Zero-Shot Reinforcement Learning from Low Quality Data

NEURIPS 2024

Scott Jeen[α], Tom Bewley[β] & Jonathan M. Cullen[α]

[α] University of Cambridge [β] University of Bristol

https://enjeeneer.io/projects/zero-shot-rl/

Motivation

- Training policies to (zero-shot) generalise to unseen tasks in an environment is hard! [1]

- Behaviour Foundation Models (BFMs) based on forward-backward representations (FB) [2] and universal successor features (USF) [3], provide principled mechanisms for performing zero-shot task generalisation

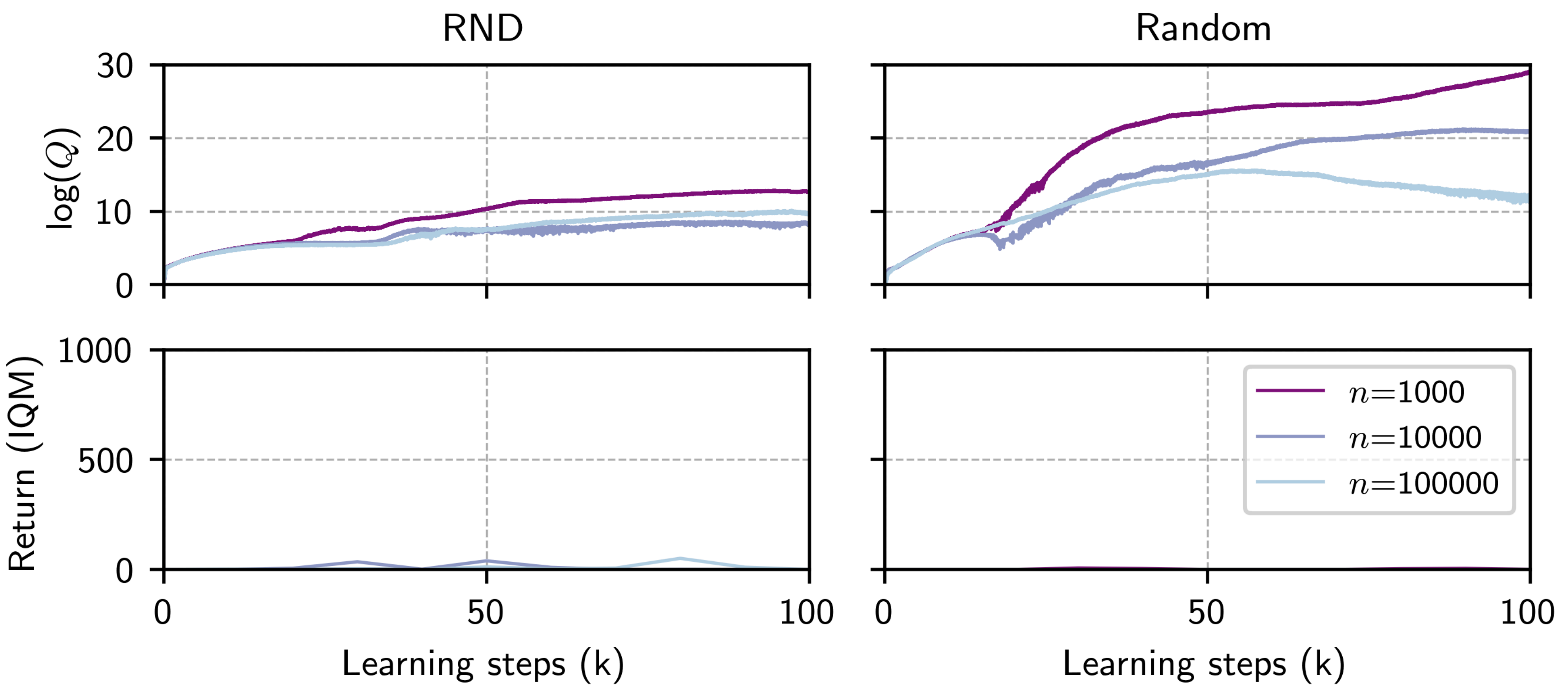
- However, BFMs assumed access to idealised (large & diverse) pre-training datasets that we can't expect for real problems

- **Can we pre-train BFMs on realistic (small & narrow) datasets?**

[1] Robert Kirk, Amy Zhang, Edward Grefenstette, and Tim Rocktäschel. *A survey of zero-shot generalisation in deep reinforcement learning*. JAIR 2023

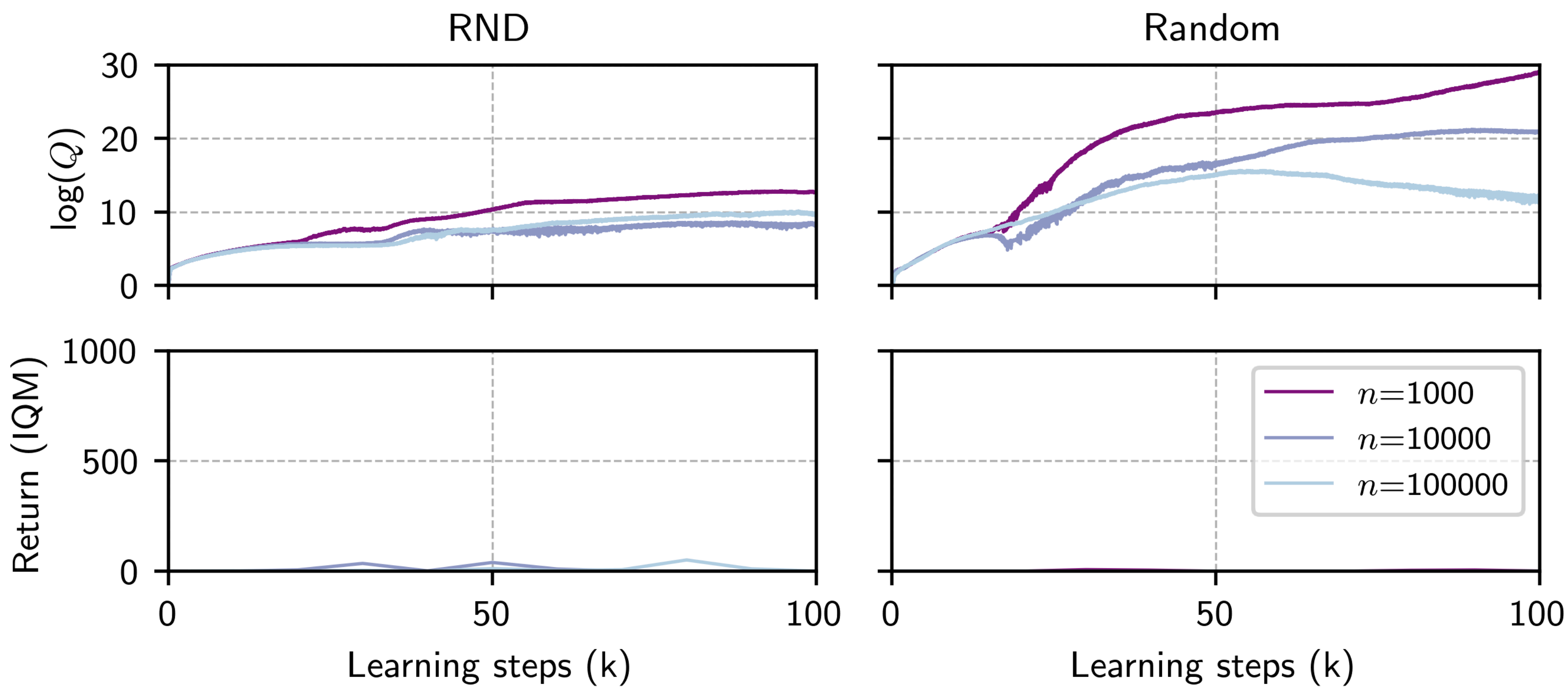
[2] Ahmed Touati, Jérémy Rapin, and Yann Ollivier. *Does zero-shot reinforcement learning exist?* ICLR 2023

Out-of-distribution Value Overestimation in BFMs



https://enjeeneer.io/projects/zero-shot-rl/

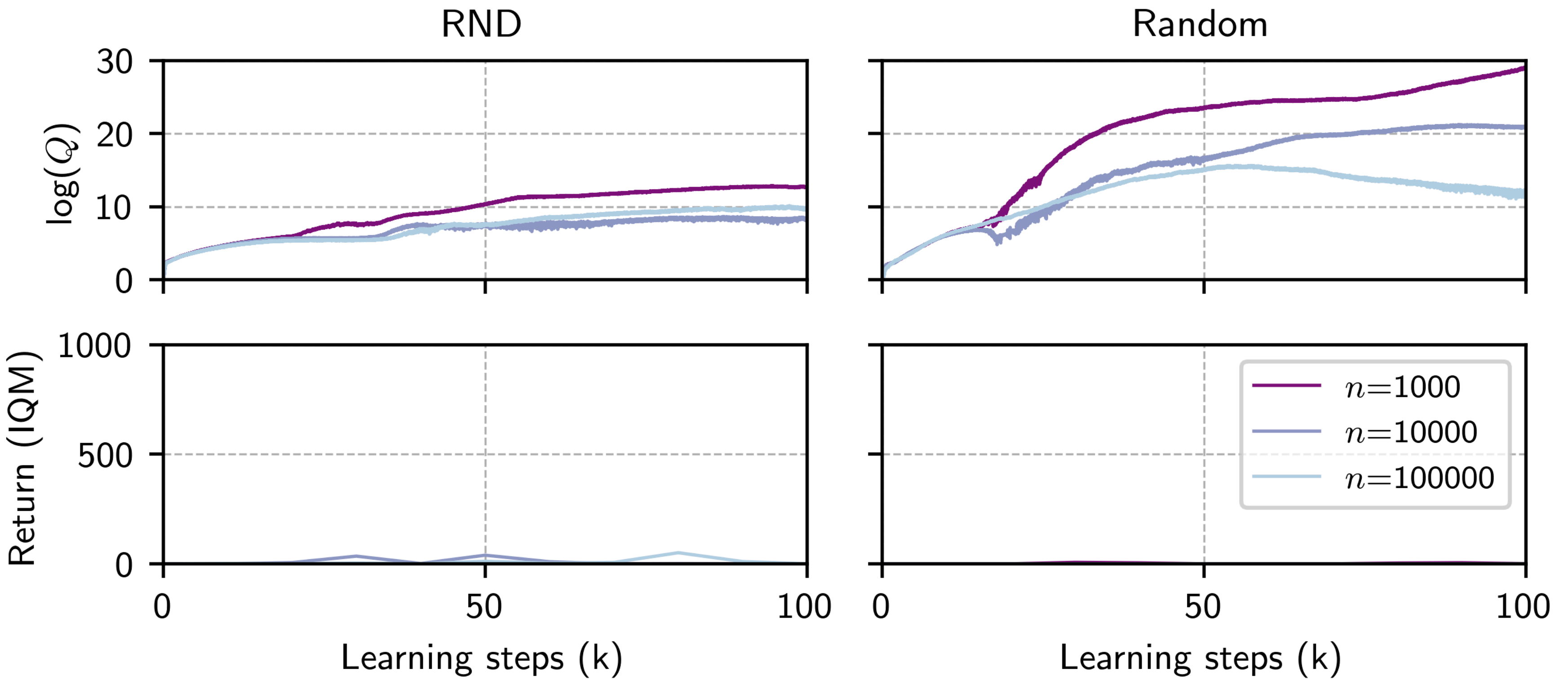
Out-of-distribution Value Overestimation in BFMs



$$\mathcal{L}_{\text{FB}} = \mathbb{E}_{(s_t,a_t,s_{t+1},s_+)\sim\mathcal{D},z\sim\mathcal{Z}}[(F(s_t,a_t,z)^\top B(s_+) - \gamma\bar{F}(s_{t+1},\pi_z(s_{t+1}),z)^\top\bar{B}(s_+))^2 - 2F(s_t,a_t,z)^\top B(s_{t+1})]$$

https://enjeeneer.io/projects/zero-shot-rl/

Out-of-distribution Value Overestimation in BFMs



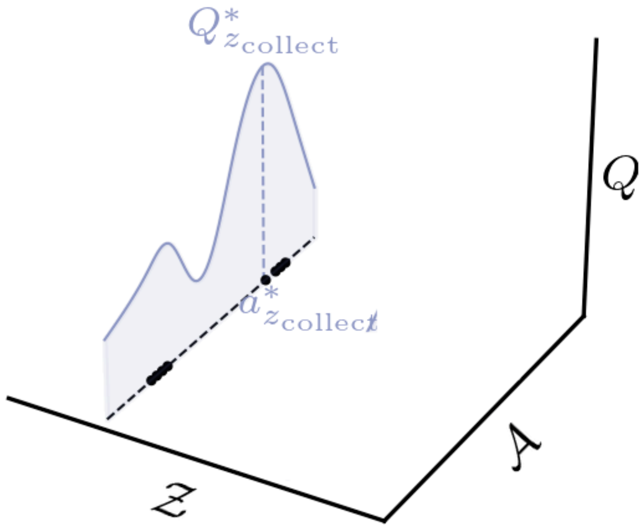
$$\mathcal{L}_{\mathrm{FB}} = \mathbb{E}_{(s_t, a_t, s_{t+1}, s_+) \sim \mathcal{D}, z \sim \mathcal{Z}}\left[\left(F(s_t, a_t, z)^\top B(s_+) - \gamma \bar{F}(s_{t+1}, \underbrace{\pi_z(s_{t+1})}_{\text{OOD}}, z)^\top \bar{B}(s_+)\right)^2 - 2F(s_t, a_t, z)^\top B(s_{t+1})\right]$$

[3] Aviral Kumar, Justin Fu, Matthew Soh, George Tucker, and Sergey Levine. *Stabilizing off-policy q-learning via*

https://enjeeneer.io/projects/zero-shot-rl/

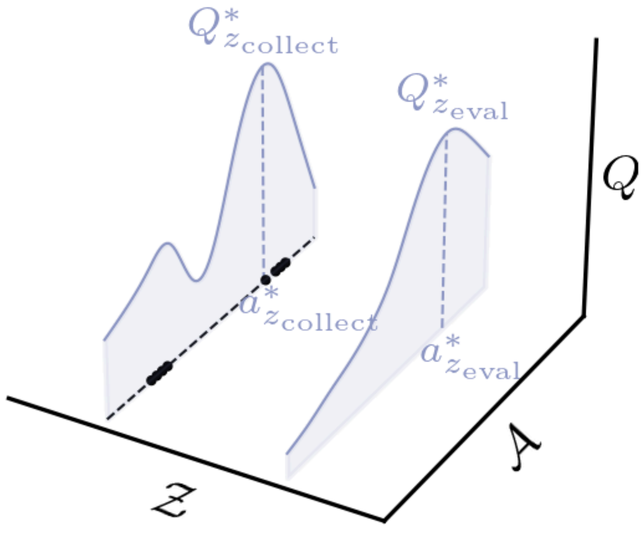
Conservative BFMs

Conservative BFMs



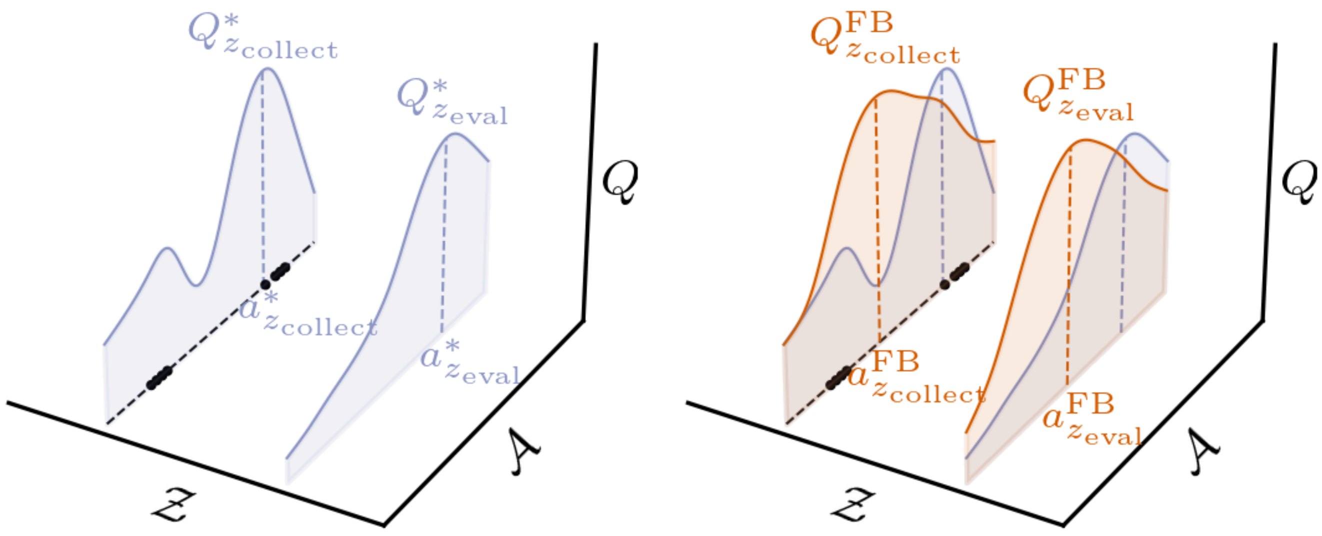
https://enjeeneer.io/projects/zero-shot-rl/

Conservative BFMs



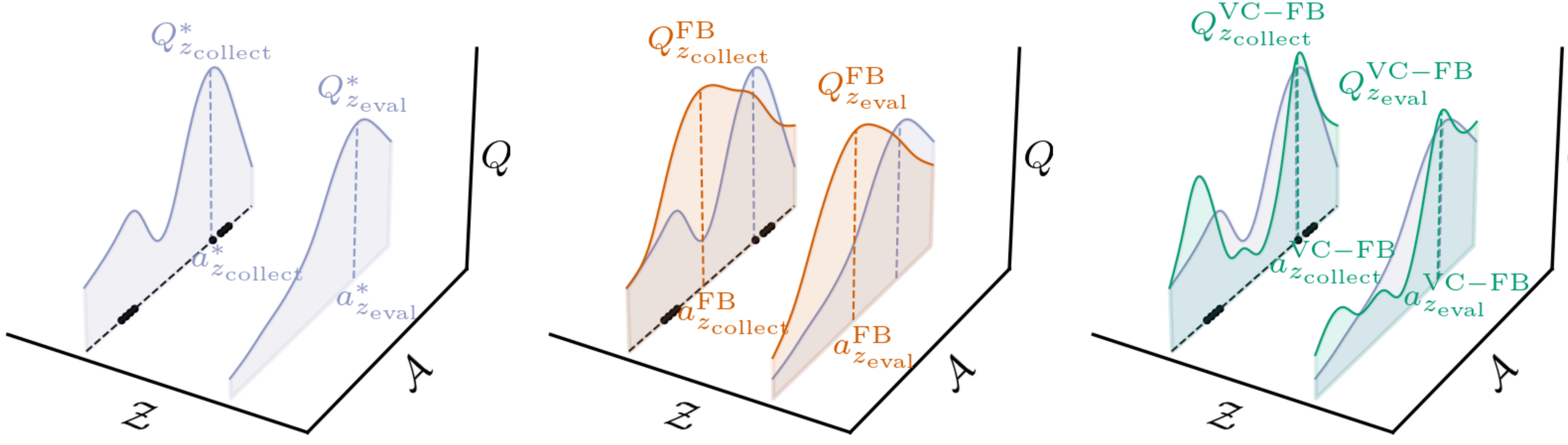
https://enjeeneer.io/projects/zero-shot-rl/

Conservative BFMs



https://enjeeneer.io/projects/zero-shot-rl/

Conservative BFMs



$$\mathcal{L}_{\text{VC-FB}} = \alpha \cdot \left(\mathbb{E}_{s\sim\mathcal{D}, a\sim\mu(a|s), z\sim\mathcal{Z}}[F(s,a,z)^\top z] - \mathbb{E}_{(s,a)\sim\mathcal{D}, z\sim\mathcal{Z}}[F(s,a,z)^\top z] - \mathcal{H}(\mu)\right) + \mathcal{L}_{\text{FB}}$$

$$\mathcal{L}_{\text{MC-FB}} = \alpha \cdot \left(\mathbb{E}_{s\sim\mathcal{D}, a\sim\mu(a|s), z\sim\mathcal{Z}, s_+\sim\mathcal{D}}[F(s,a,z)^\top B(s_+)] - \mathbb{E}_{(s,a)\sim\mathcal{D}, z\sim\mathcal{Z}, s_+\sim\mathcal{D}}[F(s,a,z)^\top B(s_+)] - \mathcal{H}(\mu)\right) + \mathcal{L}_{\text{FB}}$$

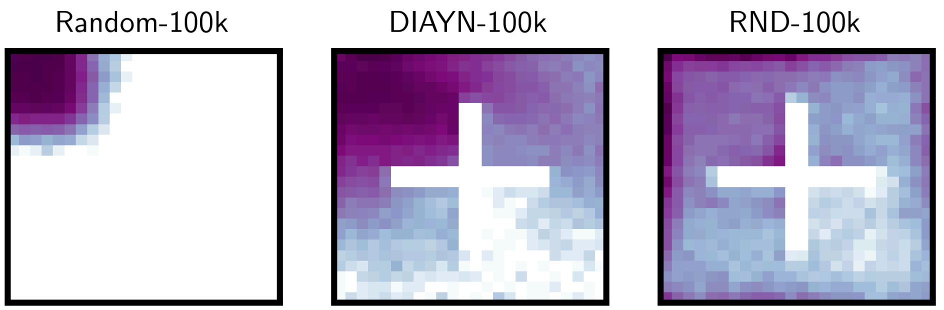
[4] Aviral Kumar, Aurick Zhou, George Tucker, and Sergey Levine. *Conservative q-learning for offline reinforcement*

ExORL Results

Baselines

- **Zero-shot RL:** FB, SF-LAP [5]
- **Goal-conditioned RL:** GC-IQL [6]
- **Offline RL:** CQL [7]

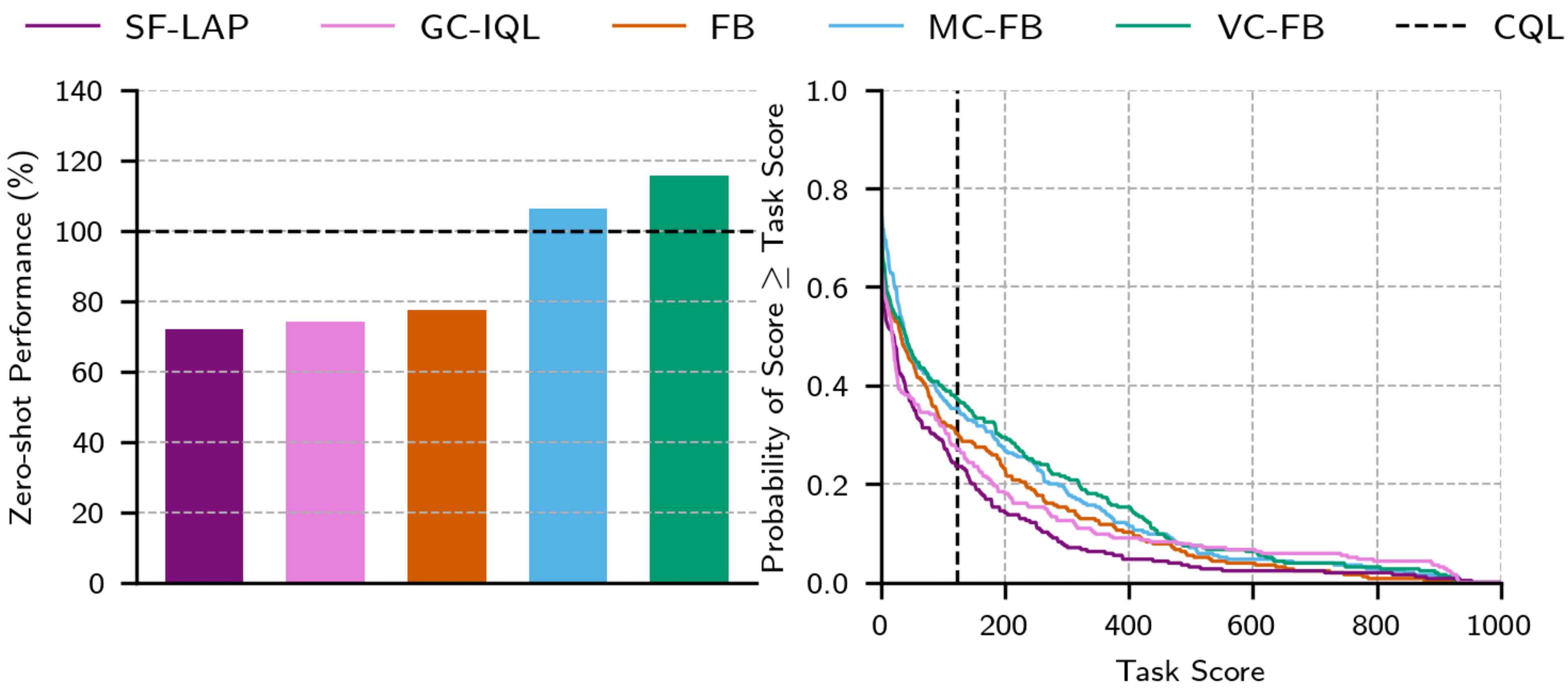
Datasets



[5] Ahmed Touati, Jérémy Rapin, and Yann Ollivier. *Does zero-shot reinforcement learning exist?* ICLR 2023

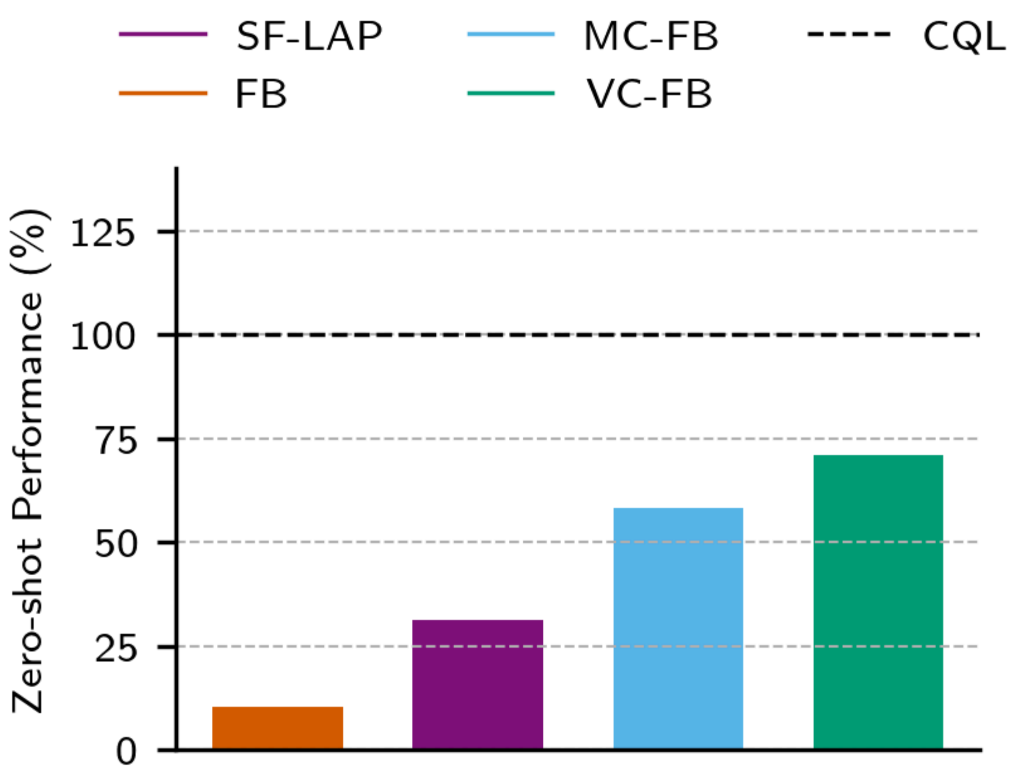
[6] Seohong Park, Dibya Ghosh, Benjamin Eysenbach, and Sergey Levine. *Hiql: Offline goalconditioned rl with latent states as actions.* NeurIPS 2023.

ExORL Results



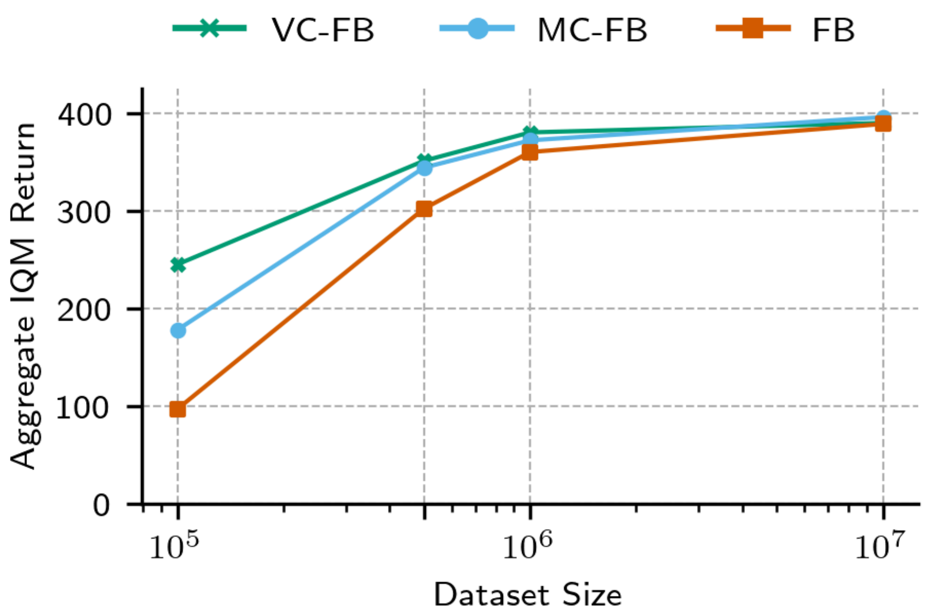
https://enjeeneer.io/projects/zero-shot-rl/

D4RL Results



https://enjeeneer.io/projects/zero-shot-rl/

Performance on Idealised Datasets is Unaffected



Dataset	Domain	Task	FB	VC-FB	MC-FB
RND	all	all	389	390	396
DIAYN	all	all	269	280	283
RANDOM	all	all	111	131	133
ALL	all	all	256	267	**271**

https://enjeeneer.io/projects/zero-shot-rl/

Conclusions

- Like standard offline RL methods, BFMs suffer from the *distribution shift*
- As a resolution, we introduce *Conservative* BFMs
- *Conservative* BFMs considerably outperform standard BFMs on low-quality datasets
- *Conservative* BFMs do not compromise performance on idealised datasets

Project page:

Twitter/X: @enjeeneer

Website: https://enjeeneer.io