

Demonstration Datasets

Training set



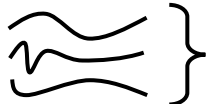
Newly collected



\mathcal{L}_{bc}

Store BC policy rollouts

π_θ



Train & Evaluate Policy

Performance Influence Estimation

Any demo

τ_1
 τ_2
 τ_3

output

Performance scores

$\psi(\tau_1)$

$\psi(\tau_2)$

$\psi(\tau_3)$

Top- k
demos

input

Estimate
Performance

$\hat{J}(\pi_\theta)$

θ

$\nabla_{\tau_1} \hat{J}$

$J(\pi)$

$\nabla_{\tau_3} \hat{J}$

$\nabla_{\tau_2} \hat{J}$



CUPID

Dataset Curation & Retraining

Curated set



Retrain & evaluate

π_θ



Original policy



After curation

