

Pythia 70M: 3,465,844 tokens randomly sampled per document position bin

