	ARC-Challenge		MMLU		HellaSwag		WinoGrande		Lambada (OpenAI)		Lambada (OpenAI)		Blimp		Lambada (Std)		Lambada (Std)	
	acc	stddev	acc	stddev	acc	stddev	acc	stddev	acc	stddev	perplexity	stddev	acc	stddev	acc	stddev	perplexity	stddev
160M Pile Dedup	0.200	0.012	0.230	0.004	0.290	0.005	0.496	0.014	0.369	0.007	31.259	1.159	0.729	0.002	0.234	0.006	172.762	7.727
160M MiniPile	0.213	0.012	0.270	0.004	0.256	0.004	0.472	0.014	0.000	0.000	3033175.269	288926.583	0.519	0.002	0.000	0.000	27067951.346	2710040.191
160M Reproduction	0.189	0.012	0.230	0.004	0.260	0.004	0.512	0.014	0.000	0.000	1854408.400	148101.598	0.548	0.002	0.000	0.000	11927123.251	1063672.928
160M Lossi	0.198	0.012	0.230	0.004	0.260	0.004	0.511	0.014	0.000	0.000	2116445.173	175403.058	0.549	0.002	0.000	0.000	14896599.925	1366937.547
160M Density	0.192	0.012	0.230	0.004	0.260	0.004	0.520	0.014	0.000	0.000	2099002.091	170652.622	0.550	0.002	0.000	0.000	13347273.608	1997894.636
160M k440	0.197	0.012	0.230	0.004	0.262	0.004	0.511	0.014	0.000	0.000	1854900.791	147593.481	0.547	0.002	0.000	0.000	11658172.431	1033012.414
160M k440 Density	0.193	0.012	0.230	0.004	0.260	0.004	0.494	0.014	0.000	0.000	2025523.777	164221.889	0.552	0.002	0.000	0.000	12959844.941	1160155.065
160M k440 Inter	0.194	0.012	0.230	0.004	0.261	0.004	0.500	0.014	0.000	0.000	1858348.205	147853.142	0.551	0.002	0.000	0.000	11655568.314	1032438.429

Table 1: Performance comparison of Pythia 160M models trained on Pile and MiniPile versions

	ARC-Challenge		MMLU		HellaSwag		WinoGrande		Lambada (OpenAI)		Lambada (OpenAI)		Blimp		Lambada (Std)		Lambada (Std)	
	acc	stddev	acc	stddev	acc	stddev	acc	stddev	acc	stddev	perplexity	stddev	acc	stddev	acc	stddev	perplexity	stddev
160M Low Density	0.189	0.011	0.230	0.004	0.251	0.004	0.507	0.014	0.000	0.000	2287598.555	192724.615	0.550	0.017	0.000	0.000	16223747.059	1503858.305
160M k440 Inter High	0.191	0.012	0.230	0.004	0.261	0.004	0.519	0.014	0.000	0.000	1976271.166	158805.423	0.544	0.002	0.000	0.000	12395759.927	1104763.293
160M Density Tiny (842k)	0.184	0.011	0.230	0.004	0.260	0.004	0.498	0.014	0.000	0.000	1934160.402	153855.866	0.536	0.002	0.000	0.000	10354382.844	900493.008
160M Density Nano (750k)	0.193	0.012	0.230	0.004	0.260	0.004	0.504	0.014	0.000	0.000	1871303.218	150515.641	0.536	0.002	0.000	0.000	10513877.858	926264.339
160M Density Pico (250k)	0.190	0.012	0.230	0.004	0.258	0.004	0.496	0.014	0.000	0.000	1964196.926	153419.785	0.538	0.002	0.000	0.000	10720344.552	925236.704
160M Density 2 Epochs	0.189	0.012	0.230	0.004	0.257	0.004	0.501	0.014	0.000	0.000	1587737.376	121555.315	0.538	0.002	0.000	0.000	8366924.760	713077.358
160M Density Pico 2 Epochs	0.193	0.012	0.230	0.004	0.257	0.004	0.493	0.014	0.000	0.000	2017680.705	159090.061	0.541	0.002	0.000	0.000	10465698.688	903166.520

Table 2: Ablation studies of Pythia 160M models trained on MiniPile versions

	ARC-Challenge		MMLU		HellaSwag		WinoGrande		Lambada (OpenAI)		Lambada (OpenAI)		Blimp		Lambada (Std)		Lambada (Std)		ARC-Easy	
	acc	stddev	acc	stddev	acc	stddev	acc	stddev	acc	stddev	perplexity	stddev	acc	stddev	acc	stddev	perplexity	stddev	acc	stddev
1.4B Pile Dedup	0.260	0.013	0.239	0.004	0.418	0.005	0.573	0.014	0.620	0.007	6.104	0.153	0.815	0.001	0.490	0.007	11.245	0.331	0.617	0.010
1.4B MiniPile	0.190	0.012	0.230	0.004	0.258	0.004	0.519	0.014	0.000	0.000	1564928.526	118691.457	0.548	0.002	0.000	0.000	8848600.941	745031.890	0.272	0.009
1.4B Reproduction	0.193	0.012	0.230	0.004	0.258	0.004	0.509	0.014	0.000	0.000	1520707.870	115261.366	0.540	0.002	0.000	0.000	8651201.888	735161.524	0.267	0.009
1.4B Density	0.185	0.011	0.230	0.004	0.259	0.004	0.504	0.014	0.000	0.000	1420846.832	106563.133	0.542	0.002	0.000	0.000	7916035.353	664805.918	0.270	0.009

Table 3: Performance comparison of Pythia 1.4B models trained on Pile and MiniPile versions