Appendix

We procide an example of generated summaries with the accompanying human scores w.r.t. the readability and quality of the summaries, as shown in Table A.1. Three graduate students working on NLP were invited to rank each summary generated by all 6 models based on their readability and quality, where 1 indicates the lowest level of readability and quality while 6 indicates the highest level. We measure the agreement as the number of people who submitted the same results, and we measure the performance of the models by using the samples that satisfy at least 2 out of 3 agreement. Thus, the final human evaluation scores for the six models (MCARS, GANsum, DeepRL, PGC, ABS, and SummaRuNNer) are (1, 4, 3, 5, 6, 2), respectively.

Input summary: The NOOKColor is awesome- it plays music, is expandable, displays gorgeous color, allows you to surf the internet (not optimally, but the ability is there), and is mostly easy to navigate. From me, it gets a solid 4 stars. The problem is that as an eReader, it's already overpriced, and the Amazon sellers who offer it are asking ridiculous amounts. It's cheaper (and ships free) from Barnes and Noble. Amazon is usually No1 at everything, including price, and I already own two Kindles (I would buy the Kindle in color if it came that way) but the price being charged for the NOOKColor on Amazon is absurd.
Ground-truth summary: Awesome eReader, Ridiculously Priced from this seller.
Summary by MCARS: Awesome NOOKColor, absurd Amazon price. **Scores by Annotator 1, 2, 3:** (1, 1, 1)
Summary by GANsum: NOOKColor is awesome music gorgeous color Amazon. **Scores by Annotator 1, 2, 3:** (4, 4, 4)
Summary by DeepRL: NOOcolor gets 4 stars. **Scores by Annotator 1, 2, 3:** (3, 3, 2)
Summary by PGC: NOOKColor easy to navigate Amazon **Scores by Annotator 1, 2, 3:** (5, 5, 5)
Summary by ABS: NOOKColor easy good easy ridiculous amount. **Scores by Annotator 1, 2, 3:** (6, 6, 6)
Summary by SummaRuNNer: The NOOKColor is awesome. **Scores by Annotator 1, 2, 3:** (2, 2, 3)

Table A.1 Examples of generated summaries and their human evaluation scores