

Average Evaluation Scores of First 20 Samples				
	Automatic Bert	Automatic Bleu	Human Factuality	Human Fluency
Greedy Search	90.22%	10.31%	2.97	2.88
Beam Search	90.10%	9.19%	2.67	2.87
Top-k sampling	90.44%	11.16%	2.92	3.18
Top-p Sampling	90.71%	12.49%	3.38	3.3