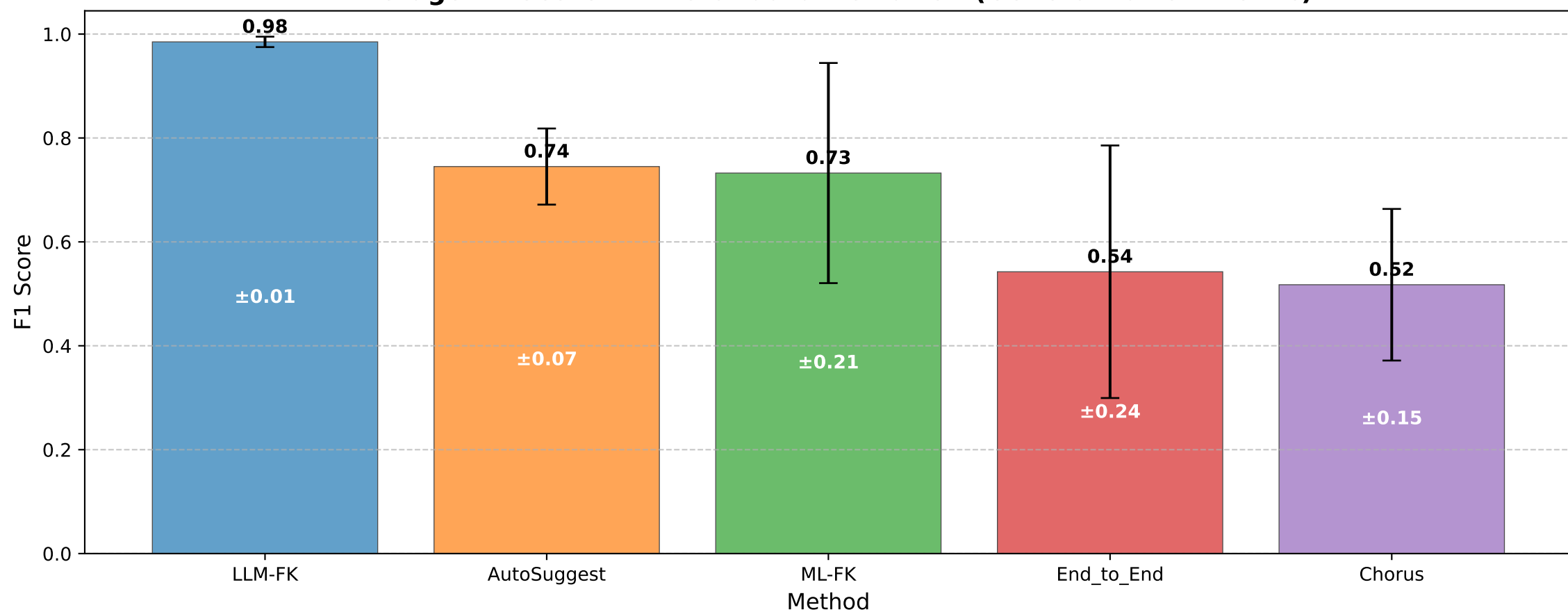


Average F1 Score with Standard Deviation (Generalization Metric)



Performance Range Across Datasets (Max F1 - Min F1)

