Average Generation Metrics per Configuration 0.7 0.6 0.5 Average Score 0.3 0.2 0.1 0.0 bleu\_score rouge\_1\_f1 rouge\_1\_precision rouge\_1\_recall

**Evaluation Metric** 

