

Supplementary Material

We have uploaded our model checkpoints at all stages through an anonymous account:

https://drive.google.com/file/d/1CJYvln97NWOXff_DxIHdWJaIZ72HYs/view?usp=sharing

The evaluation is through the publicly available evaluation repository lm-evaluation-harness with default settings:

<https://github.com/EleutherAI/lm-evaluation-harness>

We will release our BERT topic model, code for our selection method and baseline implementation, as well as labeled datasets soon.