

Pythia 12B: 4,605,750 tokens randomly sampled per token position bin

