

# Offline RL x Imitation Learning

Use imitation learning to guide and kickstart rl.

- ▶ Two papers (offline: [Wu et al. 2019]) (online: [Huang et al., 2021])
  - ▶ *Both* use a behavior policy to **guide learning**
  - ▶ *first* (offline) uses policy derived from dataset
  - ▶ uses kernel MMD, KL-divergence or Wasserstein-distance as policy distance metric
  - ▶ *second* (online) utilizes policy learned by behavioral cloning from experts
  - ▶ guides rl algorithm via KL-divergence or reward penalty
- ▶ Less exploration needed → smaller dataset possible
- ▶ offline → helps algorithm to stay in distribution
- more useful in model-based learning