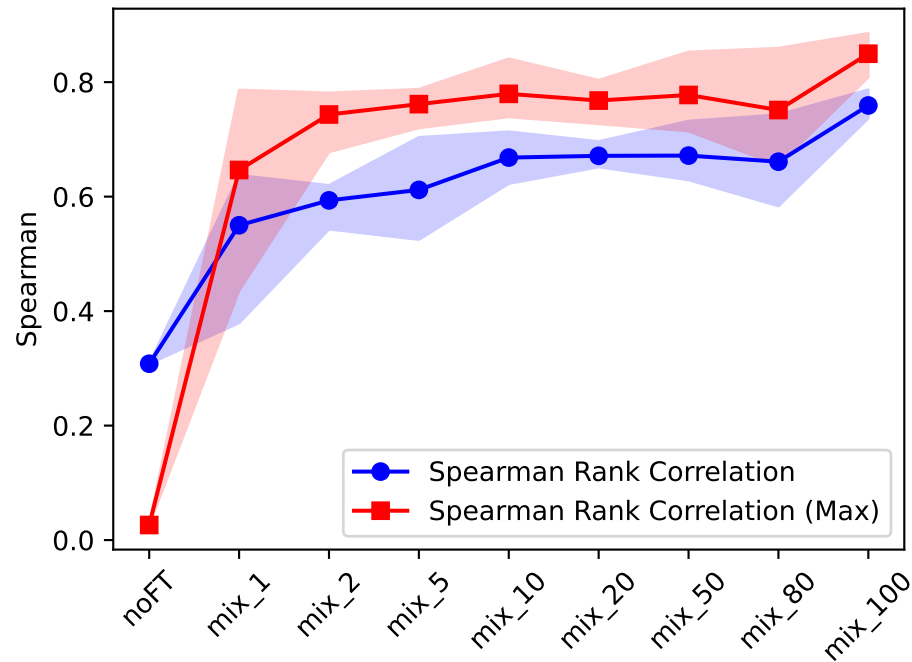


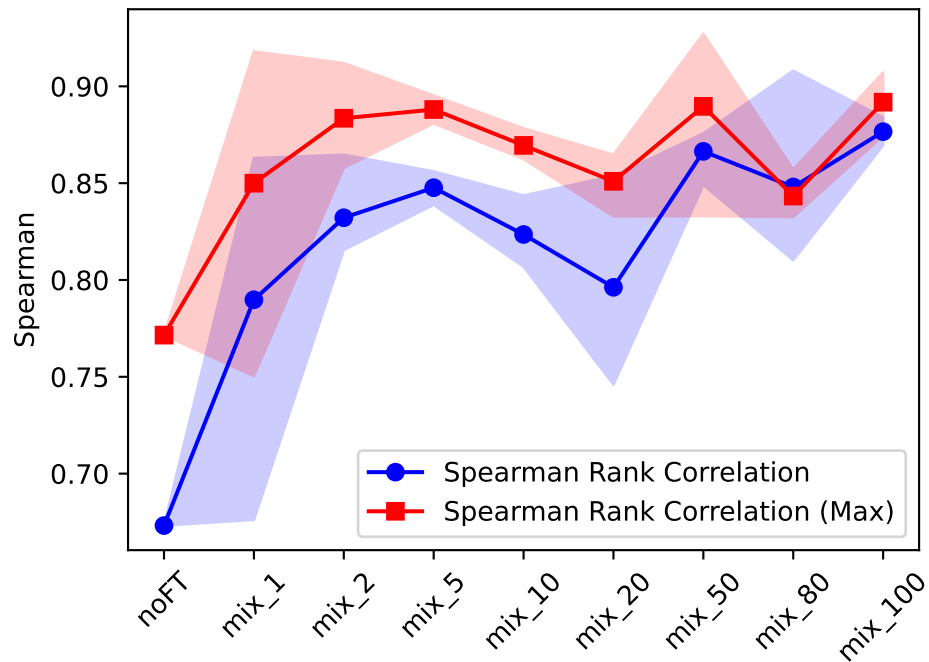
ARC Challenge



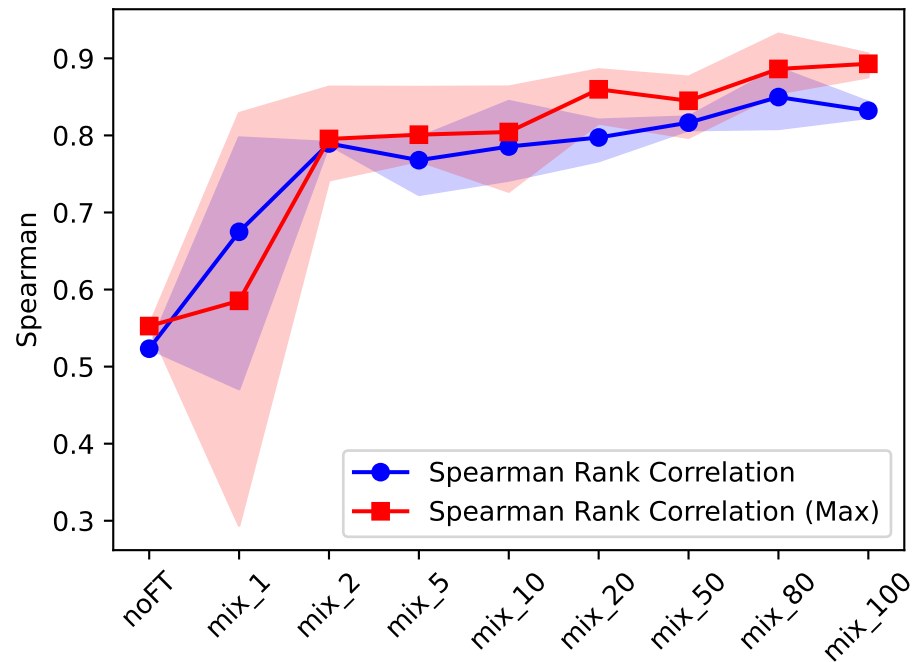
GPQA



PIQA



OpenBookQA



Note: Spearman Correlation (Max) = Spearman's rank correlation with maximum performance in standard evaluation.