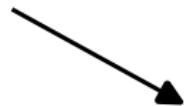
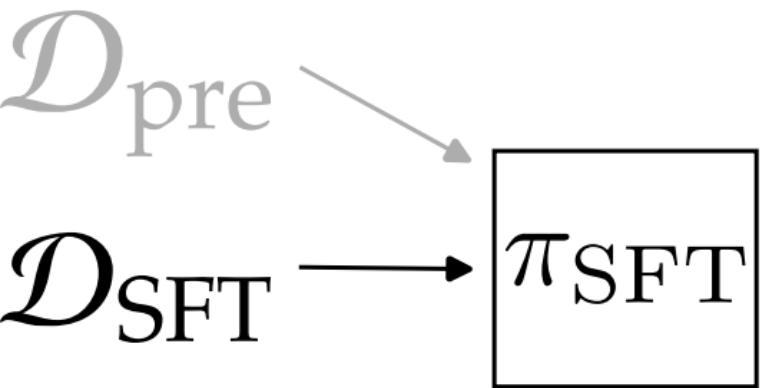
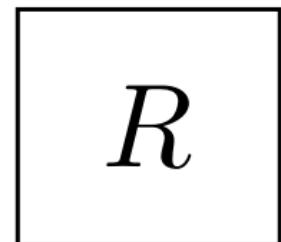
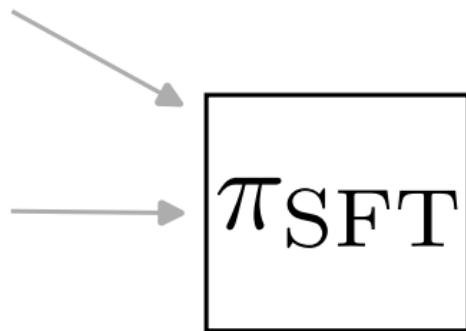


\mathcal{D}_{pre}  π_{pre}



\mathcal{D}_{pre} \mathcal{D}_{SFT} 

generation y

\mathcal{D}_{pre}

\mathcal{D}_{SFT}

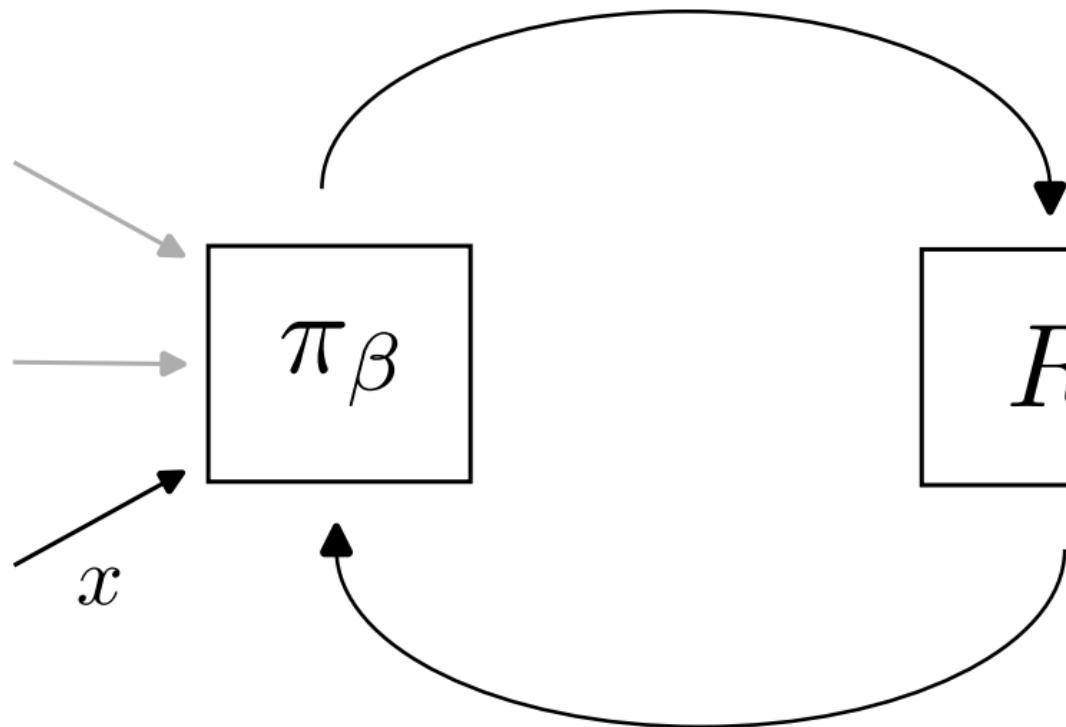
\mathcal{D}_{RL}

π_β

R

x

score $R(x, y)$



generation y

\mathcal{D}_{pre}

\mathcal{D}_{SFT}

\mathcal{D}_{RL}

π_β

R

x

LoRA

LoRA

score $R(x, y)$

