Average Generation Metrics per Configuration 0.6 0.5 0.4 Average Score o o 0.2 0.1 0.0 bleu_score answer_relevancy rouge_1_f1 rouge_1_precision

Evaluation Metric

