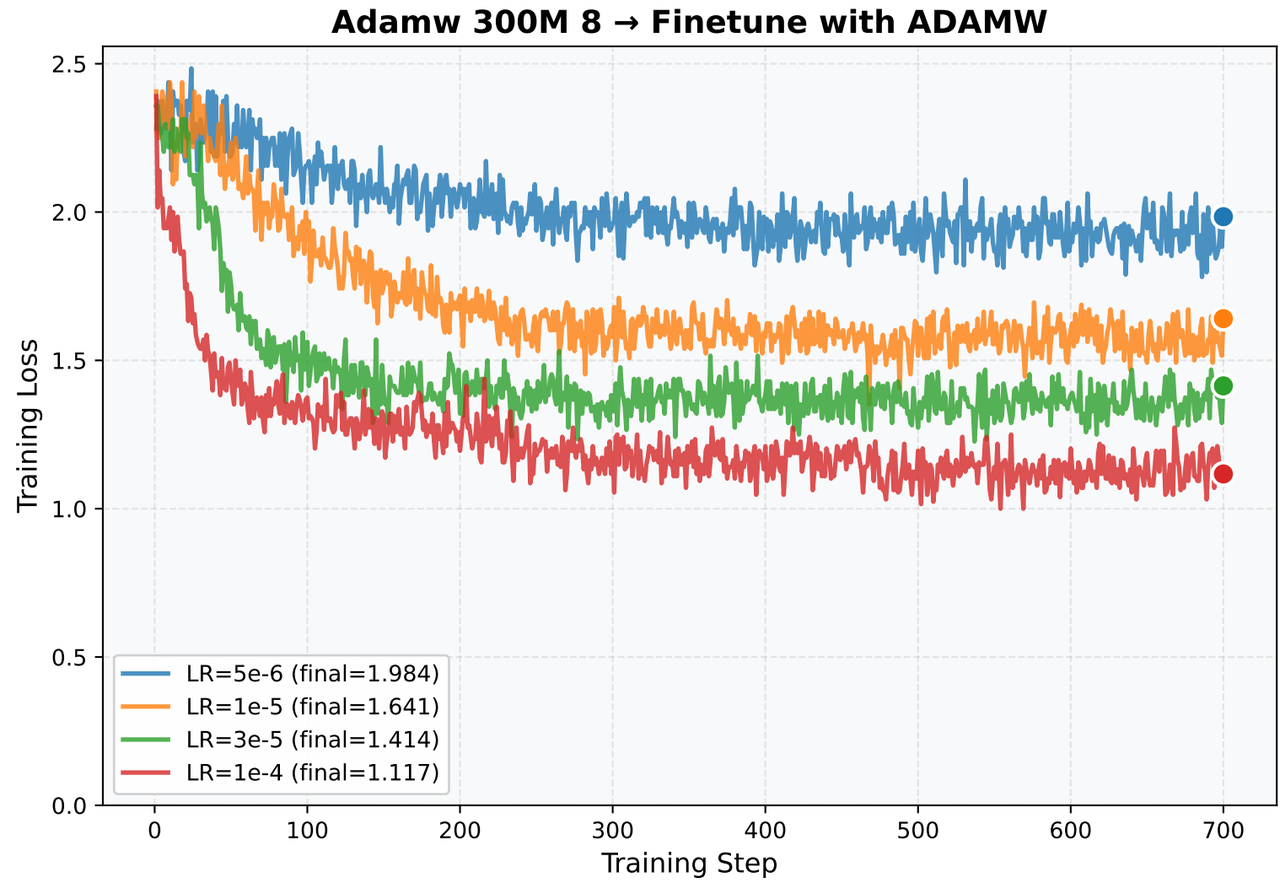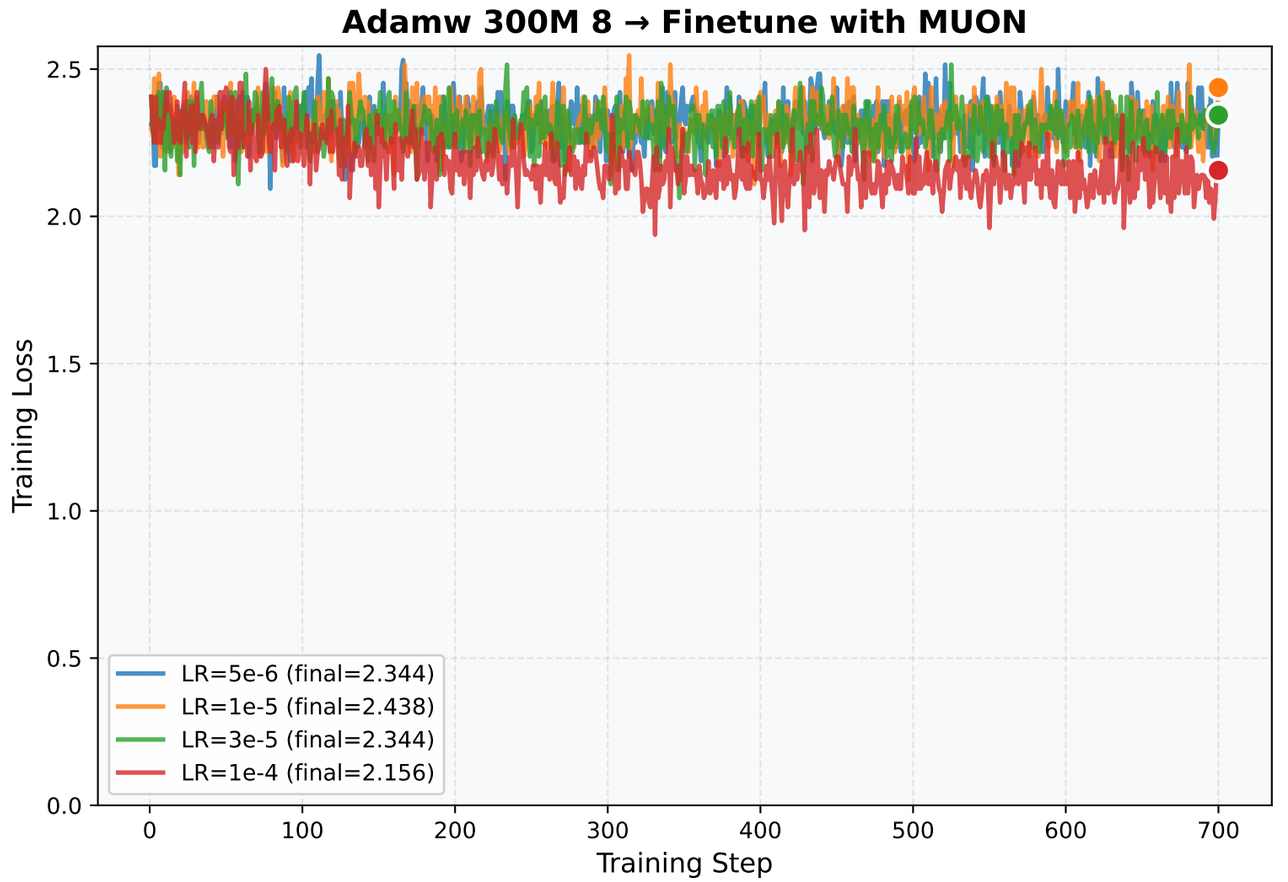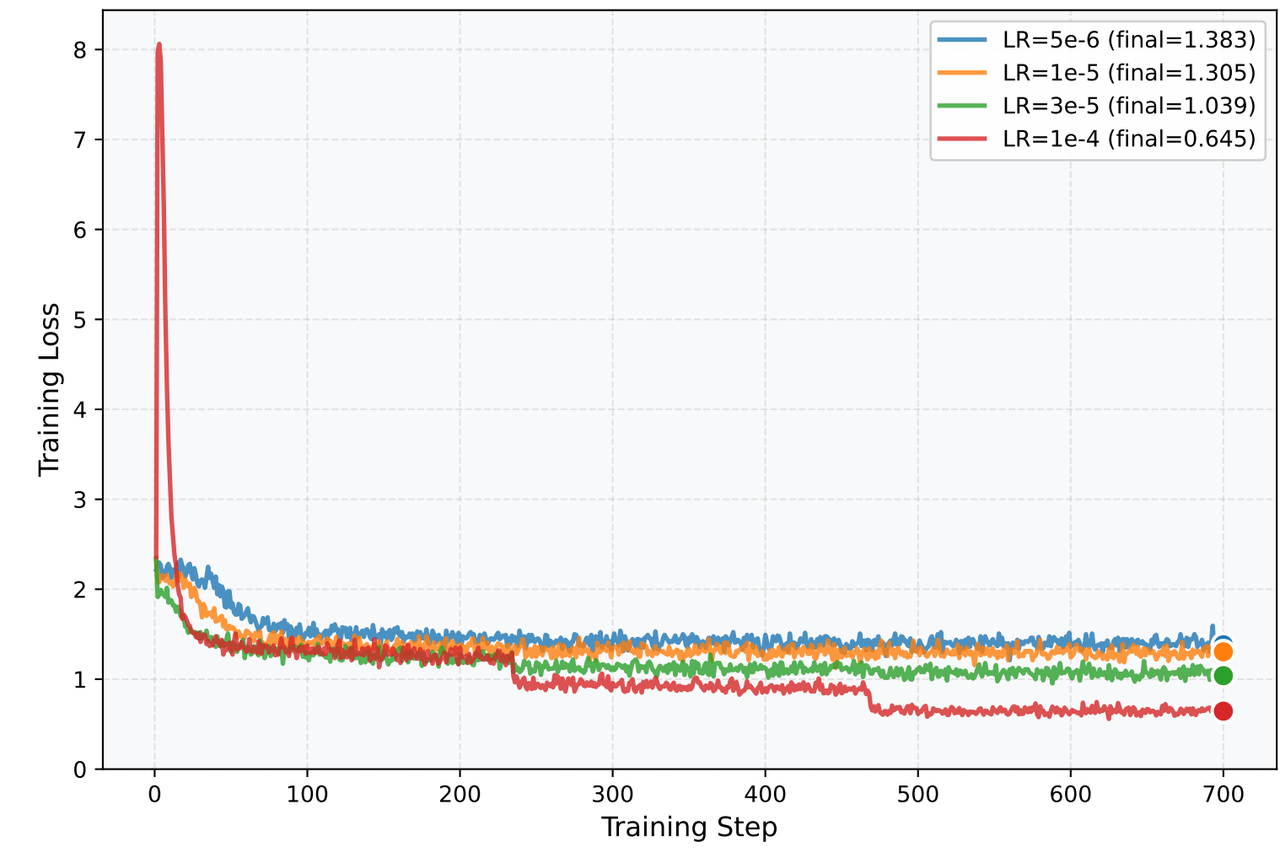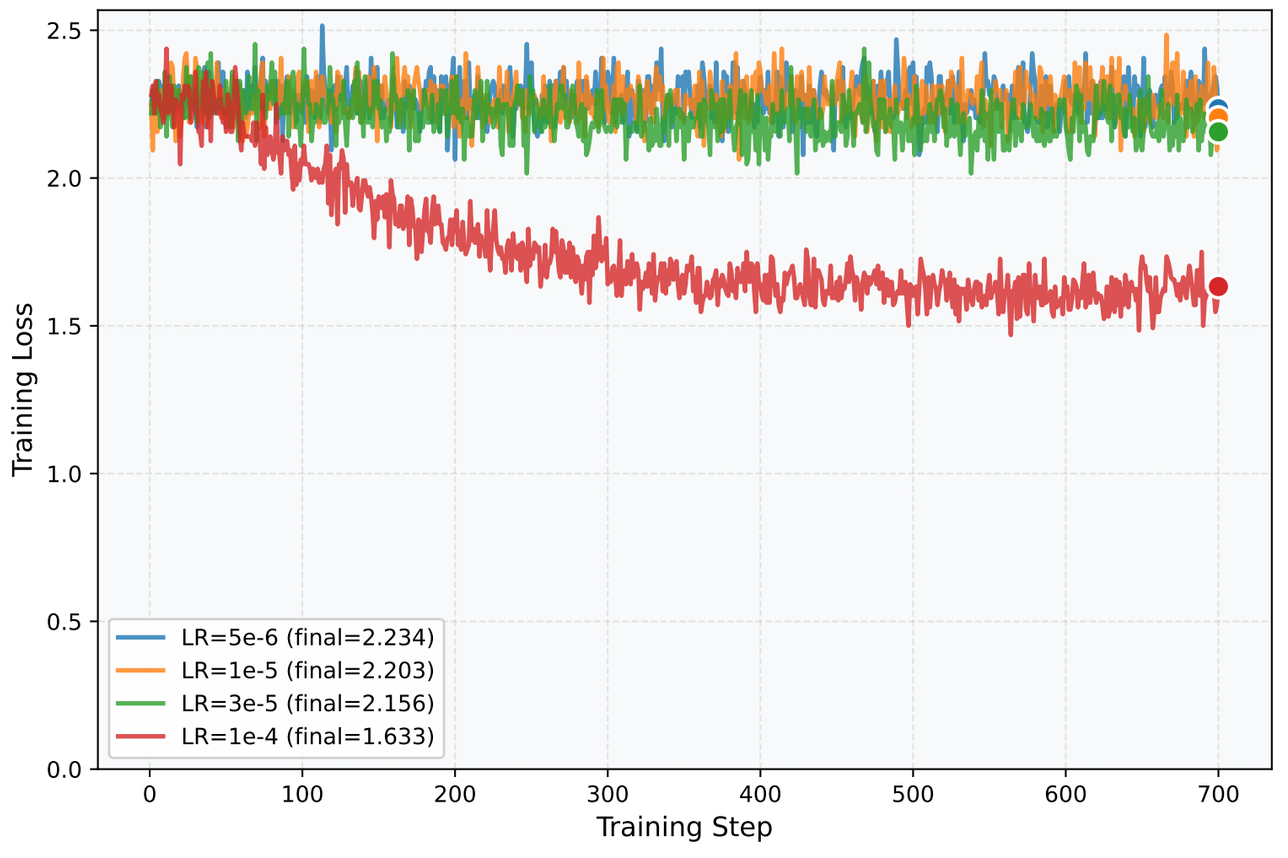GSM8K Finetuning: 300M Models
Training Loss Curves