Discriminative Adversarial Search for Abstractive Summarization
Supplementary Material

Thomas Scialom\textsuperscript{1,2} Paul-Alexis Dray\textsuperscript{1} Sylvain Lamprier\textsuperscript{2} Benjamin Piwowarski\textsuperscript{2,3} Jacopo Staiano\textsuperscript{1}

\textbf{1. Implementation details}

All models are implemented in PyText (Aly et al., 2018). For all our experiments we used a single RTX 2080 Ti GPU.

To train the discriminator, we used the Adam optimiser with the recommended parameters for BERT: learning rate of $3e^{-5}$, batch size of 4 and accumulated batch size of 32. We trained it for 5 epochs; each epoch took 100 minutes on 150k samples. During discriminator retraining, the generator is needed and thus additional memory is required: all else equal, we decreased the batch size to 2. The self-training process takes one epoch to converge, in about 500 minutes: 200 minutes for training the discriminator and 300 minutes to generate the summaries with the search procedure described in Algorithm 1.

\textbf{References}