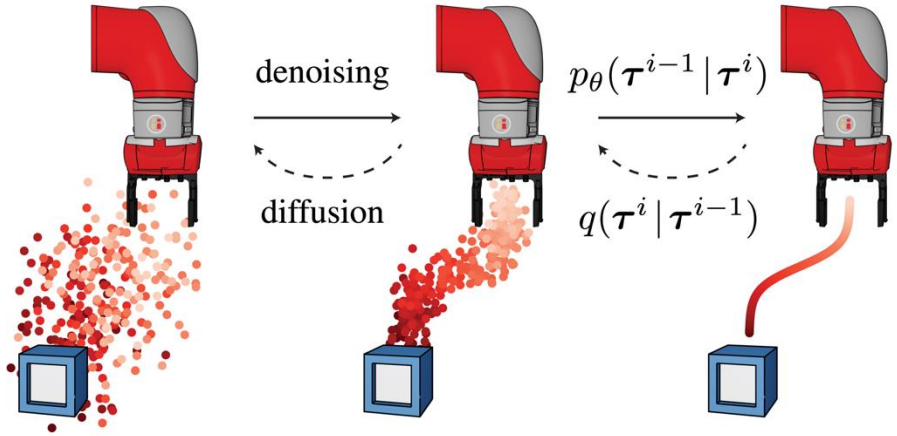
COMP 690 · DEEP GENERATIVE MODELS · ROBOTICS I

Planning with Diffusion for Flexible Behavior Synthesis

Michael Janner*, Yilun Du*, Joshua B. Tenenbaum, Sergey Levine

ICML 2022

Presented by Zhen Xu · November 17, 2026



Figures in this deck are from Janner et al. (2022) and diffusion-planning.github.io.

THE PROBLEM

A simple recipe: great in principle, a headache in practice

Algorithm 1 Model-based RL (idealized)

Inputs: Dataset of transitions $\mathcal{D} = \{(\mathbf{s}_t, \mathbf{a}_t, \mathbf{s}_{t+1}), \ldots\}$, reward function $r(\cdot, \cdot)$, current state $\mathbf{s}_0$

1: Train a predictive model

supervised learning

$$\underset{f}{\text{minimize}}\ \mathbb{E}_{\mathbf{s}_t, \mathbf{a}_t, \mathbf{s}_{t+1} \sim \mathcal{D}} \left[\| \mathbf{s}_{t+1} - f(\mathbf{s}_t, \mathbf{a}_t) \| \right]$$

2: Use model to evaluate potential plans $\mathbf{a}_{0:T}$, selecting the best one:

trajectory optimization

$$\underset{\mathbf{a}_{0:T}}{\text{maximize}}\ r(\mathbf{s}_0, \mathbf{a}_0) + r(f(\mathbf{s}_0, \mathbf{a}_0), \mathbf{a}_1) + r(f(f(\mathbf{s}_0, \mathbf{a}_0), \mathbf{a}_1), \mathbf{a}_2) + \ldots$$

Long-horizon predictions are unreliable

The optimizer finds adversarial plans

Diagram: Janner & Du, ICML 2022 talk slides.

THE CORE IDEA

Is the model really the bottleneck?

Image generators work remarkably well. So replace prediction and planning with one big generative model.

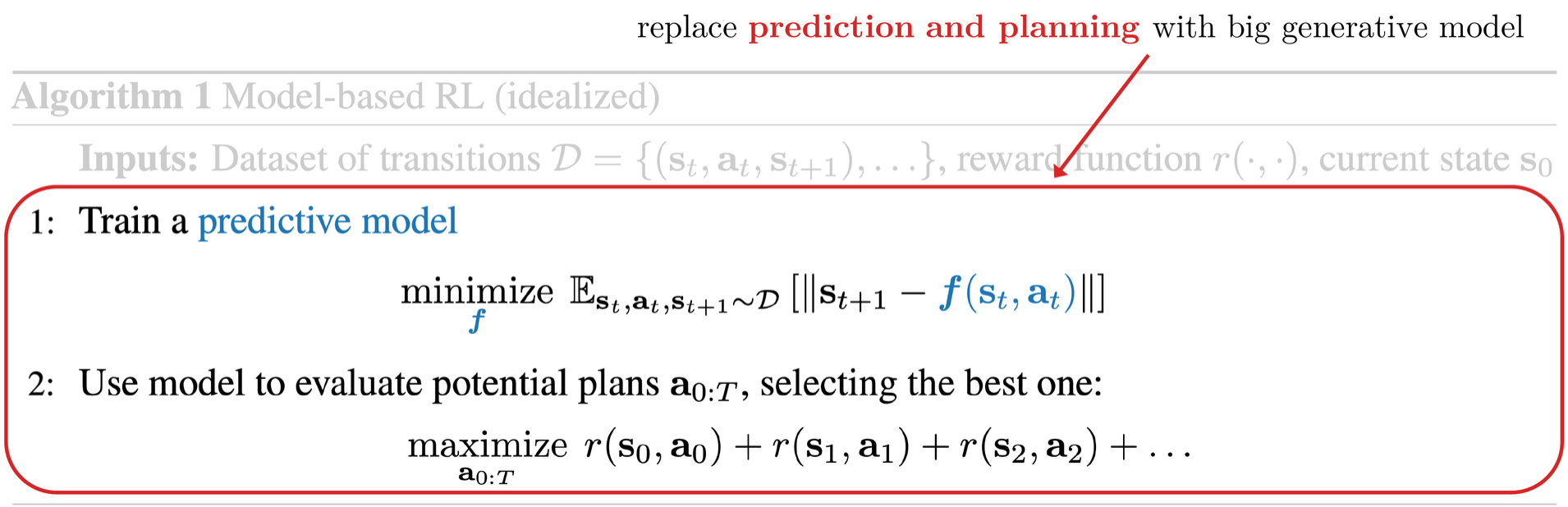


Diagram: Janner & Du, ICML 2022 talk slides.

THE MODEL · 1

A trajectory is a single-channel image

$$\tau = \begin{bmatrix} s_0 & s_1 & \cdots & s_T \\ a_0 & a_1 & \cdots & a_T \end{bmatrix}$$

Columns = time

states and actions stacked

Whole plan at once

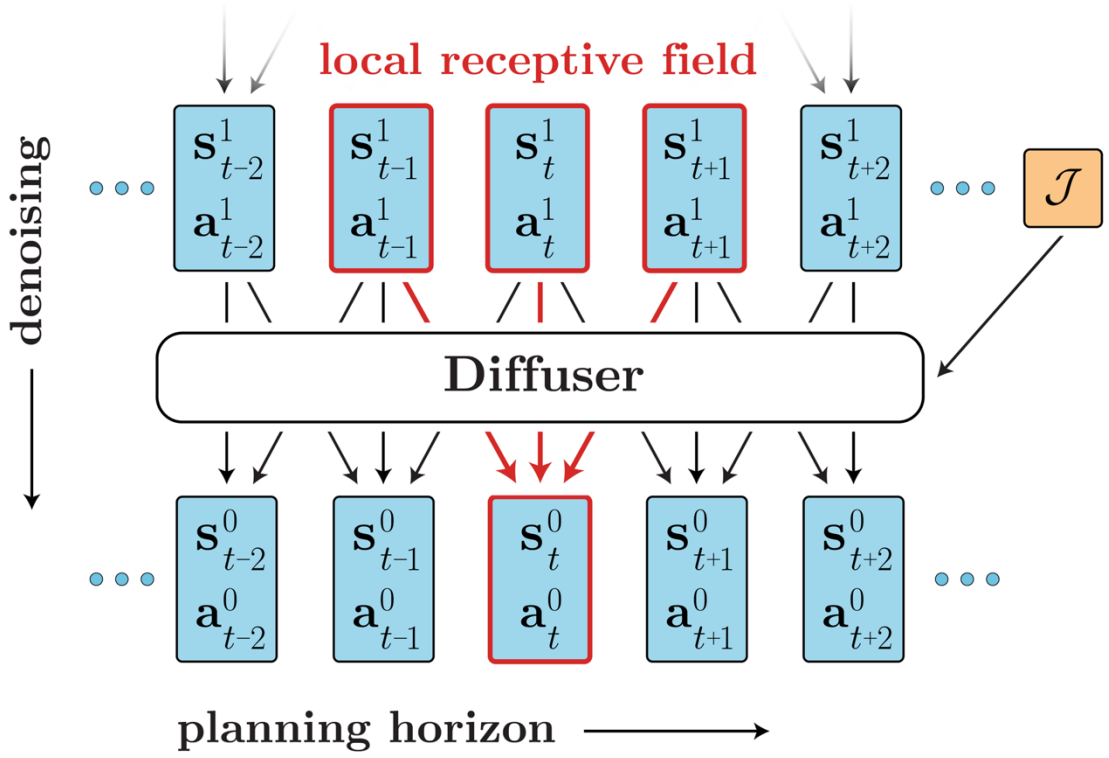
not step by step

Any length

set by the size of the noise

THE MODEL · 2

Sampling: iteratively refine a random trajectory



- U-Net with **1-D convolutions over time**
- Each step fixes **local** inconsistencies
- 100 steps → a coherent plan

STANDARD DDPM LOSS

$$\mathcal{L}(\theta) = \mathbb{E} \left\| \epsilon - \epsilon_\theta(\tau^i,\, i) \right\|^2$$

Figure: Janner et al., ICML 2022 (Figure 2).

FLEXIBLE BEHAVIOR SYNTHESIS

Planning = composing two distributions

$$\tilde{p}_\theta(\boldsymbol{\tau}) \propto p_\theta(\boldsymbol{\tau})\, h(\boldsymbol{\tau})$$

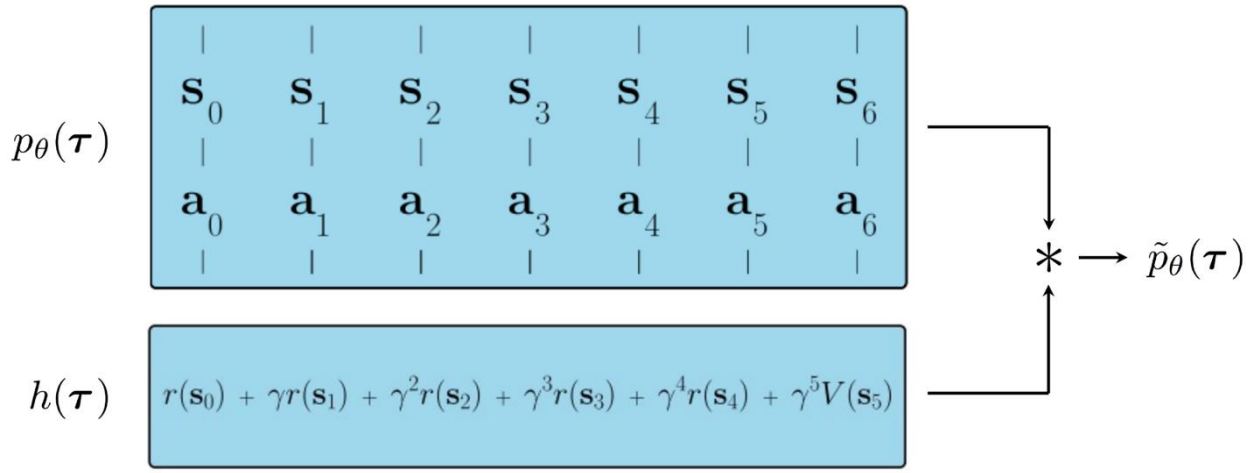
Behavior Model — Diffusion Model — Guidance Function

New task → new guidance function. No retraining.

Diagram: Janner & Du, ICML 2022 talk slides.

GUIDANCE 1 · REWARD

Offline RL through value guidance

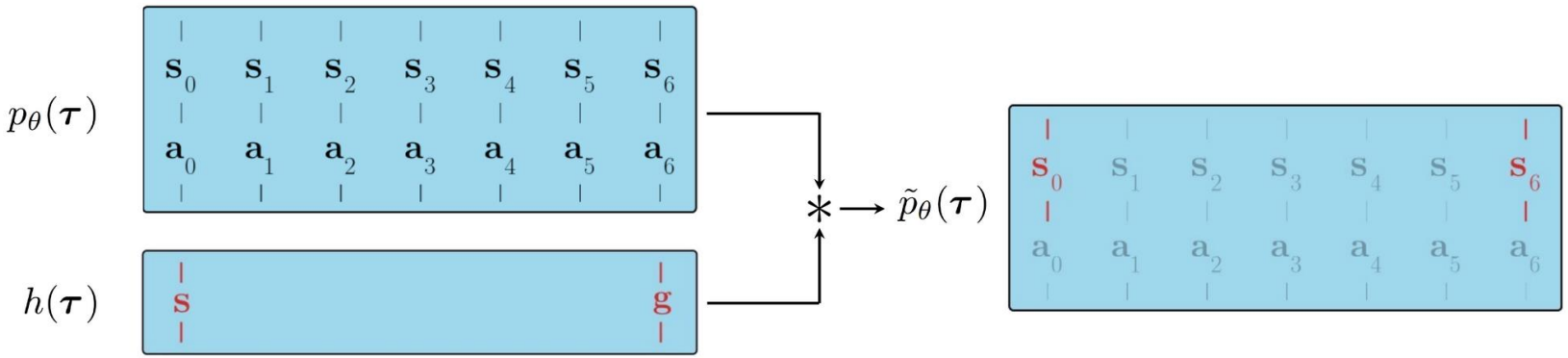


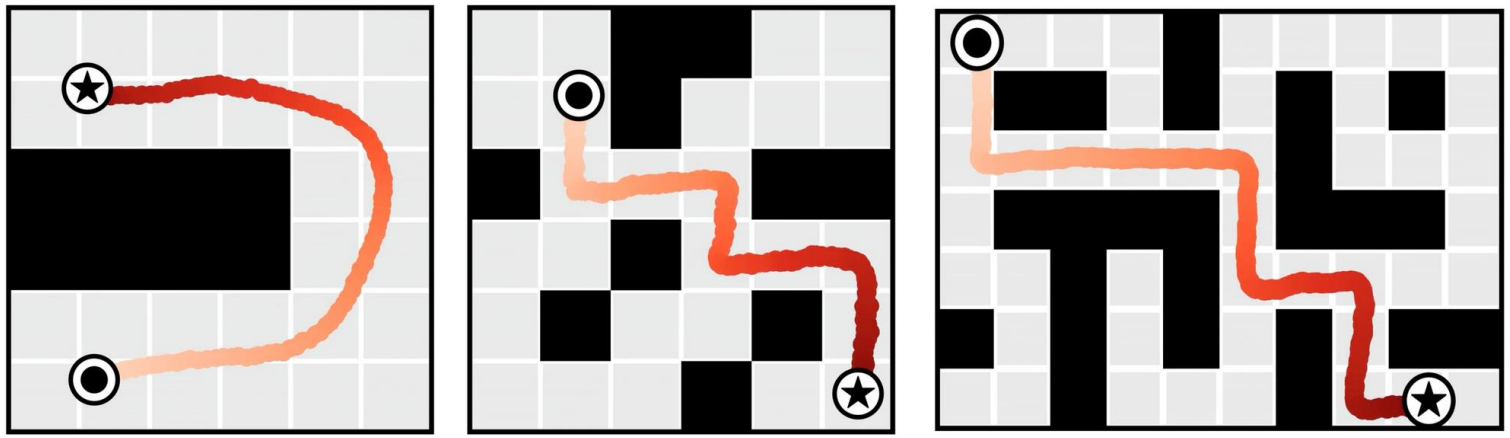
$$\tau^{i-1} \sim \mathcal{N}\Big(\underbrace{\mu}_{\text{denoise}} + \underbrace{\alpha\, \Sigma\, \nabla \mathcal{J}(\mu)}_{\text{push toward high return}},\ \Sigma \Big)$$

Top diagram: Janner & Du, ICML 2022 talk slides. Bottom: guided sampling step from the paper.

GUIDANCE 2 · GOAL

Goal reaching through inpainting

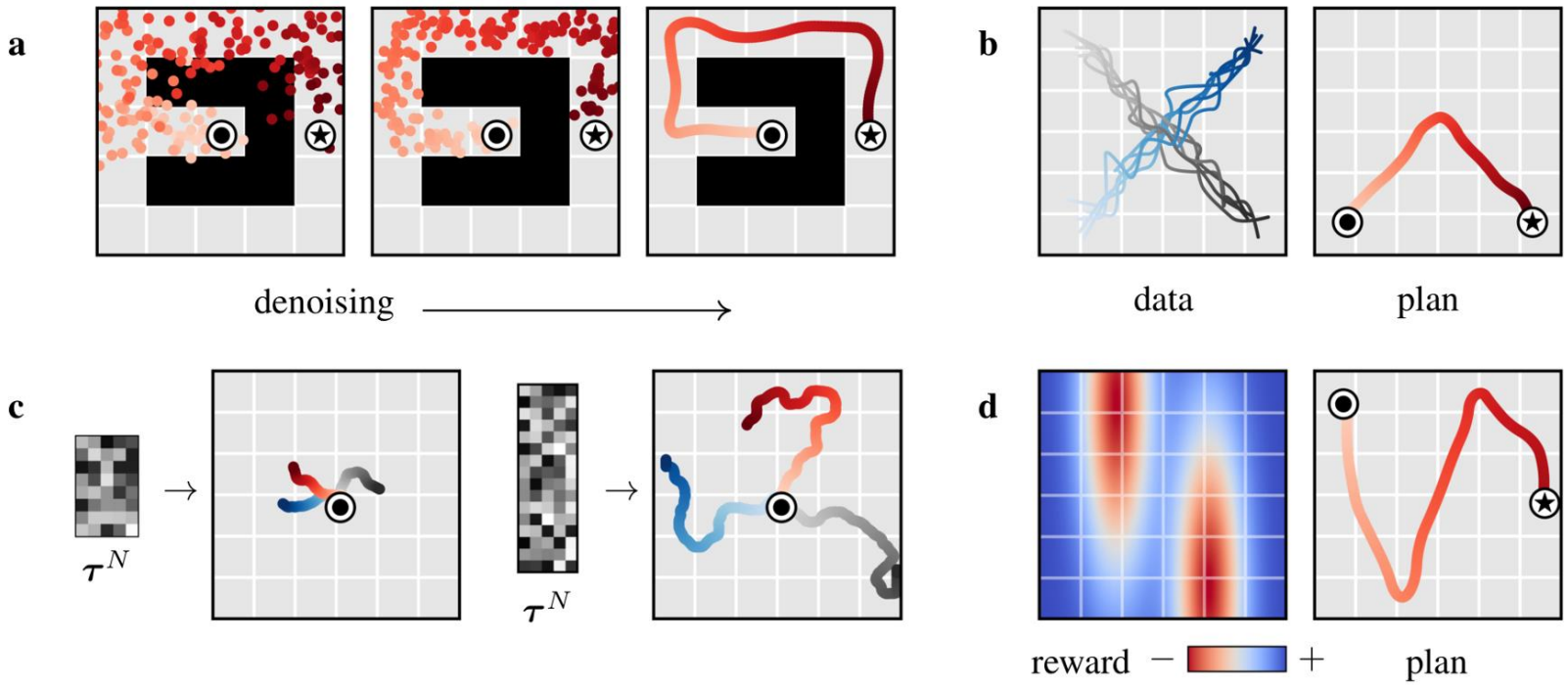




Diagrams: Janner & Du, ICML 2022 talk slides. Circle = start, star = goal.

WHAT FALLS OUT OF THIS DESIGN

Four properties for free



a · long horizons **b** · stitching **c** · variable length **d** · new rewards

Figure: Janner et al., ICML 2022 (Figure 3).

EXPERIMENT 1 · LONG HORIZON

Maze2D: reward only at the goal

Environment	IQL	Diffuser
Maze2D U-Maze	47.4	113.9
Maze2D Medium	34.9	121.5
Maze2D Large	58.6	123.0
Multi2D U-Maze	24.8	128.9
Multi2D Medium	12.1	127.2
Multi2D Large	13.9	132.1

Multi2D: a new goal every episode.

Same model; only the inpainted goal changes.

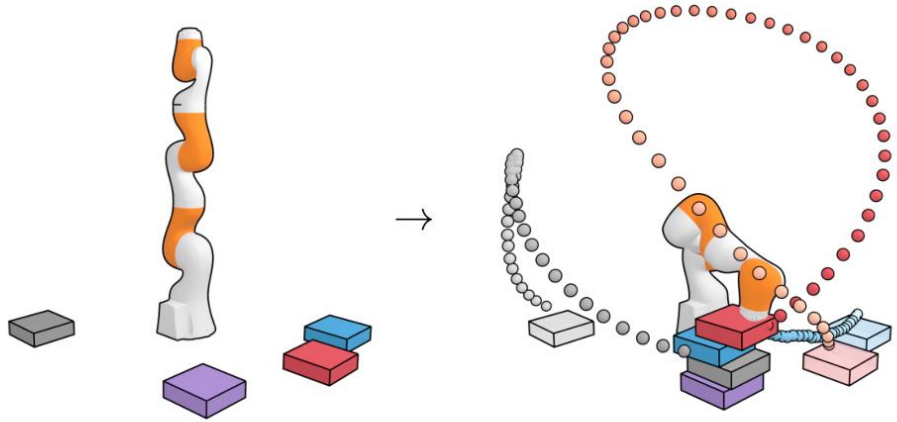
Data: Janner et al., ICML 2022 (Table 1). IQL is the strongest baseline; BCQ and CQL score below 13. 100 = reference expert.

EXPERIMENT 2 · TEST-TIME FLEXIBILITY

Block stacking: one model, three tasks

Task	CQL	Diffuser
Stack any tower	24.4	58.7
Stack in a given order	0.0	45.6
Rearrange	0.0	58.9

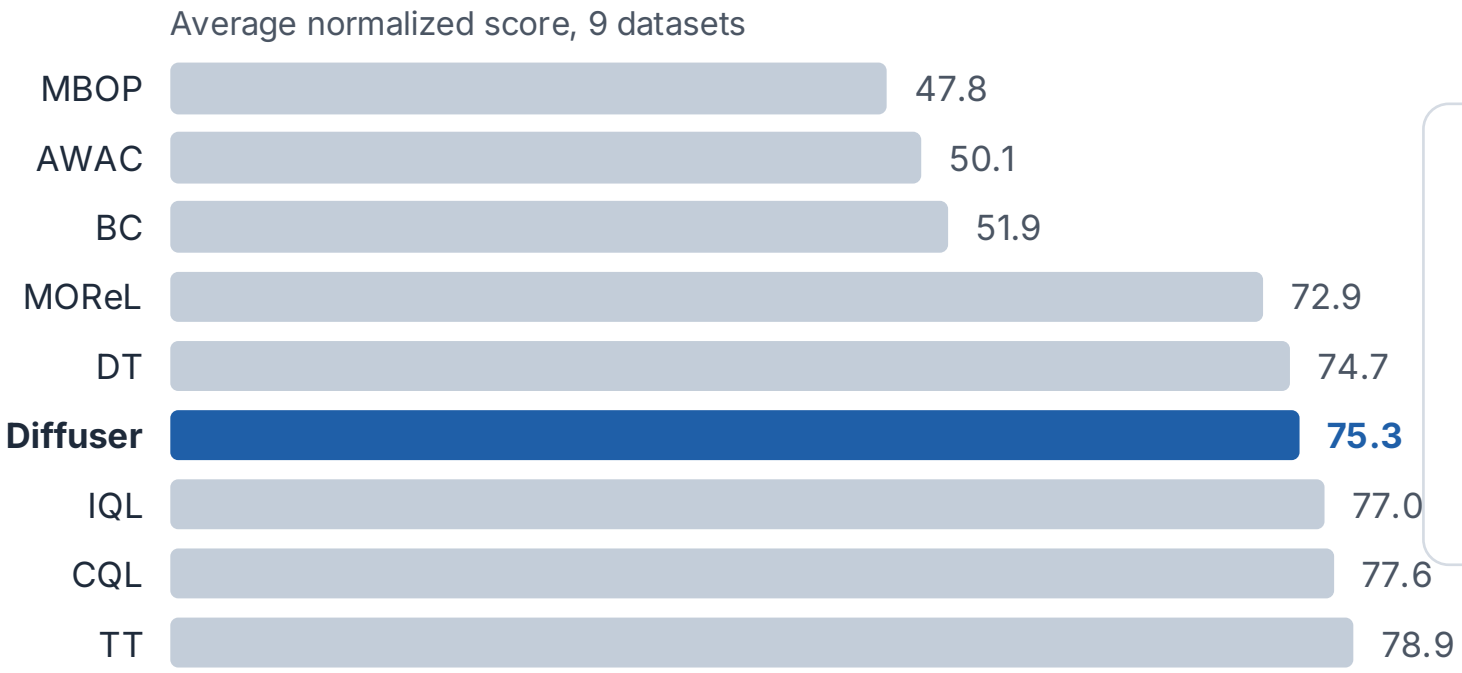
Only the guidance function changes.



Data and figure: Janner et al., ICML 2022 (Table 3, Figure 5). 100 = perfect stack. BCQ scores 0 on all three.

EXPERIMENT 3 · OFFLINE RL

D4RL locomotion: competitive, not the best



KEY ABLATION

Diffuser as a dynamics model inside MPPI ≈ random

The gain comes from planning by sampling.

Data: Janner et al., ICML 2022 (Table 2).

PRACTICAL COST

Slow, but warm-starting helps

100 denoising passes per action.

Fix: start from the last plan, slightly re-noised.

Little loss with one-tenth of the steps.

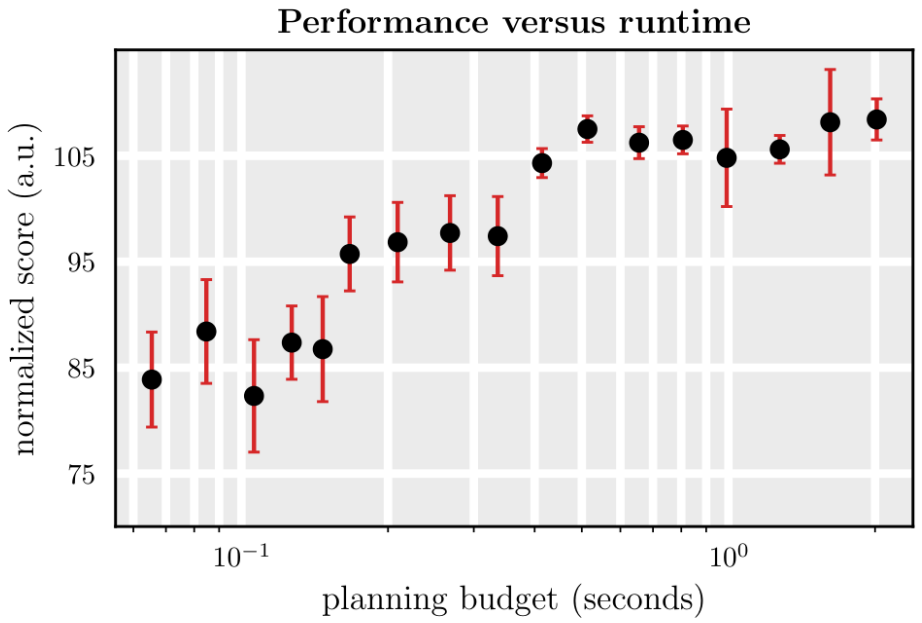


Figure: Janner et al., ICML 2022 (Figure 7), Walker2d Medium-Expert.

ASSESSMENT

Strengths and weaknesses

STRENGTHS

- One model, many tasks
- Long horizons with sparse reward
- Image-diffusion tools transfer directly

WEAKNESSES

- Slow sampling
- Generated states and actions may disagree
- Not the best on D4RL

CONTEXT

Where trajectory diffusion went next

ICLR 2023

Decision Diffuser

States only; actions from inverse dynamics.

RSS 2023 · NEXT TALK

Diffusion Policy

Short action chunks from images, on real robots.

NEURIPS 2024

Diffusion Forcing

Per-token noise levels: between autoregression and full-sequence diffusion.

DISCUSSION

Questions for the class

1 · Planning = sampling. Can a generative prior plan beyond the behaviours in its data, or does it mostly recombine them?

2 · Joint states and actions. When should we diffuse states only and recover actions with inverse dynamics, as Decision Diffuser does?