

OpenLLM Leaderboard v2

FLOPs (shared intercept)	9.4	4.5	4.2	2.6	2.8	5.6	4.9
FLOPs	5.1	3.6	2.6	2.5	4.4	4.5	3.8
Size and Tokens	4.9	3.8	3.3	3.0	2.8	6.8	4.1
PCA + FLOPs (d=1)	9.3	7.3	4.4	2.7	3.2	9.4	6.0
PCA + FLOPs (d=2)	6.1	7.4	4.3	2.7	2.8	9.3	5.4
PCA + FLOPs (d=3)	6.1	7.1	5.3	2.7	2.6	9.1	5.5
PCA + FLOPs (d=4)	6.1	7.1	5.5	2.7	3.7	8.9	5.7
Sloth basic (d=1) (shared intercept)	10.5	5.1	5.2	2.8	2.8	6.2	5.4
Sloth basic (d=2) (shared intercept)	11.3	5.0	5.5	3.0	3.0	5.6	5.6
Sloth basic (d=3) (shared intercept)	11.2	5.1	5.6	3.1	2.9	5.5	5.6
Sloth basic (d=4) (shared intercept)	11.3	5.2	5.6	3.0	3.0	5.4	5.6
Sloth (d=1) (shared intercept)	10.0	4.9	3.4	2.5	3.9	5.6	5.0
Sloth (d=2) (shared intercept)	10.2	4.2	3.0	2.3	2.4	5.0	4.5
Sloth (d=3) (shared intercept)	10.2	3.4	3.1	2.1	3.3	5.3	4.6
Sloth (d=4) (shared intercept)	10.1	3.4	3.5	2.2	3.6	5.2	4.7
Sloth basic (d=1)	9.2	6.0	3.8	2.3	2.3	7.5	5.2
Sloth basic (d=2)	4.8	4.1	3.2	2.2	1.9	5.6	3.6
Sloth basic (d=3)	4.8	4.3	2.9	2.2	3.0	5.7	3.8
Sloth basic (d=4)	4.8	5.0	3.8	2.2	3.3	5.9	4.2
Sloth (d=1)	9.9	4.2	2.6	1.9	2.0	4.2	4.1
Sloth (d=2)	4.9	2.6	2.3	2.5	2.2	2.6	2.9
Sloth (d=3)	4.8	2.5	4.1	2.4	4.0	3.1	3.5
Sloth (d=4)	4.8	2.4	4.3	2.0	2.9	3.5	3.3
	IFEval	BBH	MATH Lvl 5	GPQA	MUSR	MMLU-PRO	Average