

Open LLMLeaderboard v1/v2

|                   |      |      |           |            |            |       |        |     |            |      |      |          |         |
|-------------------|------|------|-----------|------------|------------|-------|--------|-----|------------|------|------|----------|---------|
| FLOPs             | 8.4  | 7.1  | 4.7       | 4.9        | 10.0       | 27.3  | 15.2   | 8.1 | 24.2       | 5.8  | 8.6  | 9.0      | 11.1    |
| Size and Tokens   | 3.2  | 3.1  | 3.3       | 3.2        | 7.1        | 12.2  | 5.8    | 4.7 | 8.6        | 4.3  | 3.4  | 3.9      | 5.2     |
| PCA + FLOPs (d=1) | 14.5 | 10.1 | 12.5      | 7.6        | 6.5        | 19.8  | 12.1   | 7.1 | 4.0        | 1.9  | 2.2  | 7.5      | 8.8     |
| PCA + FLOPs (d=2) | 15.1 | 8.6  | 10.3      | 5.6        | 6.2        | 19.7  | 10.8   | 7.2 | 4.1        | 2.0  | 2.2  | 7.6      | 8.3     |
| PCA + FLOPs (d=3) | 14.7 | 9.6  | 12.8      | 7.0        | 6.7        | 20.1  | 7.9    | 6.8 | 3.5        | 1.8  | 3.0  | 7.2      | 8.4     |
| PCA + FLOPs (d=4) | 13.6 | 9.8  | 12.7      | 6.9        | 7.1        | 20.6  | 10.3   | 7.2 | 3.8        | 2.0  | 3.1  | 6.3      | 8.6     |
| Sloth basic (d=1) | 7.8  | 4.9  | 6.9       | 4.6        | 5.4        | 17.9  | 8.3    | 6.8 | 5.7        | 1.8  | 2.8  | 7.6      | 6.7     |
| Sloth basic (d=2) | 5.4  | 5.8  | 7.8       | 4.3        | 5.8        | 12.7  | 8.7    | 5.0 | 4.4        | 1.5  | 2.2  | 4.9      | 5.7     |
| Sloth basic (d=3) | 4.9  | 5.4  | 7.3       | 3.8        | 6.3        | 12.0  | 6.1    | 4.9 | 4.9        | 1.5  | 2.6  | 4.6      | 5.4     |
| Sloth basic (d=4) | 3.0  | 5.8  | 5.8       | 3.2        | 6.8        | 10.7  | 4.2    | 5.2 | 6.5        | 1.8  | 2.9  | 4.3      | 5.0     |
| Sloth (d=1)       | 3.1  | 6.7  | 9.7       | 6.4        | 4.7        | 16.7  | 8.6    | 4.8 | 6.4        | 1.9  | 2.5  | 3.9      | 6.3     |
| Sloth (d=2)       | 5.9  | 3.1  | 3.0       | 2.1        | 4.9        | 14.9  | 8.2    | 4.1 | 3.6        | 1.9  | 1.9  | 4.0      | 4.8     |
| Sloth (d=3)       | 3.1  | 4.8  | 3.4       | 3.2        | 5.6        | 8.0   | 7.0    | 3.5 | 2.9        | 2.1  | 2.2  | 2.3      | 4.0     |
| Sloth (d=4)       | 2.3  | 2.7  | 2.4       | 1.9        | 6.2        | 9.5   | 5.5    | 4.6 | 4.1        | 2.3  | 3.6  | 2.5      | 4.0     |
|                   | MMLU | ARC  | HellaSwag | Winogrande | TruthfulQA | GSM8K | IFEval | BBH | MATH Lvl 5 | GPQA | MUSR | MMLU-PRO | Average |