

OpenLLM Leaderboard v2

FLOPs	13.9	7.4	26.6	4.8	9.8	8.6	11.9
Size and Tokens	7.1	2.2	4.6	4.2	3.9	1.8	4.0
PCA + FLOPs (d=1)	9.0	8.1	4.1	1.7	3.9	8.2	5.8
PCA + FLOPs (d=2)	11.3	6.8	4.5	1.4	3.5	6.0	5.6
PCA + FLOPs (d=3)	11.2	7.0	3.3	1.4	3.8	6.1	5.5
PCA + FLOPs (d=4)	11.2	6.9	3.2	1.3	10.6	6.5	6.6
Sloth basic (d=1) (shared intercept)	9.2	5.1	5.0	2.4	3.3	4.9	5.0
Sloth basic (d=2) (shared intercept)	9.5	5.2	6.0	2.5	3.3	5.1	5.3
Sloth basic (d=3) (shared intercept)	9.2	5.2	5.9	2.7	4.1	5.0	5.4
Sloth basic (d=4) (shared intercept)	9.2	5.2	5.9	2.8	4.0	4.8	5.3
Sloth (d=1) (shared intercept)	9.6	5.3	4.3	2.4	3.3	4.7	4.9
Sloth (d=2) (shared intercept)	9.3	5.1	4.5	2.6	3.4	4.9	5.0
Sloth (d=3) (shared intercept)	9.5	5.1	4.5	2.5	3.6	4.9	5.0
Sloth (d=4) (shared intercept)	9.4	4.8	4.2	2.5	3.6	4.8	4.9
Sloth basic (d=1)	8.3	5.7	3.3	1.4	3.2	5.0	4.5
Sloth basic (d=2)	4.9	4.8	3.2	1.4	3.2	4.5	3.7
Sloth basic (d=3)	4.8	4.6	6.4	2.0	3.6	4.2	4.3
Sloth basic (d=4)	4.7	3.9	6.8	2.5	3.1	4.7	4.3
Sloth (d=1)	8.1	4.4	7.9	1.5	2.8	4.3	4.8
Sloth (d=2)	5.1	4.3	8.2	2.0	2.8	2.3	4.1
Sloth (d=3)	6.7	2.4	3.8	1.7	5.4	1.8	3.6
Sloth (d=4)	5.8	2.1	5.1	1.6	3.3	1.6	3.2
	IFEval	BBH	MATH Lvl 5	GPQA	MUSR	MMLU-PRO	Average