

Pythia 12B deduped: 2,261,935 tokens randomly sampled per token position bin

