

	Coherence		Consistency		Fluency		Relevance	
Metrics	sys	sum	sys	sum	sys	sum	sys	sum
ROUGE-1	0.67	0.49	0.38	0.25	0.66**	0.46	0.73**	0.42
ROUGE-2	0.52	0.4	0.36	0.24	0.49	0.36	0.66**	0.4
ROUGE-3	0.41	0.32	0.34	0.23	0.38	0.28	0.60*	0.37
ROUGE-4	0.35	0.28	0.31	0.22	0.31	0.23	0.56*	0.34
ROUGE-L	0.66**	0.49	0.35	0.23	0.64*	0.45	0.70**	0.41
BERTScore-p	0.84**	0.66	0.09	0.07	0.78**	0.58	0.53	0.32
BERTScore-r	0.45	0.36	0.4	0.28	0.44	0.33	0.68**	0.43
BERTScore-f1	0.69**	0.55	0.24	0.17	0.65*	0.49	0.63*	0.39
MoverScore	0.57*	0.45	0.34	0.24	0.55*	0.41	0.67**	0.41
SMS	0.33	0.24	0.32	0.22	0.30*	0.21	0.56	0.34
BARTScore-s-h+	-0.44	-0.36	0.65*	0.44	-0.26	-0.21	0.26	0.17
BARTScore-h-r	0.15	0.09	-0.08	-0.06	0.23	0.17	-0.04	-0.04
BARTScore-h-r	0.48	0.36	0.5	0.32	0.48	0.34	0.76**	0.45
BARTScore-r-h	0.85**	0.6	0.29	0.21	0.85**	0.59	0.69**	0.38
BLANC-help+	-0.79**	-0.56	0.52	0.36	-0.64*	-0.43	0.11	0.11
BLANC-tune+	-0.82**	-0.59	0.48	0.3	-0.68**	-0.46	0.06	0.07
FEQA+	0.89**	0.41	0.37	0.15	0.92**	0.4	0.63**	0.22
QuestEval+	0.1	0.04	0.89**	0.3	0.31	0.08	0.72**	0.24
SummaQA-conf+	-0.66*	-0.34	0.64*	0.33	-0.52	-0.26	0.23	0.13
SummaQA-fscore+	-0.74**	-0.32	0.55*	0.24	-0.59*	-0.24	0.16	0.07
PPL-	0.17	0.06	-0.48	-0.25	-0.02	-0.08	-0.35	-0.17
CHRF	0.43	0.34	0.41	0.28	0.42	0.32	0.68**	0.42
BLEU-1	0.41	0.33	0.29	0.2	0.37	0.28	0.57*	0.35
BLEU-2	0.34	0.28	0.29	0.21	0.29	0.23	0.54*	0.35
BLEU-3	0.3	0.25	0.27	0.2	0.25	0.2	0.51	0.33
BLEU-4	0.25	0.21	0.27	0.19	0.2	0.16	0.49	0.32
METEOR	0.38	0.31	0.36	0.26	0.35	0.28	0.62*	0.4
EmbeddingAverage	0.78**	0.53	0.12	0.09	0.77**	0.51	0.59*	0.31
VectorExtrema	0.56*	0.43	0.3	0.2	0.53*	0.39	0.64*	0.38
GreedyMatching	0.56*	0.42	0.3	0.2	0.54*	0.4	0.65*	0.37
FactCC+	-0.83**	-0.5	0.45	0.21	-0.73**	-0.41	0	-0.04
DAE+	-0.79**	-0.53	0.51	0.27	-0.67**	-0.43	0.04	0