Open LLM Leaderboard v2 (Average error across benchmarks)

	bloom	pythia	falcon	gemma-2	gpt-j-neo-neox	meta-llama-3	olmo	opt	qwen2	starcoder2	smollm	yi-1.5
FLOPs (shared intercept)	2.4	1.4	4.0	7.1	1.4	14.6	4.2	2.4	6.2	4.8	1.4	8.3
FLOPs	2.3	4.0	3.9	**3.5**	1.3	4.3	2.5	5.3	9.0	**1.5**	3.3	4.5
Size and Tokens	2.9	4.1	4.0	4.3	**0.8**	**4.1**	3.0	4.5	7.6	3.3	6.3	**3.9**
PCA + FLOPs (d=1)	4.1	8.0	**1.9**	8.2	2.9	8.1	5.0	10.2	9.6	2.2	6.7	5.3
PCA + FLOPs (d=2)	5.2	8.2	**1.9**	6.0	2.3	4.8	5.0	10.1	8.5	2.1	6.9	**3.9**
PCA + FLOPs (d=3)	5.2	8.3	2.3	5.6	2.4	4.8	5.1	10.3	8.3	2.3	7.1	4.0
PCA + FLOPs (d=4)	5.5	8.8	2.5	5.4	2.5	5.3	5.6	9.9	8.6	2.2	7.4	4.2
Sloth basic (d=1)	**1.3**	4.8	4.4	8.9	1.6	11.4	2.1	5.3	11.3	3.5	2.4	5.3
Sloth basic (d=2)	3.0	3.0	3.6	5.7	1.4	**4.1**	1.9	2.7	10.2	1.7	2.2	4.2
Sloth basic (d=3)	3.1	3.0	4.0	5.7	1.5	4.4	2.5	3.5	9.8	**1.5**	2.5	4.6
Sloth basic (d=4)	2.9	2.9	4.1	5.8	1.5	6.4	2.6	5.0	10.1	1.8	2.3	4.6
Sloth (d=1)	2.1	**1.2**	4.7	6.5	1.1	8.1	5.0	1.3	7.7	4.2	2.3	5.3
Sloth (d=2)	3.0	1.3	4.2	4.2	1.2	4.8	**0.7**	**0.9**	**5.0**	3.4	1.6	**3.9**
Sloth (d=3)	2.5	**1.2**	5.0	3.6	2.0	5.1	1.8	1.6	8.3	5.0	**1.2**	4.7
Sloth (d=4)	2.0	1.4	4.3	4.3	0.9	5.0	3.4	1.3	8.4	2.9	1.6	4.1