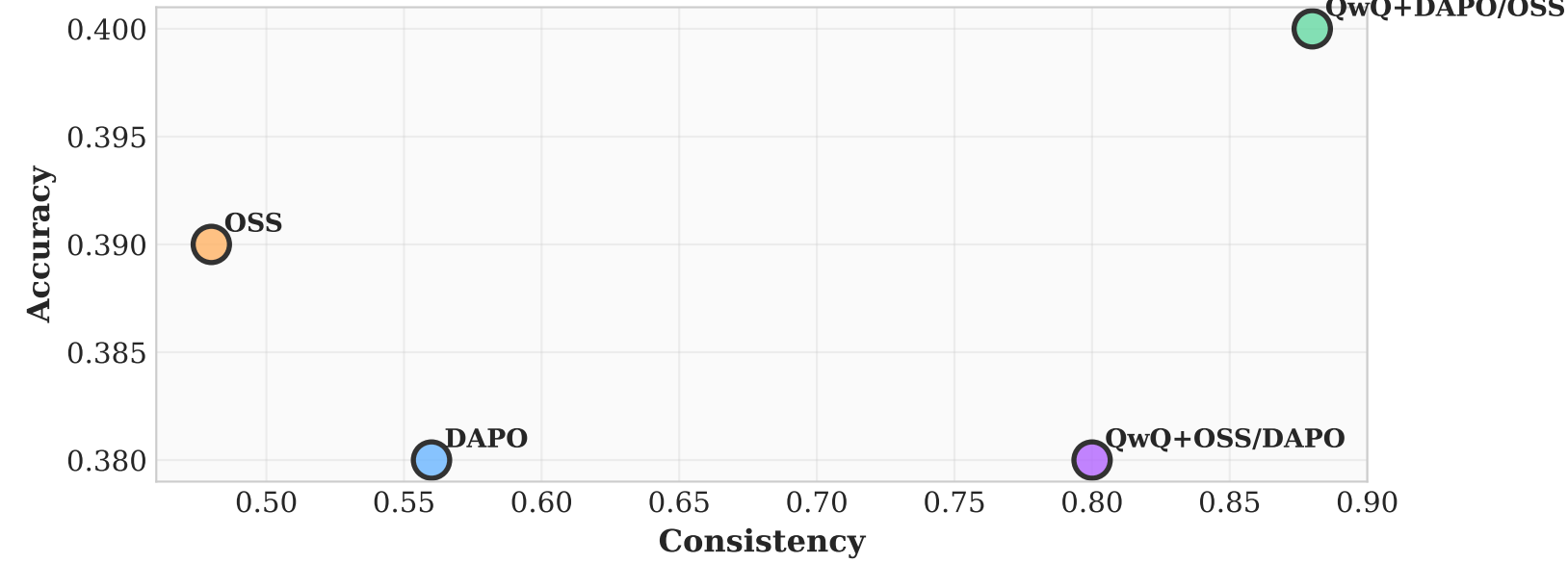


Comprehensive Model Evaluation Dashboard

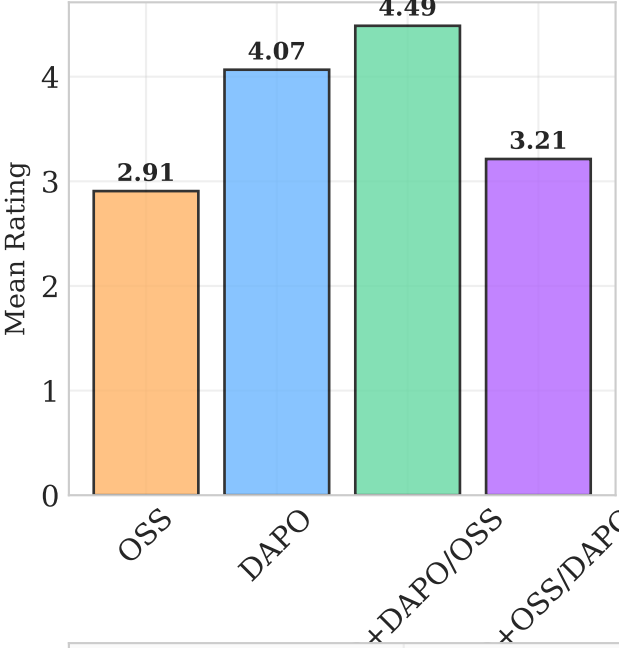
Performance Metrics



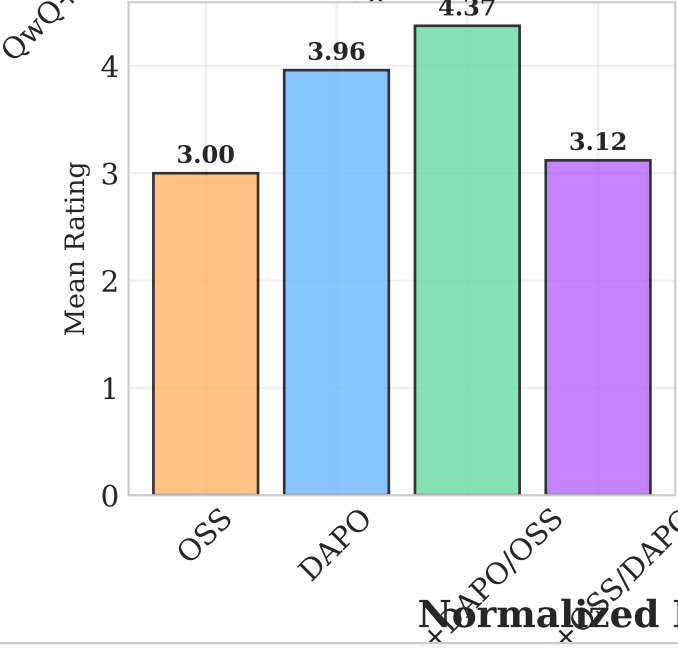
Consistency vs Accuracy



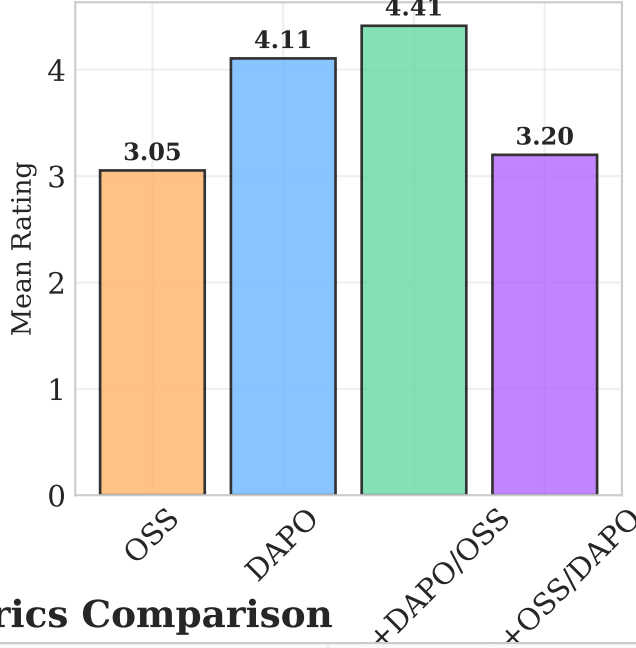
Clarity of Steps



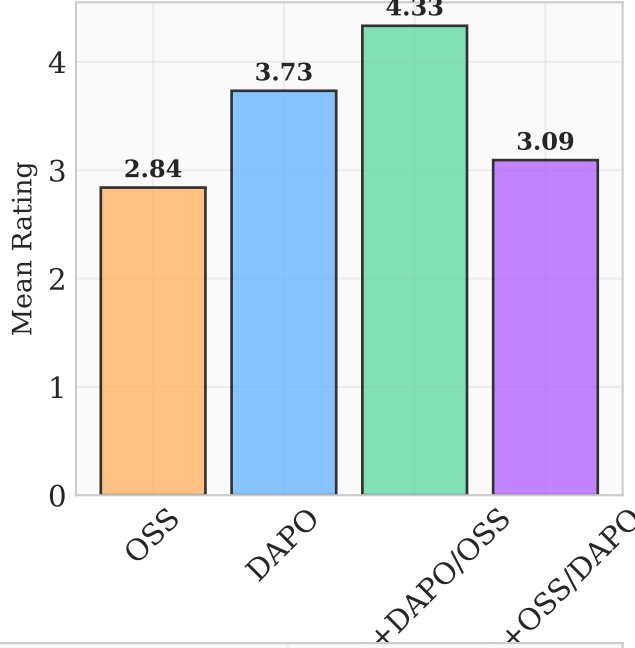
Ease of following



Confidence



Best Overall



Normalized Metrics Comparison

