

# GSM8K Finetuning Final Accuracy (%)

## 300M Models, All Configurations

Base Model → Finetune Optimizer

ADAMW 300M 8  
→ ADAMW

2.35

1.60

3.14

4.88

ADAMW 300M 8  
→ MUON

2.36

2.36

2.36

2.35

MUON 300M 8  
→ ADAMW

2.94

1.95

7.62

13.48

MUON 300M 8  
→ MUON

2.34

2.34

2.34

2.26

LR=5e-6

LR=1e-5

LR=3e-5

LR=1e-4

Learning Rate

Accuracy (%)

14

12

10

8

6

4

2

0