Average Generation Metrics per Configuration 0.6 0.5 0.4 Average Score 5.0 8.0 8.0 0.2 0.1 0.0 rouge\_L\_precision rouge\_L\_recall bertscore\_precision bertscore\_recall

**Evaluation Metric** 

