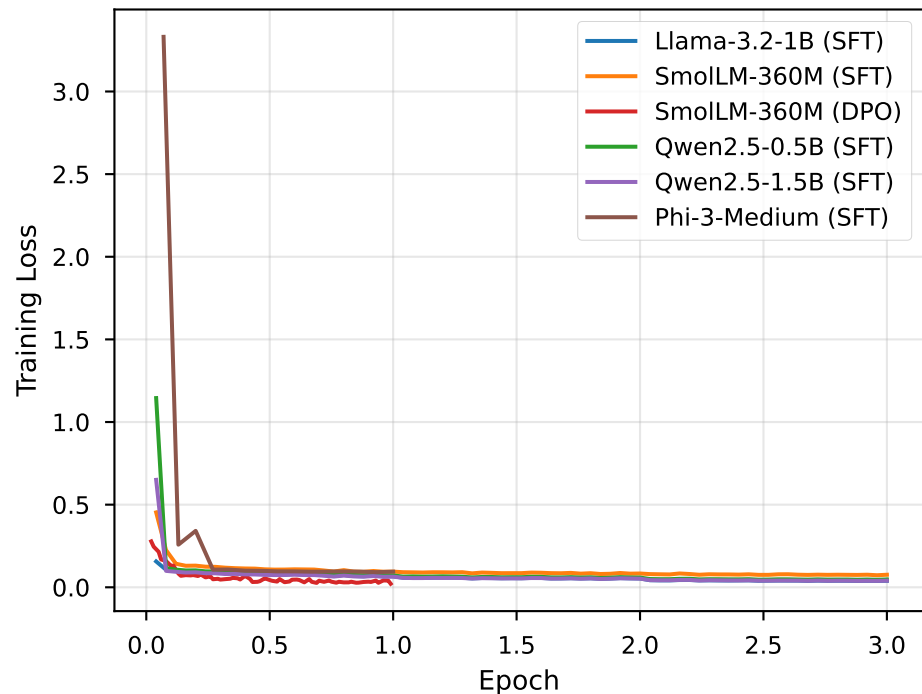
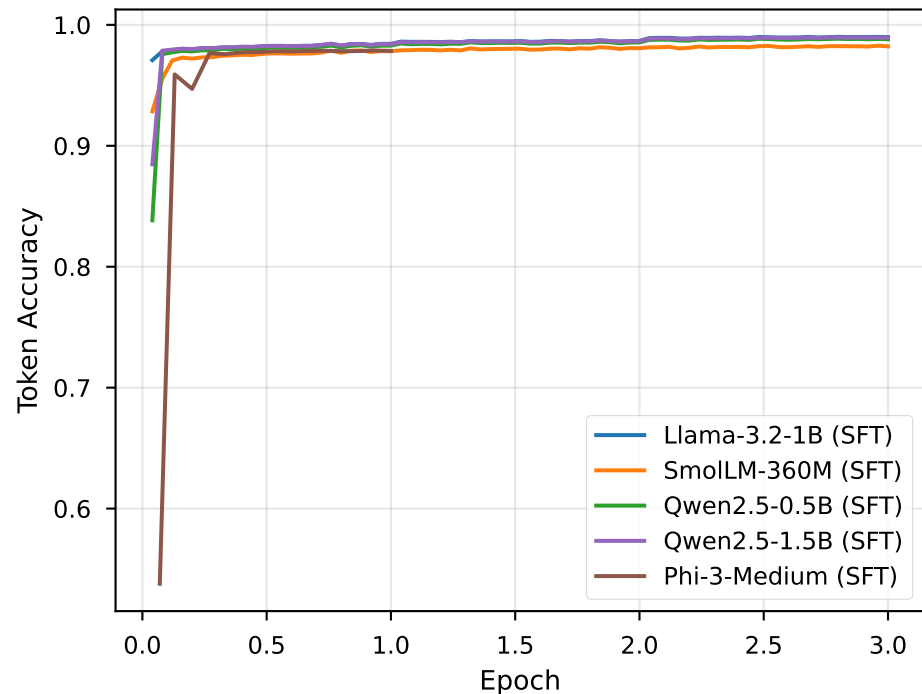


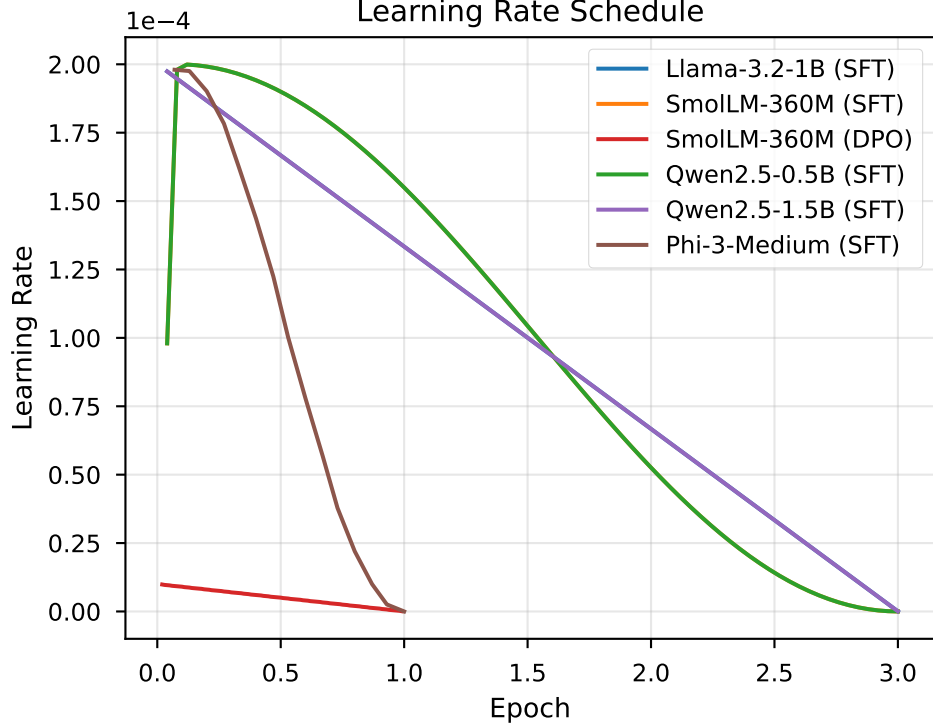
Training Loss vs Epoch



Token Accuracy vs Epoch



Learning Rate Schedule



Entropy vs Epoch

