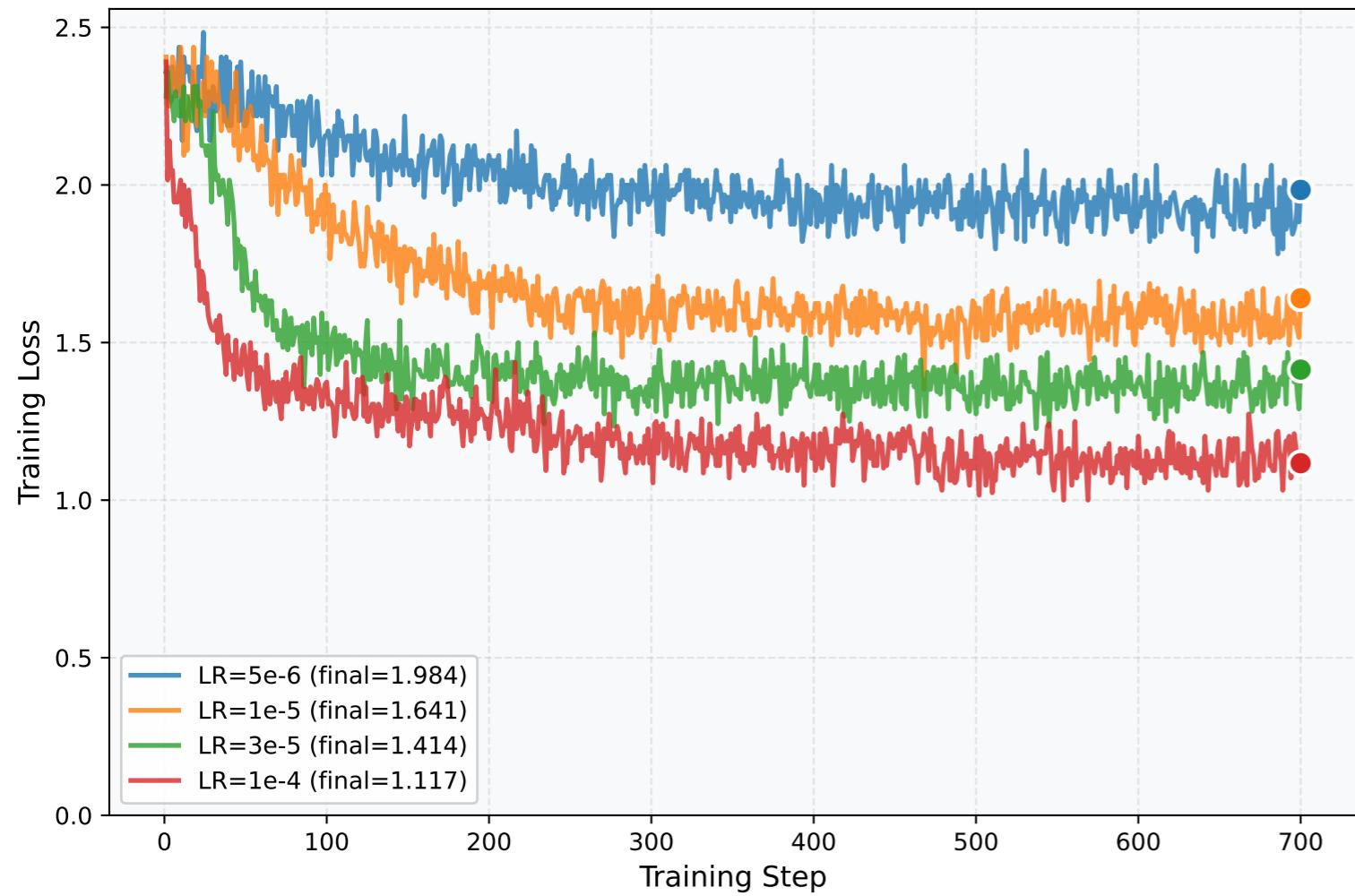


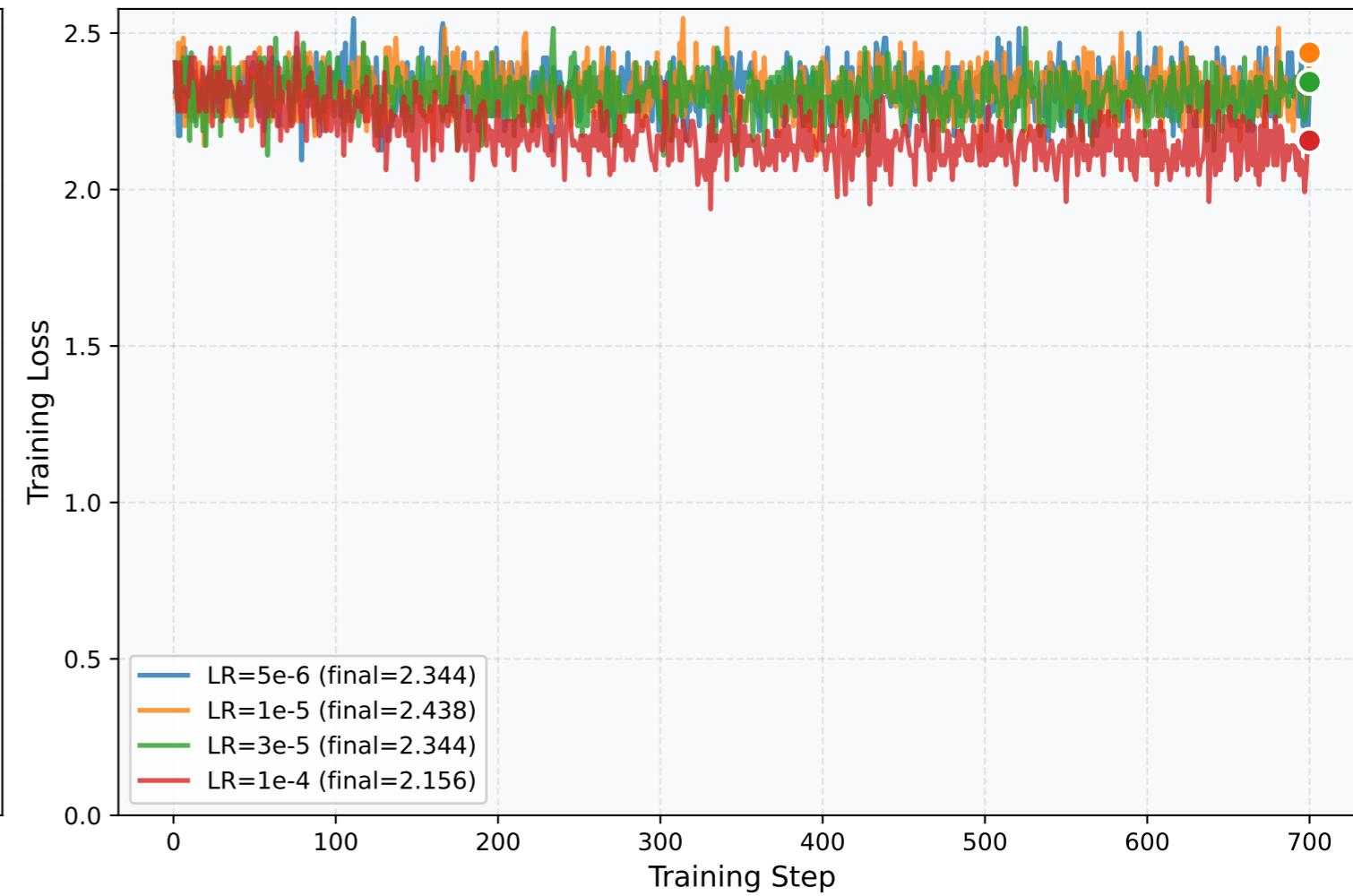
# GSM8K Finetuning: 300M Models

## Training Loss Curves

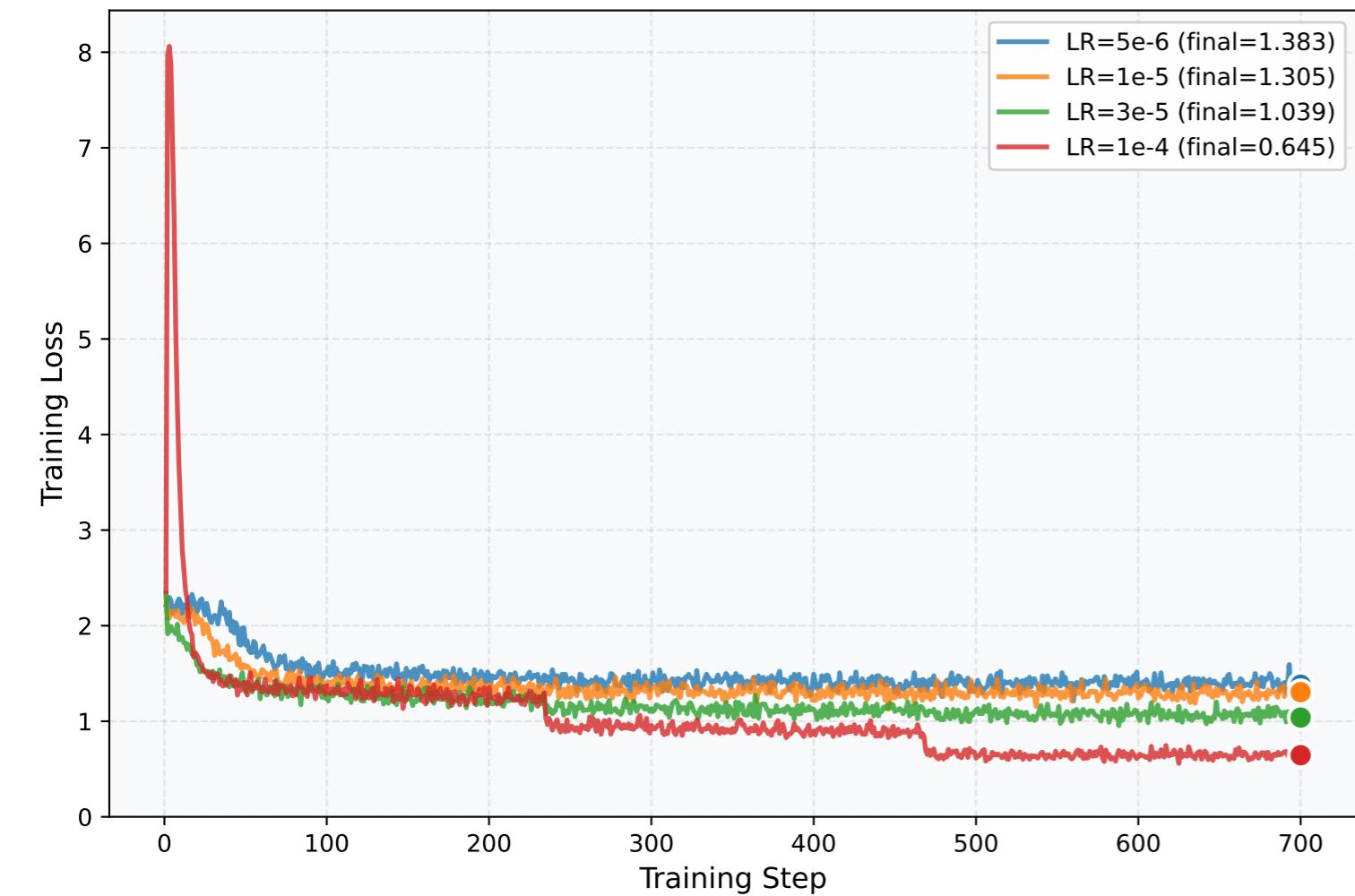
**Adamw 300M 8 → Finetune with ADAMW**



**Adamw 300M 8 → Finetune with MUON**



**Muon 300M 8 → Finetune with ADAMW**



**Muon 300M 8 → Finetune with MUON**

