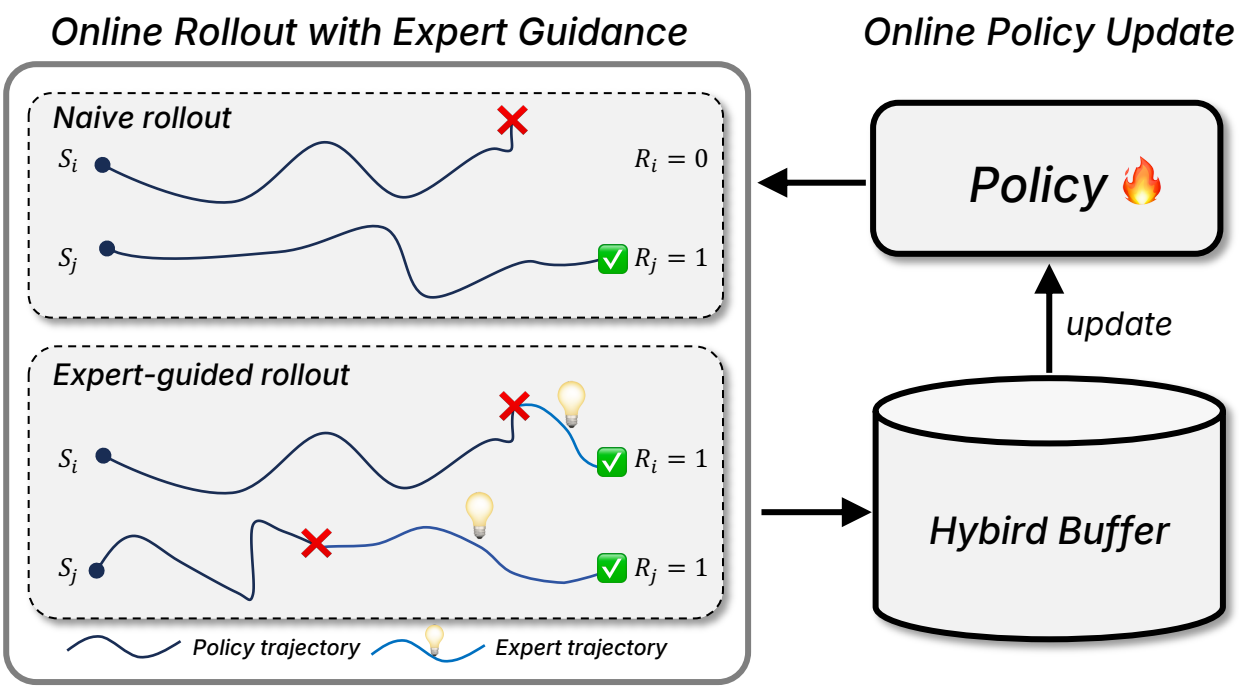
Online Rollout with Expert Guidance
Online Policy Update
Naive rollout
S_i
$R_i = 0$
S_j
$R_j = 1$
Expert-guided rollout
S_i
$R_i = 1$
S_j
$R_j = 1$
Policy trajectory
Expert trajectory
Policy
update
Hybird Buffer