

Pythia 12B: 1,559,914 tokens randomly sampled per document position bin

