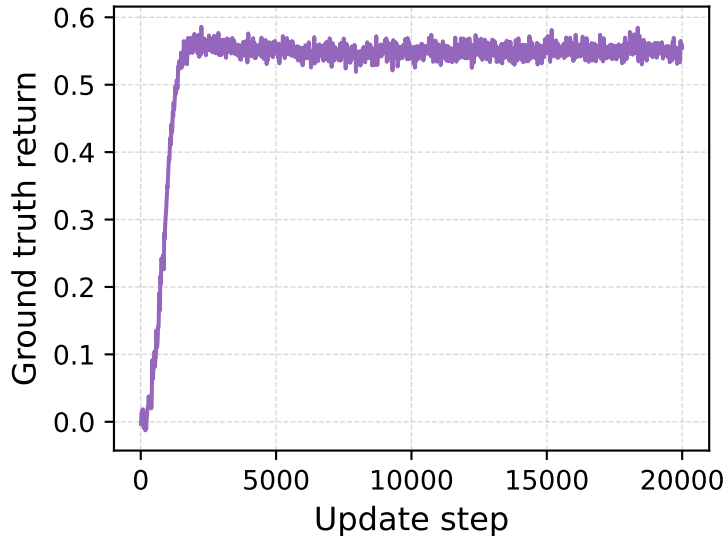


Learning curves at 10% oversight



Gradient-routed MoE