Average Generation Metrics per Configuration 0.6 0.5 0.4 Average Score 0.2 0.1 0.0 bertscore\_f1

Evaluation Metric

