Open LLM Leaderboard v1/v2 (Average error across benchmarks)

	bloom	pythia	falcon	gemma	gpt-j-neo-neox	meta-llama-3	olmo	opt	qwen2	starcoder2	yi-1.5
FLOPs (shared intercept)	2.2	3.0	4.8	4.9	1.4	14.4	6.9	4.2	7.0	9.2	9.2
FLOPs	2.4	5.3	6.7	6.5	1.2	3.9	**1.5**	5.1	6.9	**2.4**	**3.4**
Size and Tokens	**2.0**	4.0	6.8	5.0	1.6	4.0	3.3	3.4	6.3	5.3	**3.4**
PCA + FLOPs (d=1)	5.2	8.8	6.1	6.5	3.1	7.2	5.7	9.1	8.6	4.7	4.3
PCA + FLOPs (d=2)	4.2	8.1	5.4	6.2	2.8	8.0	5.0	9.2	9.1	2.8	4.0
PCA + FLOPs (d=3)	4.1	8.2	4.5	6.0	2.0	6.3	5.0	9.4	9.2	2.8	3.6
PCA + FLOPs (d=4)	5.1	9.5	4.8	6.1	2.5	5.2	4.3	9.3	9.2	3.0	3.8
Sloth basic (d=1) (shared intercept)	2.3	2.9	5.1	4.9	1.7	17.2	6.7	4.1	6.9	9.1	9.2
Sloth basic (d=2) (shared intercept)	2.5	2.5	3.1	5.1	1.9	18.0	7.0	4.4	5.5	8.9	8.9
Sloth basic (d=3) (shared intercept)	2.4	2.6	**2.5**	5.1	1.6	19.9	6.9	4.5	5.2	8.8	8.8
Sloth basic (d=4) (shared intercept)	2.4	2.6	**2.5**	5.1	1.5	20.5	6.9	4.5	5.3	8.8	8.9
Sloth (d=1) (shared intercept)	2.6	2.4	7.7	5.9	2.2	8.3	7.7	3.0	6.2	9.8	7.2
Sloth (d=2) (shared intercept)	2.5	2.0	5.3	6.7	1.3	15.2	6.4	2.1	5.6	12.7	6.4
Sloth (d=3) (shared intercept)	2.6	2.3	4.2	5.9	**1.1**	6.8	6.9	2.3	6.6	12.7	7.0
Sloth (d=4) (shared intercept)	2.6	2.4	3.9	6.2	**1.1**	7.5	6.8	2.4	5.7	11.2	7.1
Sloth basic (d=1)	3.9	4.6	5.6	7.9	2.5	5.3	3.2	4.6	23.5	10.0	4.2
Sloth basic (d=2)	2.9	4.8	7.2	7.3	1.8	7.1	1.7	5.6	8.6	2.9	4.2
Sloth basic (d=3)	3.0	4.3	7.3	7.2	1.6	4.8	1.9	4.8	8.6	2.9	3.6
Sloth basic (d=4)	3.9	5.7	7.7	6.4	1.9	6.0	1.7	6.1	7.8	2.8	3.7
Sloth (d=1)	3.7	3.7	6.2	6.8	1.2	5.4	6.1	3.7	9.0	8.2	4.0
Sloth (d=2)	4.5	**1.7**	8.2	6.0	1.2	5.7	2.3	1.5	6.8	4.6	3.5
Sloth (d=3)	2.1	3.3	7.6	5.0	1.3	3.7	2.4	1.4	**4.3**	3.7	3.7
Sloth (d=4)	5.6	4.8	6.8	**4.3**	2.2	**3.5**	2.1	**1.1**	5.5	5.7	**3.4**