

Test set KL divergence of TM & LM after the same amount of training steps

