Open LLM Leaderboard v1 (Average error across benchmarks)

	bloom	codellama	deepseek-coder-base	falcon	gpt-j-neo-neox	llama-2	open_llama_	opt	pythia	qwen1.5	qwen2	rwkv-4-pile	starcoder2	xglm	yi-1.5
FLOPs	6.5	13.3	10.1	20.3	2.9	8.8	4.5	11.9	8.9	10.6	18.5	7.9	7.9	2.1	6.5
Size and Tokens	5.2	6.9	11.9	7.2	2.0	4.2	3.5	5.7	3.1	9.0	6.0	1.4	3.8	1.3	1.8
PCA + FLOPs (d=1)	5.2	9.0	6.4	17.2	5.5	4.4	9.3	6.9	4.7	7.2	46.1	4.9	8.4	2.3	10.9
PCA + FLOPs (d=2)	4.5	9.3	5.1	18.5	4.4	5.5	4.0	7.0	4.3	6.9	45.4	3.5	8.6	2.1	9.0
PCA + FLOPs (d=3)	4.7	9.2	4.8	20.5	3.0	5.1	4.5	6.4	4.3	9.0	43.8	2.4	8.4	2.5	8.8
PCA + FLOPs (d=4)	5.0	11.9	4.8	18.8	2.9	8.7	4.4	6.5	4.6	9.1	46.0	2.7	8.3	2.5	9.2
Sloth basic (d=1) (shared intercept)	6.2	10.9	14.5	11.5	3.6	5.6	6.1	5.5	3.4	8.1	5.4	3.7	13.2	3.3	7.4
Sloth basic (d=2) (shared intercept)	4.3	11.1	13.5	12.5	2.3	4.7	4.2	3.0	1.9	7.8	3.9	2.2	14.0	4.3	6.5
Sloth basic (d=3) (shared intercept)	4.8	11.6	14.3	13.4	2.2	4.4	4.3	3.0	1.7	7.4	5.3	2.1	13.7	4.1	5.7
Sloth basic (d=4) (shared intercept)	4.9	11.6	14.3	12.9	2.2	4.4	4.3	3.0	1.7	7.4	4.7	2.1	13.8	4.1	5.8
Sloth (d=1) (shared intercept)	4.4	12.2	19.3	12.4	3.3	7.6	7.3	6.0	2.4	9.1	8.4	2.8	12.5	4.2	6.1
Sloth (d=2) (shared intercept)	3.5	12.2	15.5	12.2	2.6	4.1	4.8	2.9	2.4	7.3	6.8	2.2	15.6	4.7	6.3
Sloth (d=3) (shared intercept)	3.7	12.1	16.0	12.1	2.5	4.3	4.2	2.9	1.7	7.4	6.6	1.7	15.2	4.5	6.5
Sloth (d=4) (shared intercept)	4.2	12.2	16.2	12.2	2.2	4.3	4.1	2.6	1.7	7.3	6.7	1.8	15.4	4.6	6.6
Sloth basic (d=1)	4.5	12.9	8.4	8.5	4.2	5.6	5.1	5.1	6.3	21.2	23.9	5.7	17.6	3.6	5.6
Sloth basic (d=2)	3.9	9.8	6.0	7.0	3.5	6.6	4.3	3.8	7.7	10.0	4.5	5.4	7.3	2.5	4.5
Sloth basic (d=3)	3.8	8.0	4.1	9.3	3.1	5.7	4.3	3.2	6.6	10.3	6.7	4.5	6.8	2.4	3.3
Sloth basic (d=4)	3.6	8.2	4.9	5.7	2.6	4.9	4.0	3.8	6.2	10.2	5.8	4.6	7.0	2.5	3.5
Sloth (d=1)	3.7	5.9	8.9	9.5	2.3	7.5	5.8	4.0	7.7	20.8	6.2	2.7	14.9	2.8	3.1
Sloth (d=2)	2.8	4.7	5.9	5.2	2.8	6.6	3.1	2.9	3.0	7.9	5.9	2.8	4.2	1.8	2.1
Sloth (d=3)	2.6	6.5	11.4	3.9	2.2	2.8	4.3	2.7	3.2	8.3	4.8	2.3	3.9	1.4	2.6
Sloth (d=4)	2.5	7.0	11.5	3.6	1.5	4.1	3.7	2.2	2.9	8.1	5.9	2.5	5.0	1.4	3.2