

KL(BERT||DistilBERT) for 10,000 random examples from Wikipedia and Bookcorpus
mean: 5.85 (± 0.06)

