

Open LLM Leaderboard v2 (Average error across benchmarks)

FLOPs	10.5	5.3	20.4	13.3	11.4	23.1	1.9	4.6	16.2
Size and Tokens	1.2	3.7	4.0	2.0	5.1	7.3	1.5	3.7	7.2
PCA + FLOPs (d=1)	3.9	6.4	5.7	2.2	6.0	10.3	2.3	3.5	12.4
PCA + FLOPs (d=2)	4.8	3.9	5.8	3.0	5.6	9.6	2.0	3.5	12.2
PCA + FLOPs (d=3)	4.6	3.0	4.8	2.7	5.4	9.4	1.9	3.3	14.1
PCA + FLOPs (d=4)	5.6	4.7	5.6	4.9	6.4	12.4	2.0	4.1	13.7
Sloth basic (d=1) (shared intercept)	2.3	7.3	5.0	1.6	3.2	10.3	1.8	6.5	6.8
Sloth basic (d=2) (shared intercept)	2.6	7.4	4.8	1.7	3.3	12.1	1.8	6.8	6.8
Sloth basic (d=3) (shared intercept)	2.6	7.2	4.7	1.5	3.3	13.0	2.0	6.6	7.4
Sloth basic (d=4) (shared intercept)	2.6	7.1	4.6	1.5	3.3	13.0	2.0	6.7	7.3
Sloth (d=1) (shared intercept)	2.1	7.0	5.4	1.5	4.9	7.7	1.5	7.5	6.7
Sloth (d=2) (shared intercept)	2.5	6.9	4.5	1.5	4.5	7.4	1.4	8.8	7.3
Sloth (d=3) (shared intercept)	2.5	6.9	5.2	1.4	4.4	6.9	1.4	9.0	7.3
Sloth (d=4) (shared intercept)	2.5	6.8	5.1	1.4	4.4	6.5	1.4	8.8	7.1
Sloth basic (d=1)	1.4	5.7	5.3	1.6	8.0	6.9	2.0	3.0	6.6
Sloth basic (d=2)	1.6	3.7	4.6	2.5	6.4	5.4	1.5	2.3	5.0
Sloth basic (d=3)	1.8	3.4	4.3	2.7	5.7	12.0	2.2	1.5	4.7
Sloth basic (d=4)	1.5	3.6	4.0	2.5	5.7	12.6	1.7	2.3	4.6
Sloth (d=1)	1.8	6.0	5.0	1.2	6.7	11.4	3.5	1.8	6.1
Sloth (d=2)	1.4	4.7	4.7	1.6	4.4	11.2	1.0	2.2	6.0
Sloth (d=3)	2.3	4.2	4.2	2.0	3.9	6.4	1.9	2.6	5.0
Sloth (d=4)	1.9	3.3	2.8	1.6	4.8	7.0	1.4	2.3	4.1
	bloom	llama-2	orca_mini_v3	pythia	qwen1.5	qwen2	smollm	starcoder2	yi-1.5