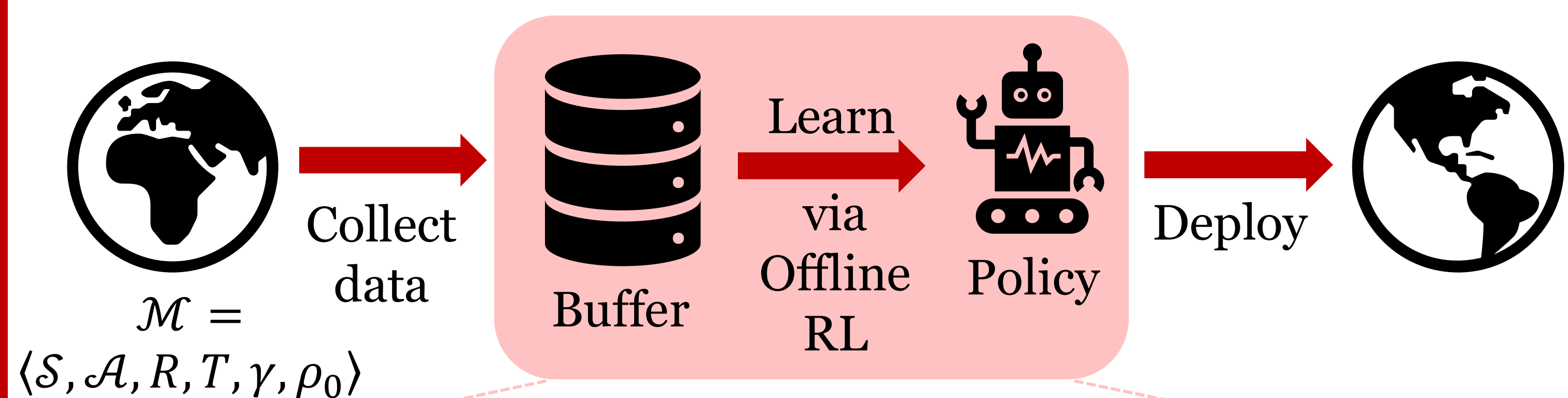


Neural Stochastic Differential Equations for Uncertainty-Aware Offline RL

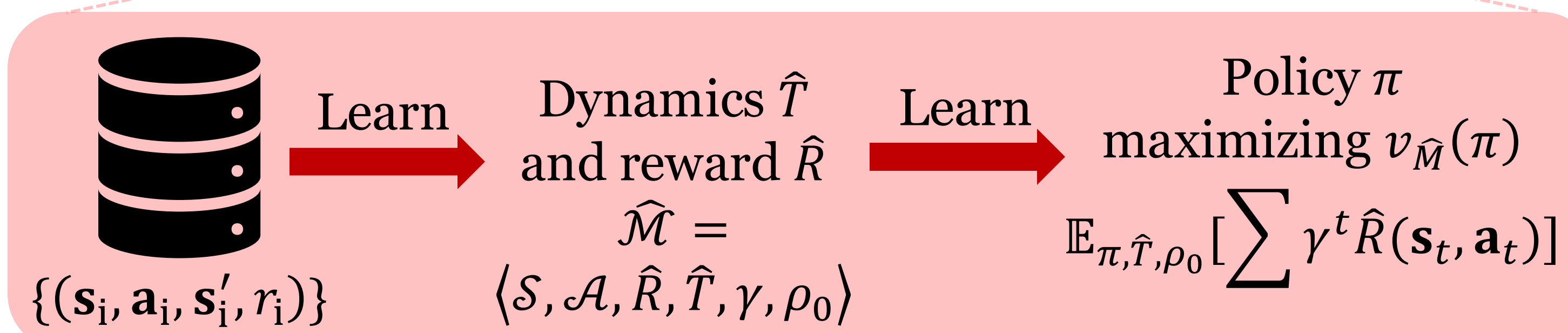
Cevahir Koprulu¹, Franck Djeumou², Ufuk Topcu¹



Problem Setting



Offline Model-based RL



Model Exploitation

Estimated return exceeds the true return $v_{\hat{M}}(\pi) - v_M(\pi) > 0$. Policy π learns to **exploit the regions of state-action space with high model uncertainty** as well as high estimated return.

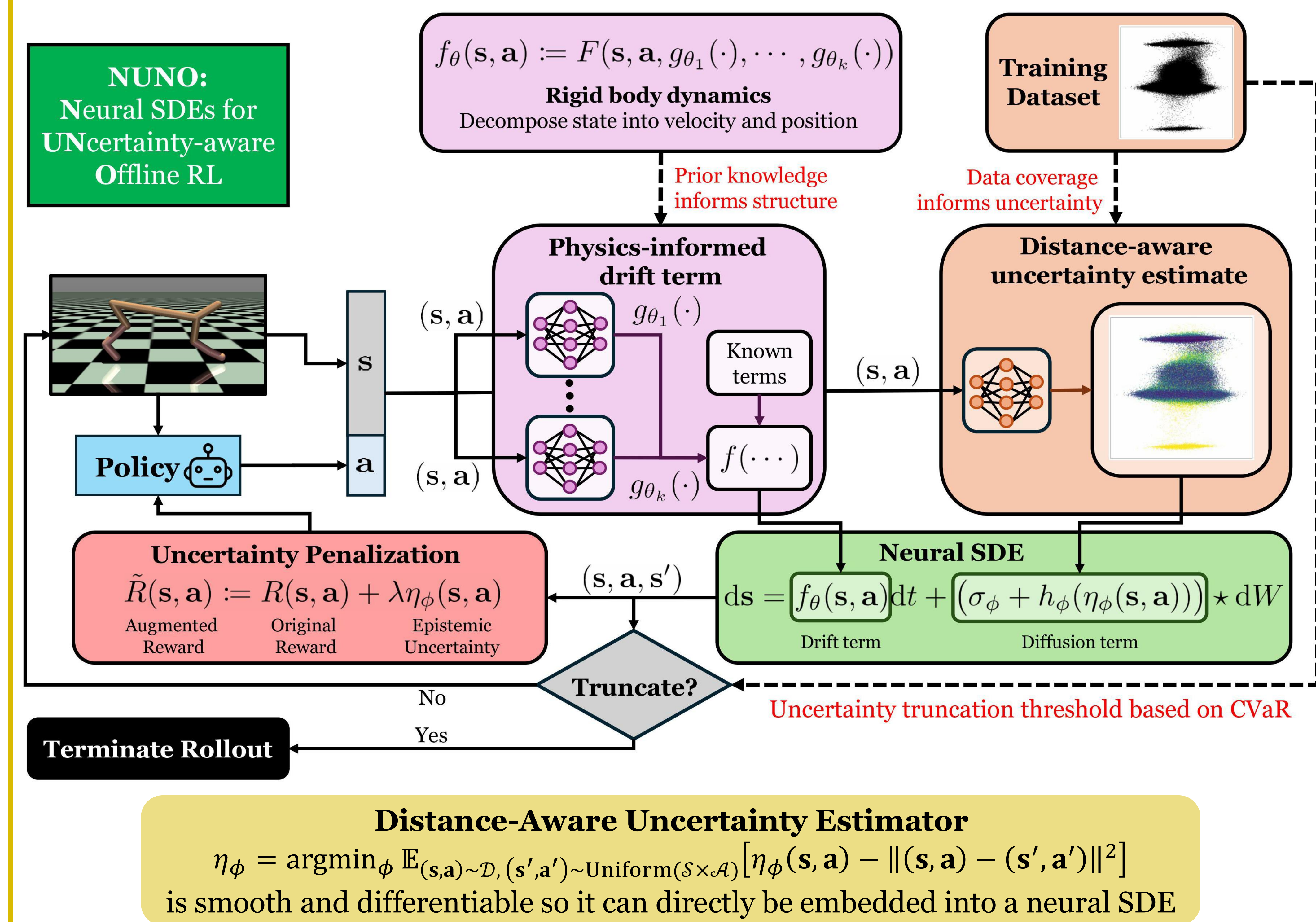
Popular Remedies

Enforce conservatism by uncertainty penalization/truncation. **However**, they suffer especially when data quality is low.

Use neural SDEs with distance-aware uncertainty estimates when your data sucks!



NUNO

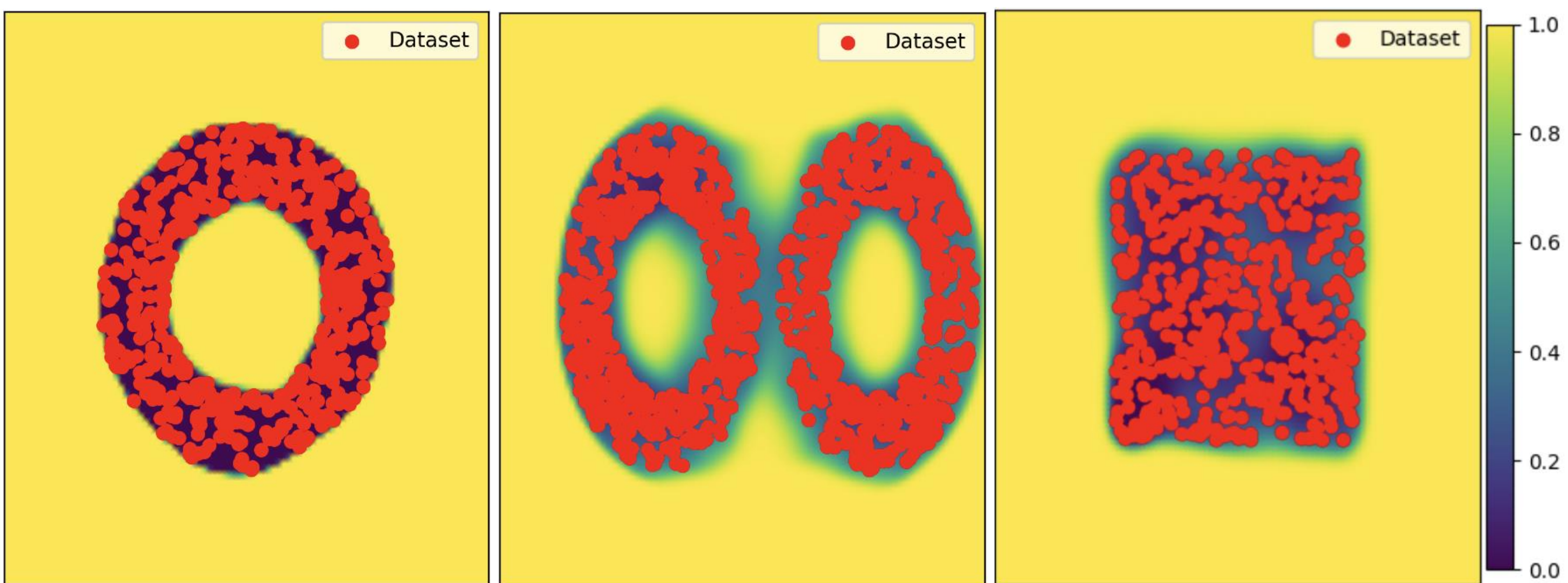


Results

Theoretical Result

Penalization via a distance-aware uncertainty estimator enables conservatism by encouraging the policy to stay close to the convex hull of the offline data.

To address multiple clusters, we enforce low values with almost-zero gradients when evaluated on near points and provide high but bounded values when evaluated far from the training data.



The distance-aware uncertainty estimate η_{ϕ} on three synthetic datasets. The red points represent the state-action samples in the dataset. Yellow indicates high uncertainty, while dark blue represents low uncertainty. The x and y axes denote the states of the system, respectively.

Experimental Result 1: Average human-normalized scores

NUNO outperforms SOTA in low-quality datasets and overall.

D4RL MuJoCo

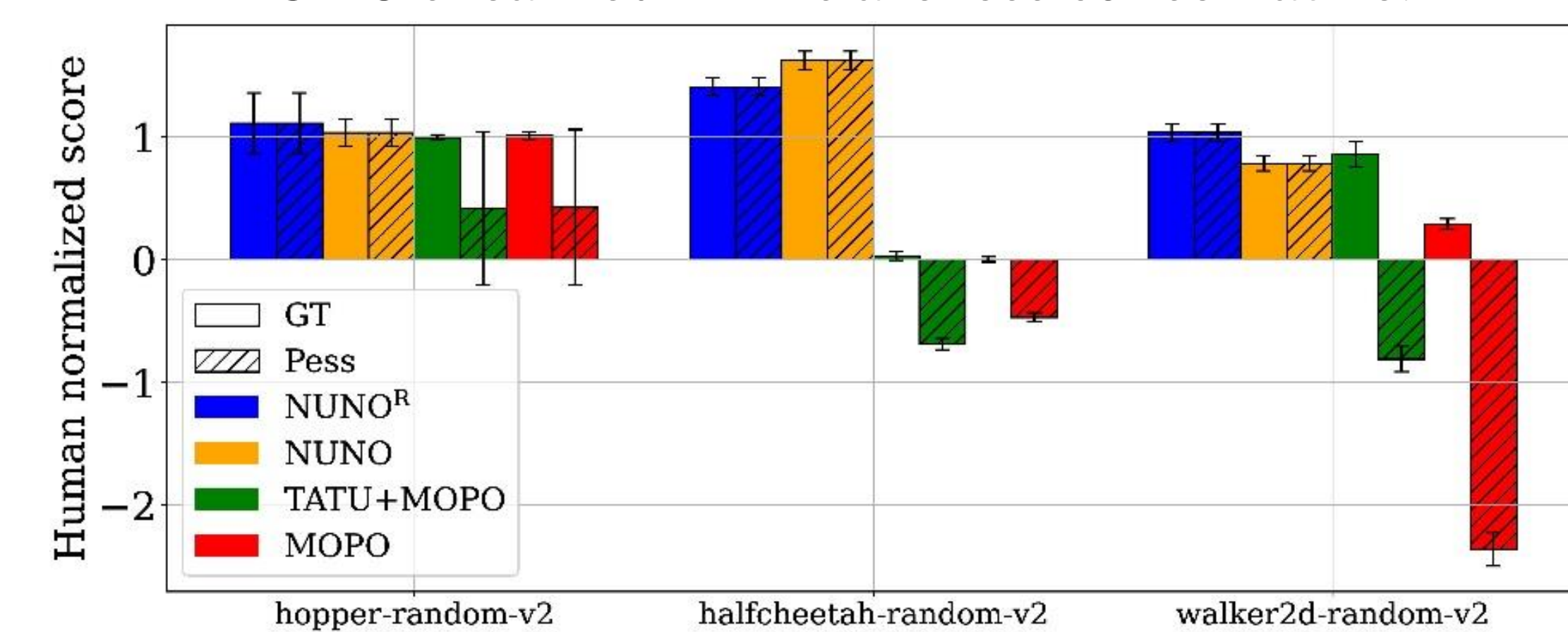
Task	NUNO	NUNO ^R	MOBILE	MOPO ^T	MOPO
hc-r	52.7 ± 3.4	52.2 ± 0.5	39.3 ± 3.0	33.3	35.9
hp-r	73.2 ± 9.8	53.7 ± 13.9	31.9 ± 0.6	31.9	16.7
wk-r	27.7 ± 0.9	28.1 ± 1.2	17.9 ± 6.6	10.4	4.2
Overall	83.8	82.4	80.0	71.3	49.4

NeoRL MuJoCo

Task	NUNO	NUNO ^R	MOBILE	MOPO
hc-L	52.5 ± 0.6	58.4 ± 0.5	54.7 ± 3.0	40.1
hp-L	26.9 ± 3.8	26.4 ± 6.8	17.4 ± 3.9	6.2
wk-L	52.5 ± 2.5	49.4 ± 1.9	37.6 ± 2.0	11.6
Overall	70.6	68	60.7	38.5

Experimental Result 2: Model exploitation

NUNO's learned MDPs are less conservative.



Experimental Result 3: Model accuracy

NSDEs are more accurate than Gaussian ensembles.

