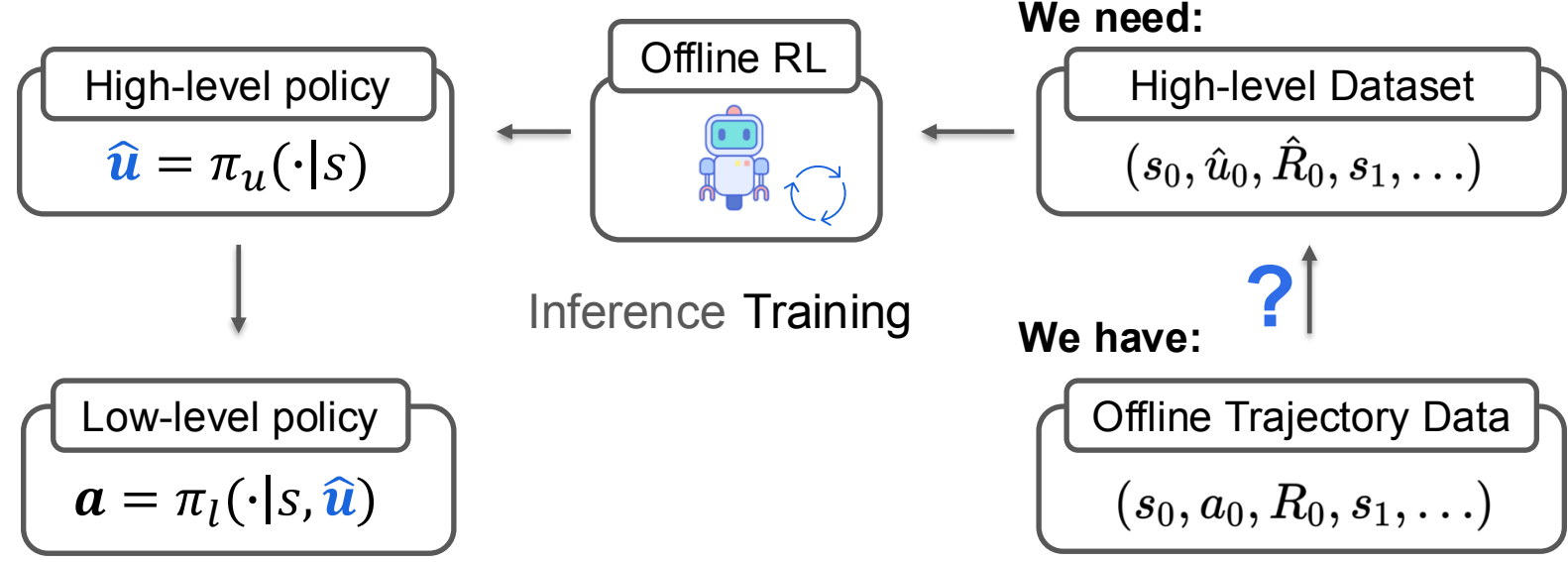
High-level policy
$\hat{u} = \pi_u(\cdot|s)$
Offline RL
Inference
Training
We need:
High-level Dataset
$(s_0, \hat{u}_0, \hat{R}_0, s_1, \ldots)$
?
We have:
Low-level policy
$a = \pi_l(\cdot|s, \hat{u})$
Offline Trajectory Data
$(s_0, a_0, R_0, s_1, \ldots)$