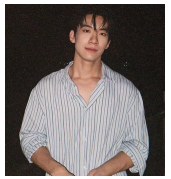
PrivORL: Differentially Private Synthetic Dataset for Offline Reinforcement Learning

Network and Distributed System Security (NDSS) Symposium 2026

Presenter: Chen Gong (University of Virginia)

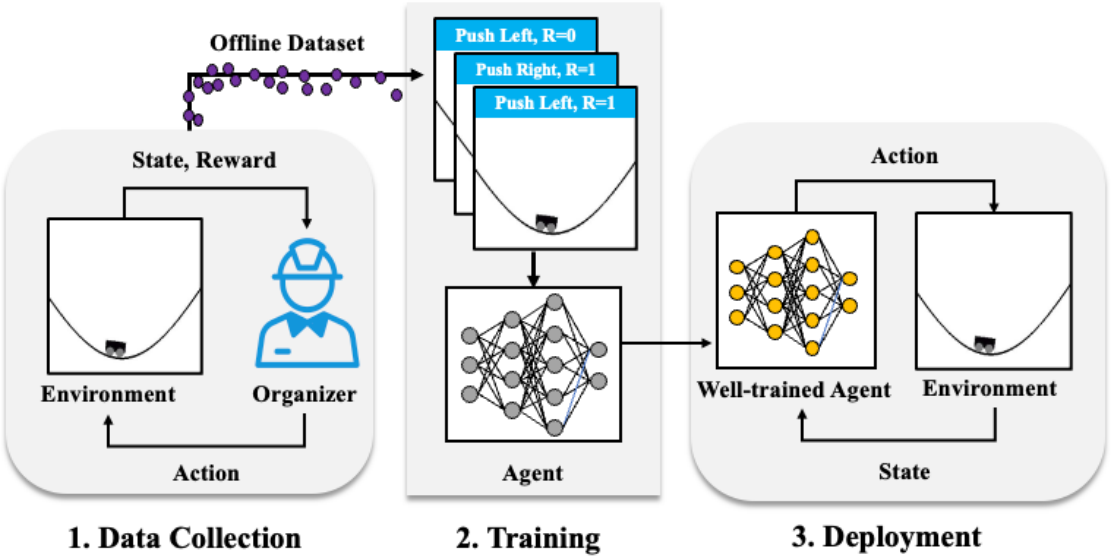
Chen Gong*, Zheng Liu*, Kecen Li, Tianhao Wang

*Indicates Equal Contribution

University of Virginia

Offline Reinforcement Learning

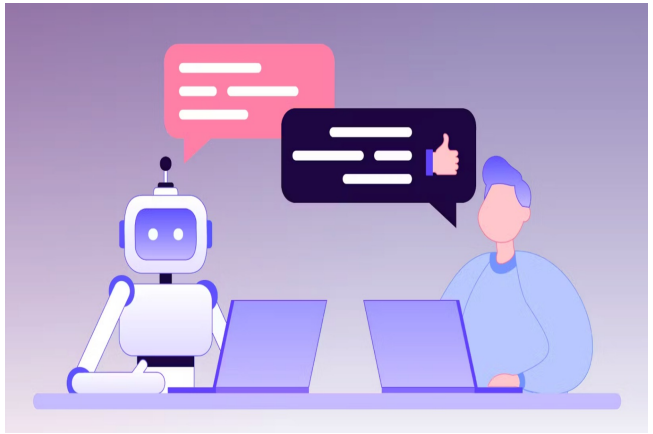


An Agent trained by interacting with a fixed dataset.

Trajectory:

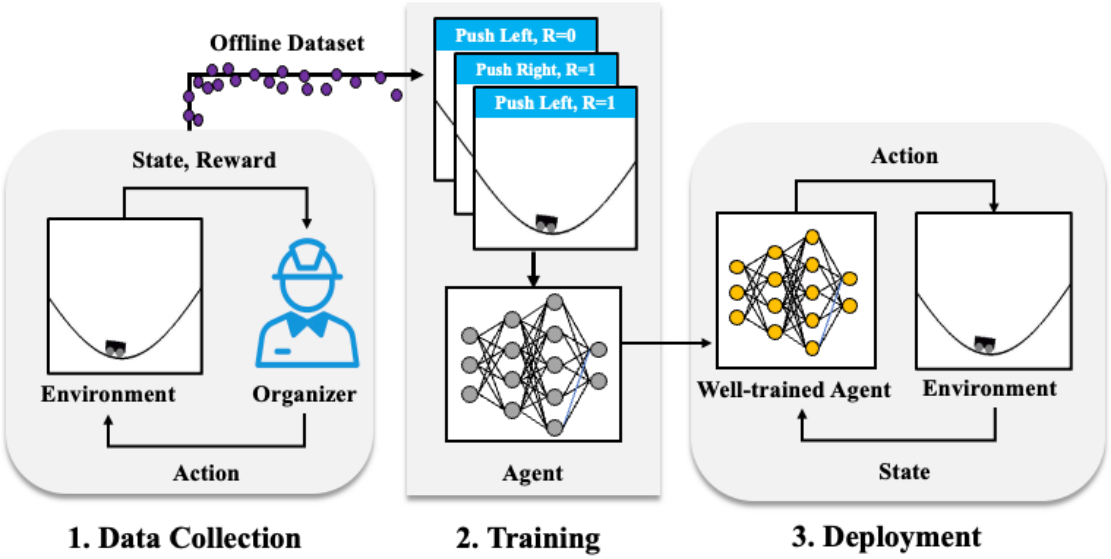
$$(s_0, a_0, r_0, s_1, a_1, r_1, s_2, a_2, r_2, \cdots, s_{|\tau|}, a_{|\tau|}, r_{|\tau|})$$

s_t: the current state;
a_t : current action executed by agents;
r_t : current reward feedback from environment;
s_{t+1}: the next time state;



Levine, et al, 2020.

Offline Reinforcement Learning

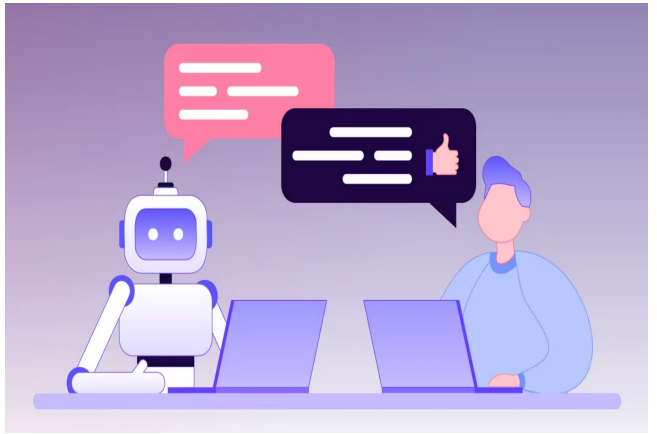


An Agent trained by interacting with a fixed dataset.

Transition:

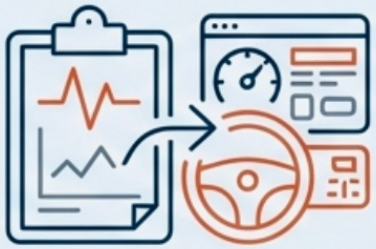
$$(s_0, a_0, r_0, s_1, a_1, r_1, s_2, a_2, r_2, \cdots, s_{|\tau|}, a_{|\tau|}, r_{|\tau|})$$

s_t: the current state;
a_t : current action executed by agents;
r_t : current reward feedback from environment;
s_{t+1}: the next time state;



Levine, et al, 2020.

Privacy Concerns



Strength of Offline RL

Allow the agent to learn from a **static dataset** without needing to interact with the real environment in potentially dangerous ways.

Autonomous Driving: Decision-making in complex road scenarios — reduces the risks associated with real-world road testing.

Privacy Concerns

A static dataset contains sensitive user information.

Membership Inference Attacks: An attacker can infer whether a specific user's medical record is included in the dataset.

Environment Information Leakage: An attacker may reverse-engineer private environment configurations.

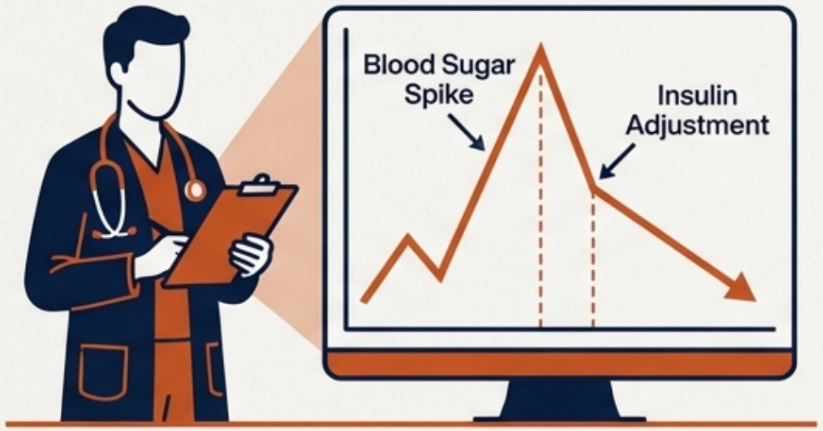
"We need large amounts of data to train intelligent agents, but directly sharing such data is ethically and legally infeasible."

Levine, et al, 2020.

Offline Reinforcement Learning Datasets

Two Distinct Forms of User Information

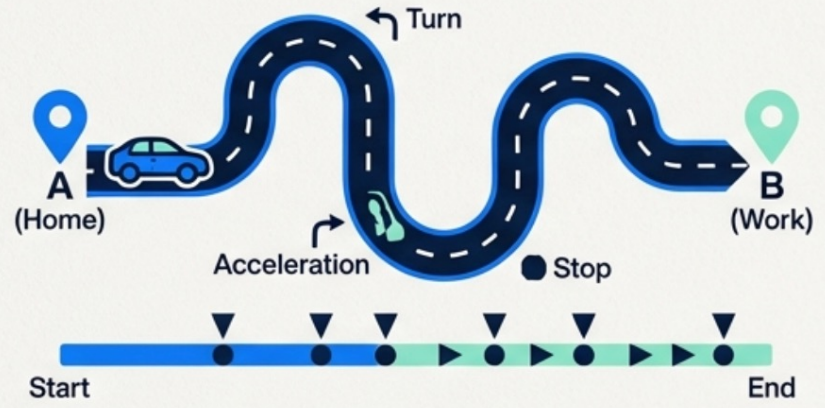
Transition Level (The Micro View)



$$\{s_t, a_t, r_t, s_{t+1}\}$$

A single user contributing a transition, such as a doctor observing a diabetic patient's blood sugar level, adjusting insulin dosage, and recording the response. Practical for analyzing the immediate impact of isolated decisions.

Trajectory Level (The Macro View)

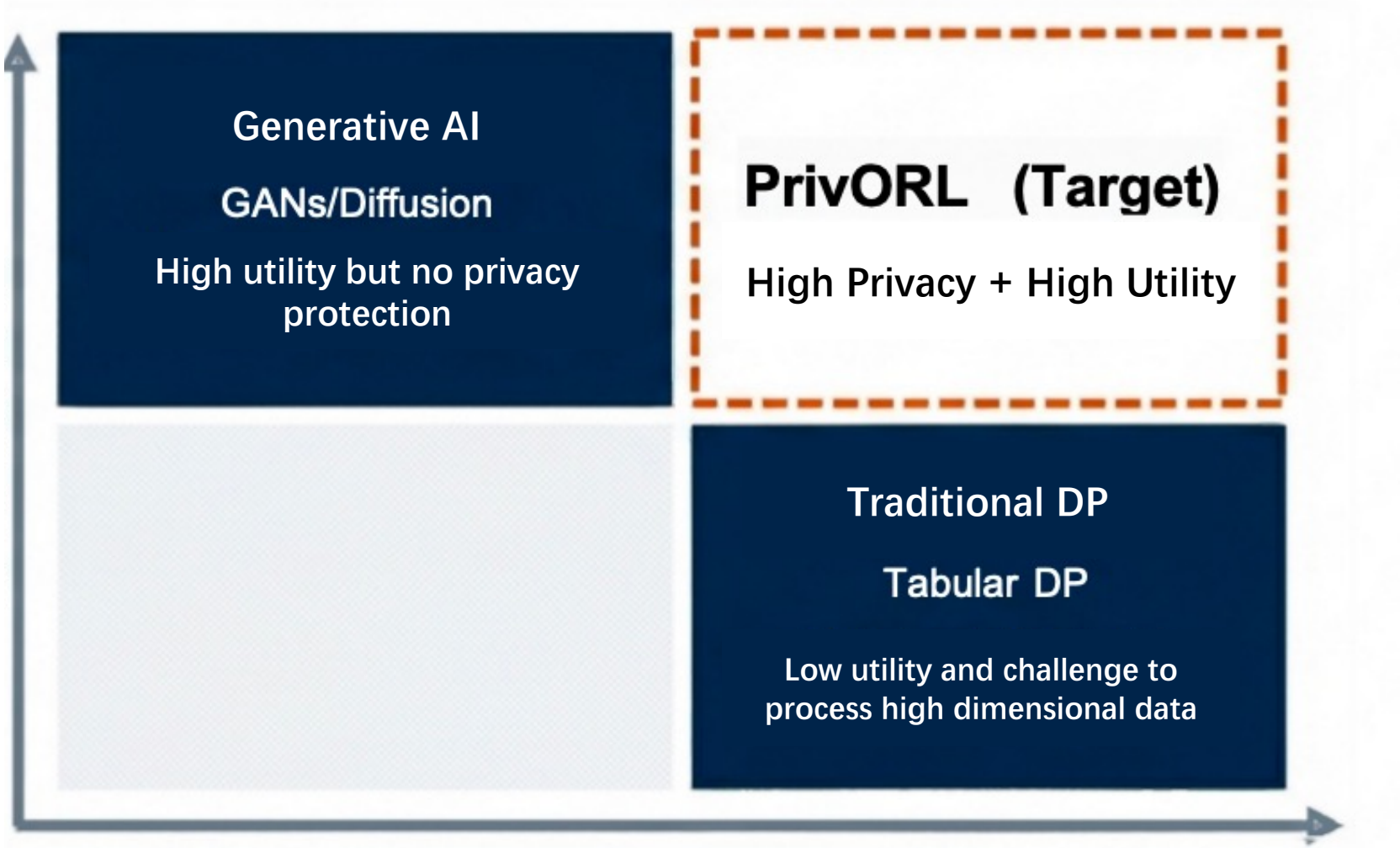


$$\{(s_0, a_0, r_0), \ldots, (s_T, a_T, r_T)\}$$

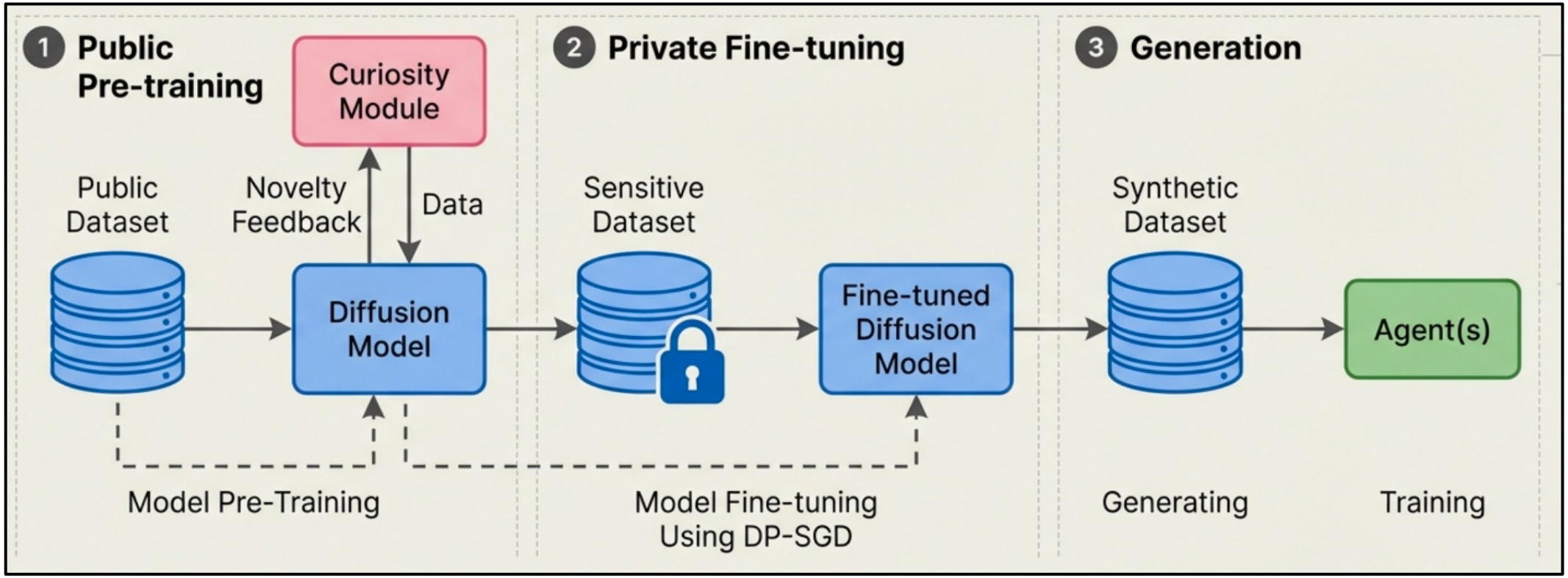
A user contributing an entire trajectory, like a driver documenting a full trip—capturing every turn, acceleration, and stop. Reasonable for long-term outcomes and analyzing how actions interplay over time.

Levine, et al, 2020.

Previous Works

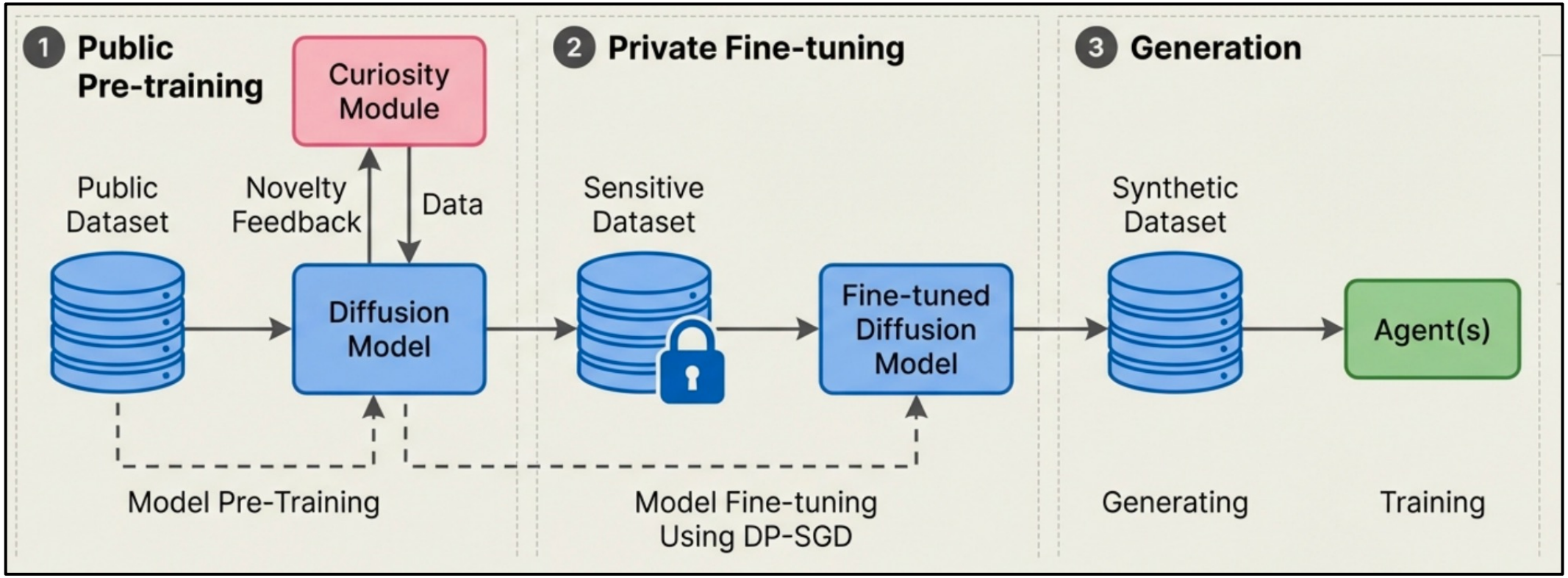


PrivORL: A Unified framework for Private Synthesis



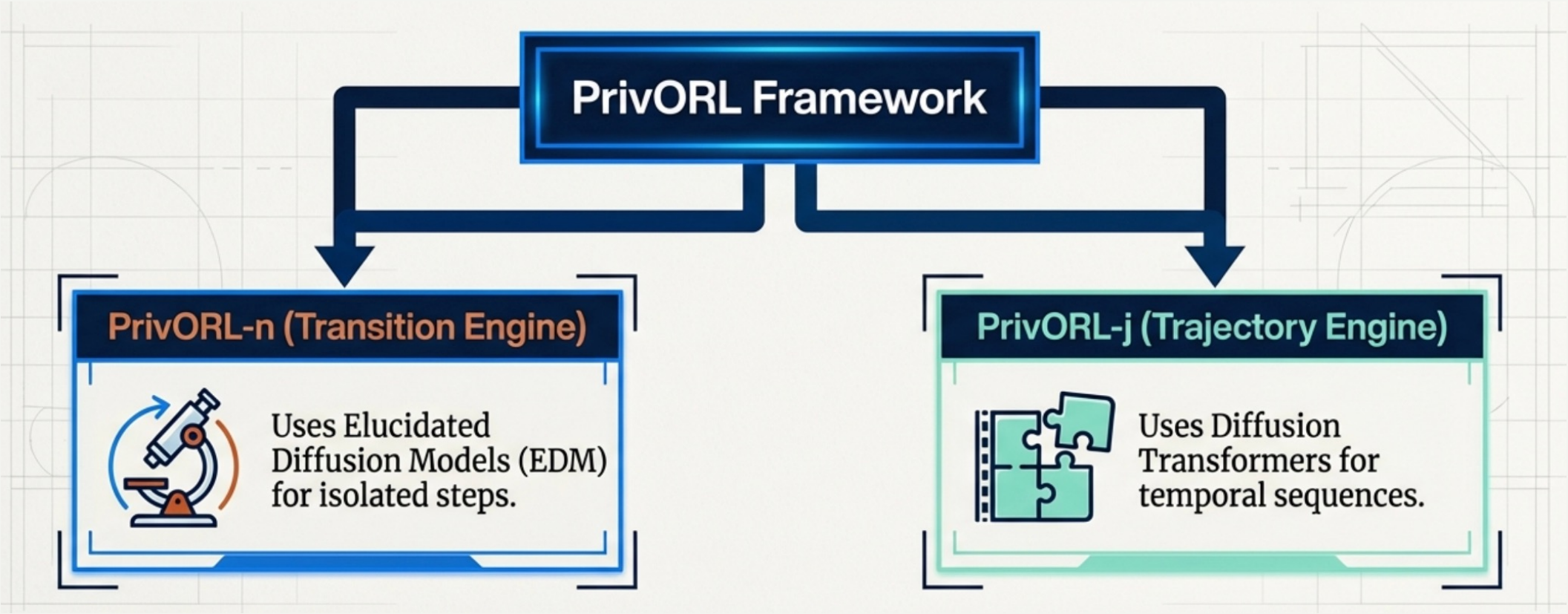
PrivORL-n: Specialized for Transition-level synthesis (single steps).
PrivORL-j: Specialized for Trajectory-level synthesis (full sequences).

PrivORL: A Unified framework for Private Synthesis

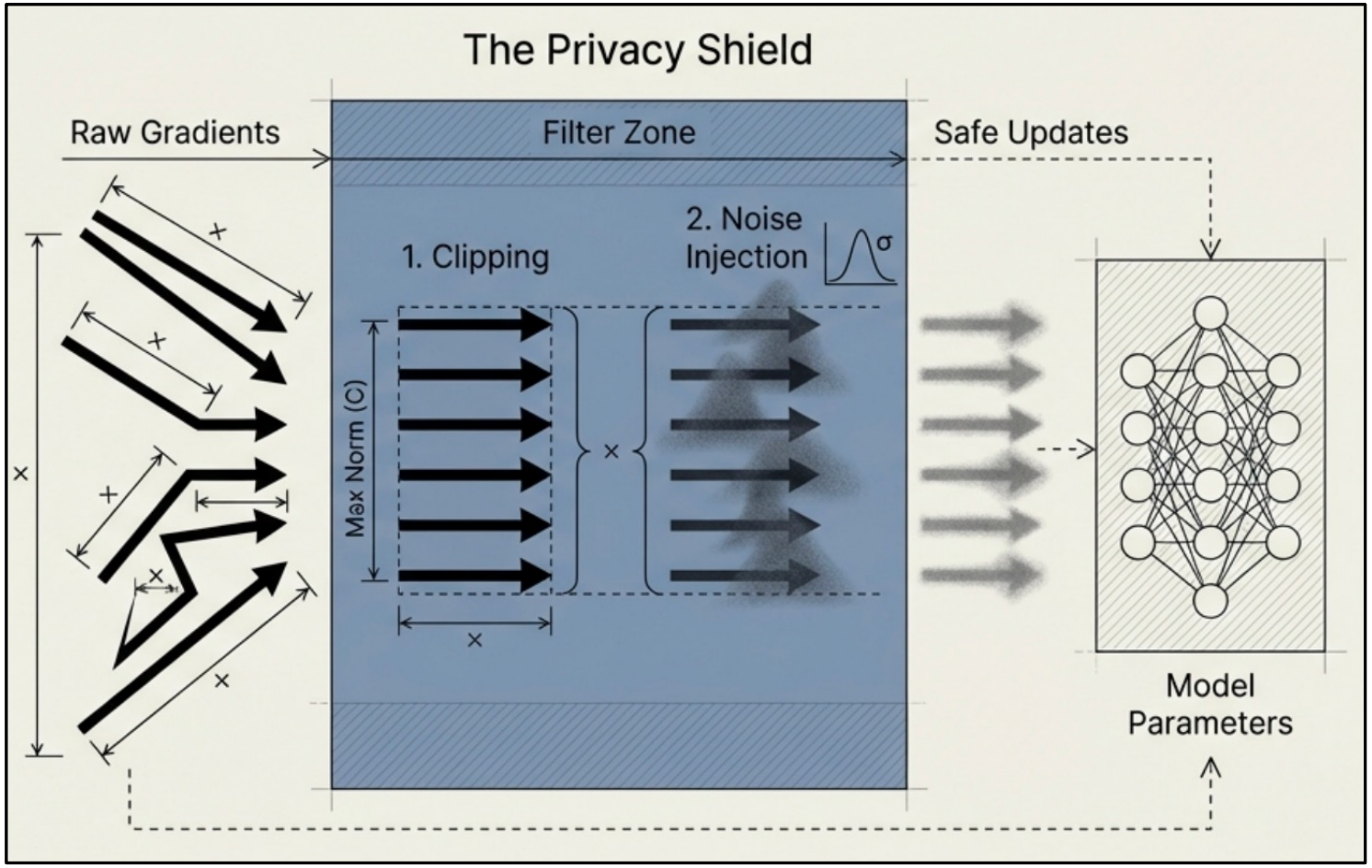


PrivORL-n: Specialized for Transition-level synthesis (single steps).
PrivORL-j: Specialized for Trajectory-level synthesis (full sequences).

PrivORL: A Unified framework for Private Synthesis



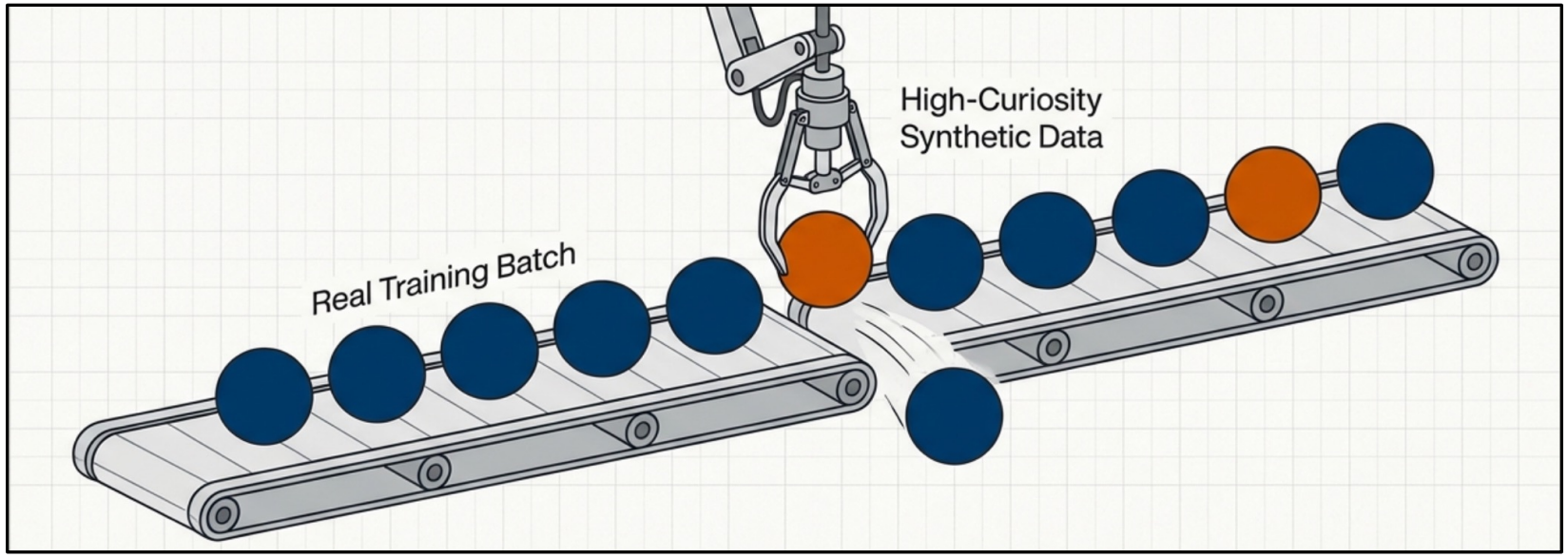
PrivORL: A Unified framework for Private Synthesis



Mechanism: DPSGD

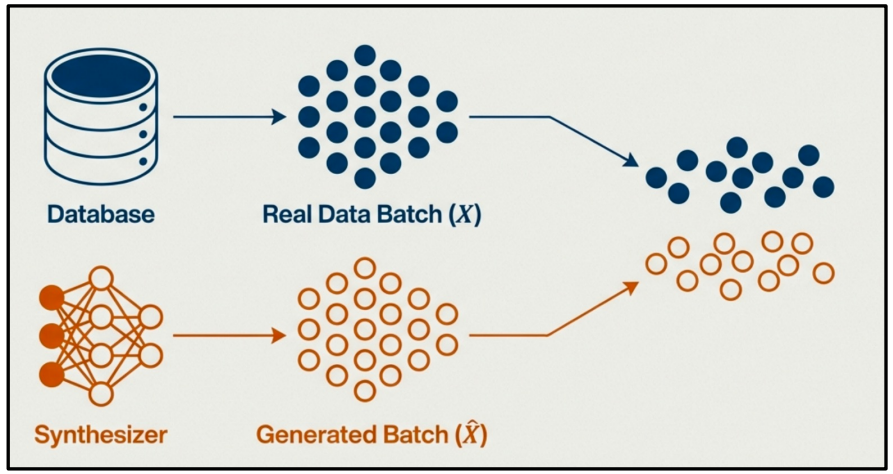
- Gradients are clipped to a maximum norm C.
- Gaussian noise is added during updates.

Curiosity as a Mechanism for Diversity



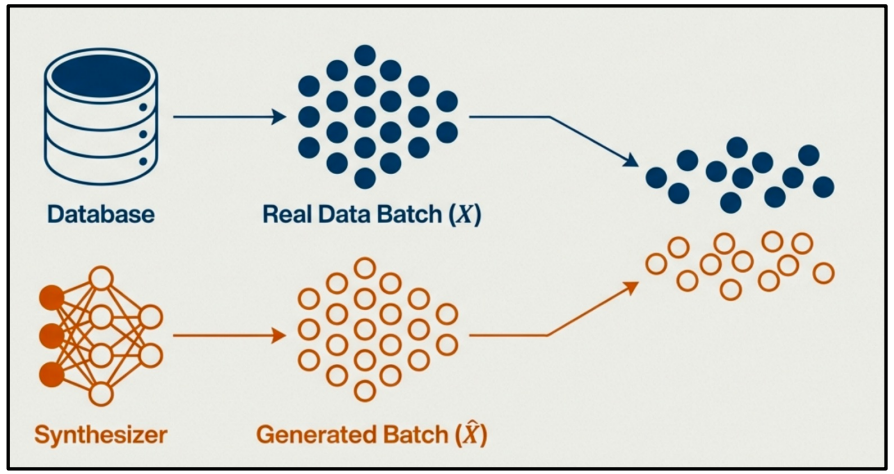
To force them to learn diversity, we manipulate the training data itself. We inject the model‘s own most “curious” creations back into the training batch.

Curiosity Driven Module

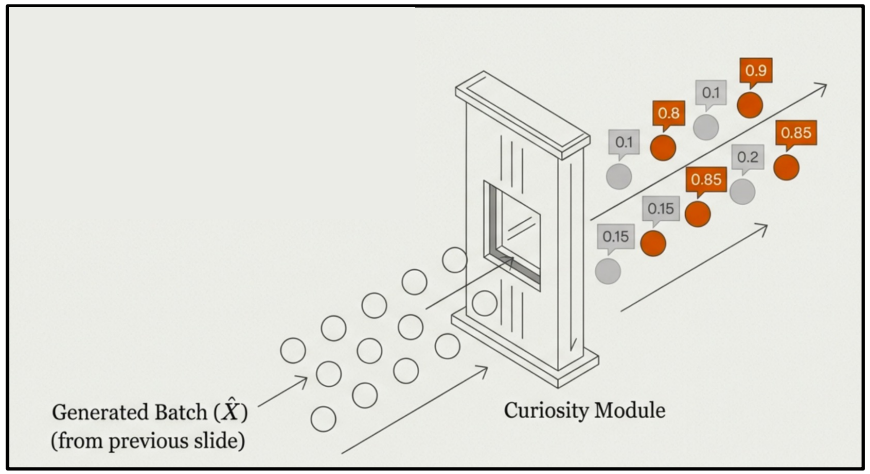


Sampling and Generation

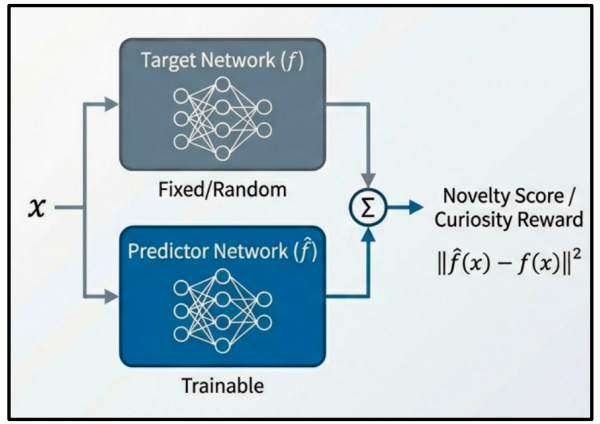
Curiosity Driven Module



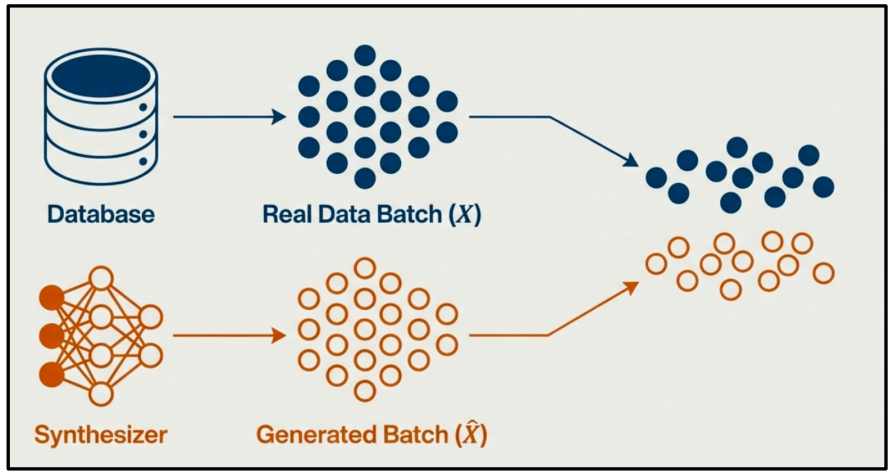
Sampling and Generation



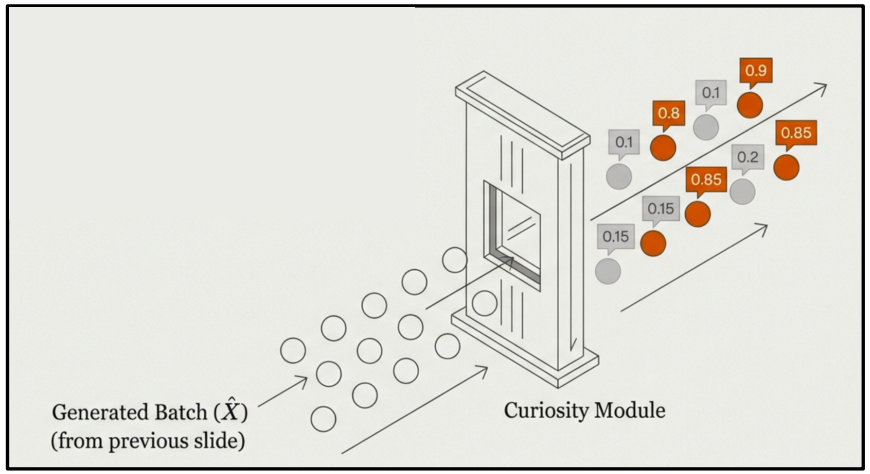
Scoring the Synthetic Data



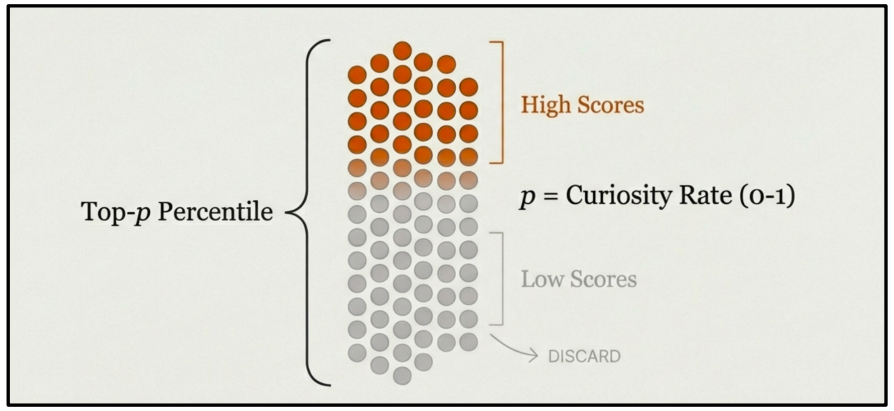
Curiosity Driven Module



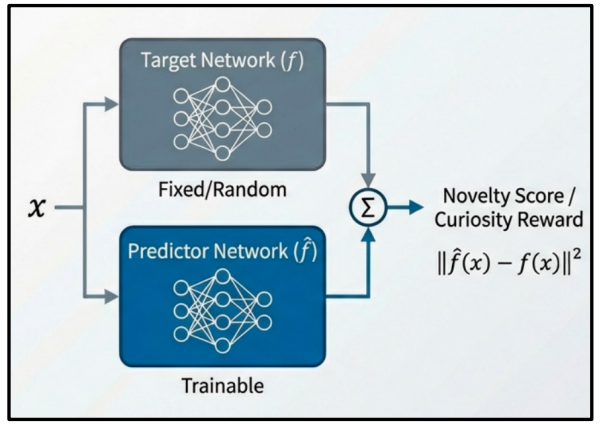
Sampling and Generation



Scoring the Synthetic Data

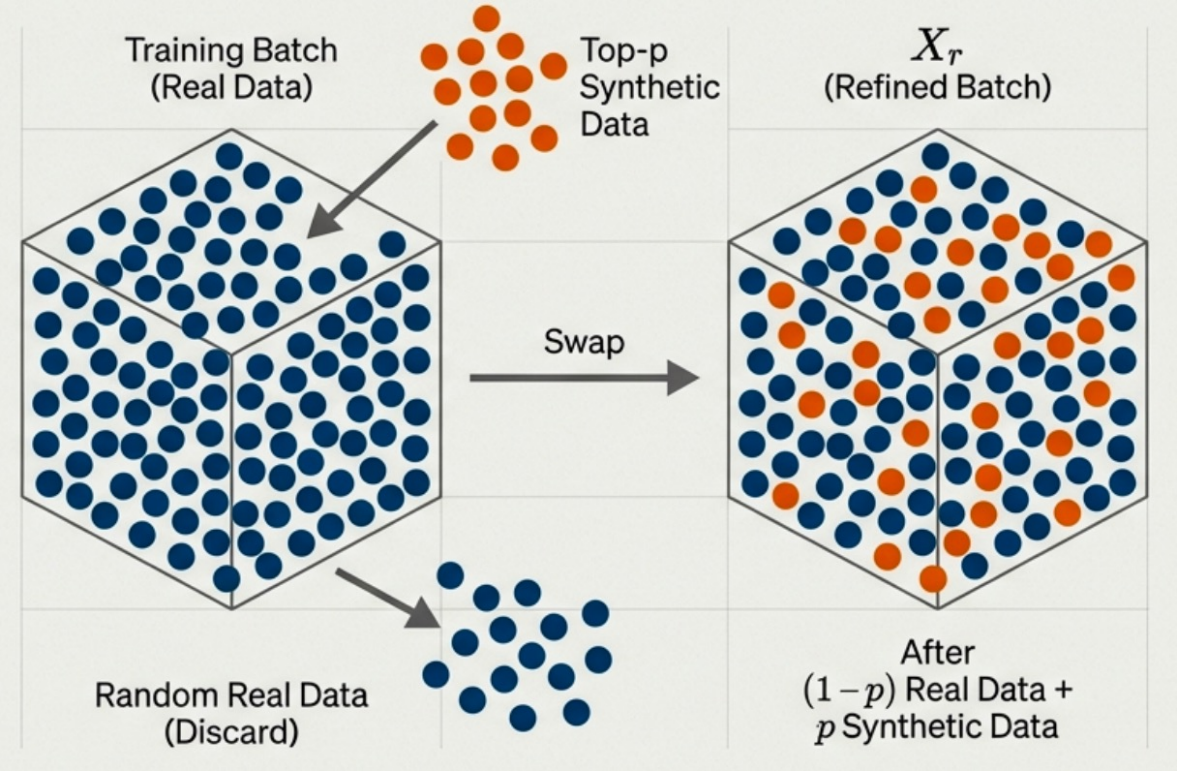


Filter

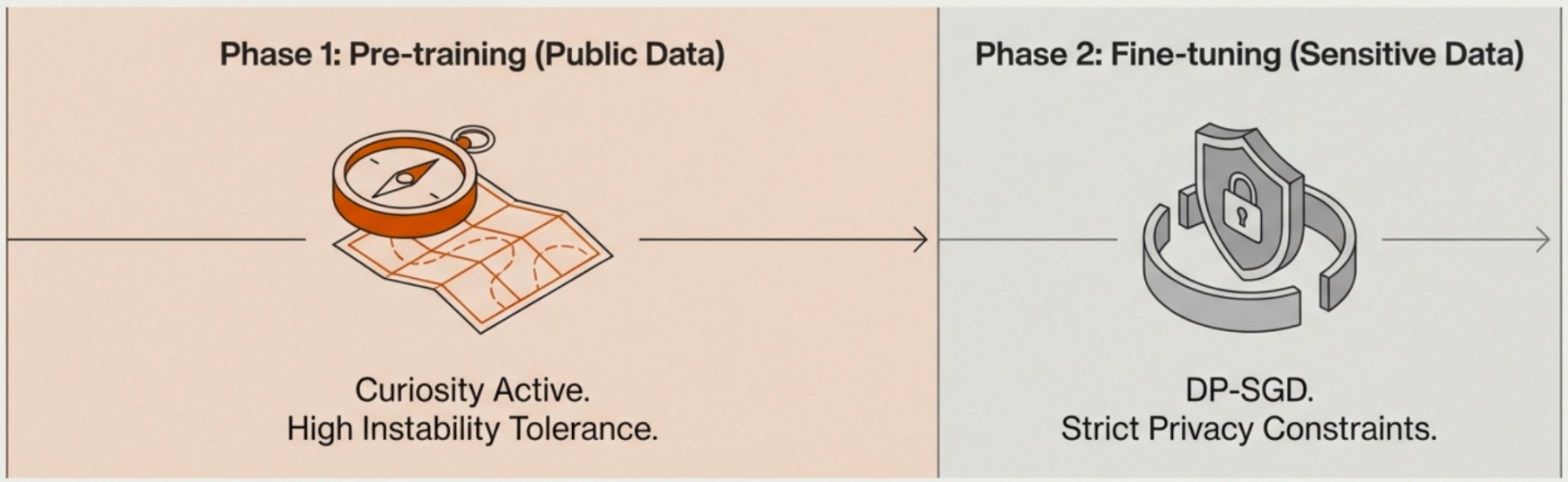


Curiosity Driven Module

We replace a portion of the real data with our high-curiosity synthetic samples.



Pretraining not Finetuning

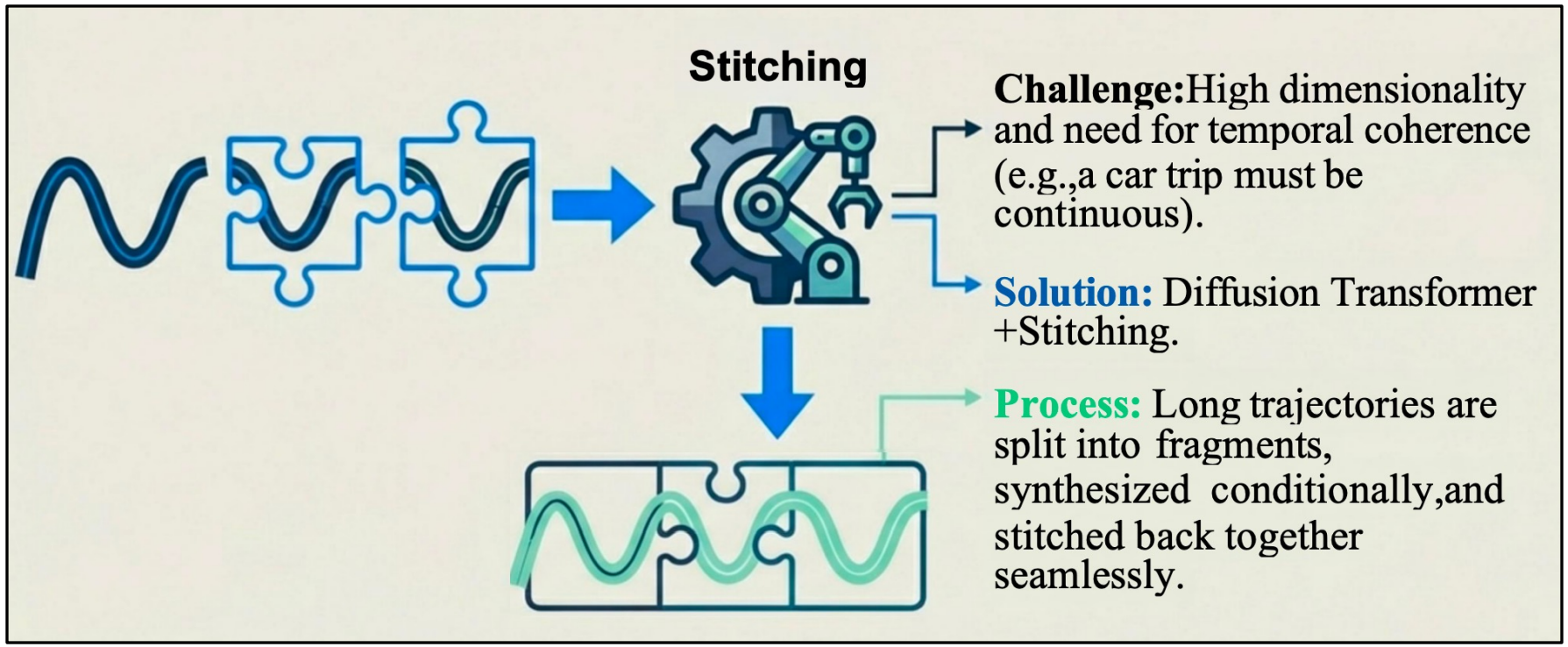


Method	Maze			Kitchen	Mujoco
	umaze	medium	large	umaze	halfcheetah
FineCurPrivTranR	54.1	65.9	54.1	5.2	18.7
Ours	70.3	90.7	81.0	25.5	36.9

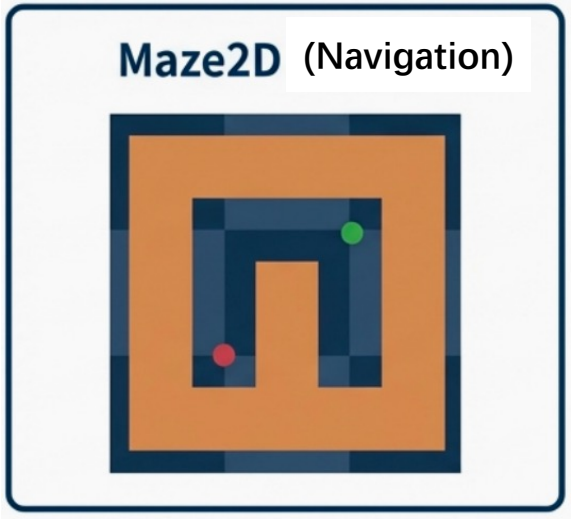
PrivORL-n: Modeling Isolated Decisions (Transition-Level)

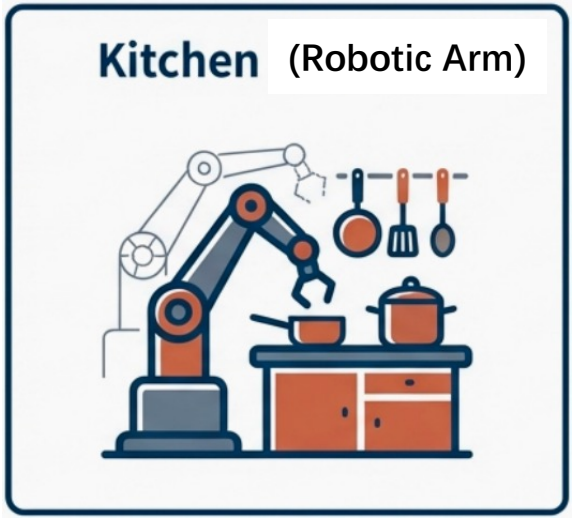
Uses Elucidated Diffusion Models (EDM)to model each transition independently.
Ideal for datasets where immediate feedback loops are the priority.

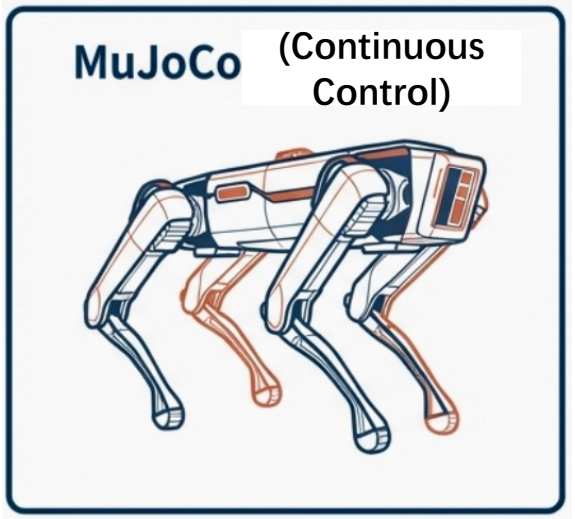
PrivORL-j: Capturing Temporal Dependencies (Trajectory-Level)



Benchmark







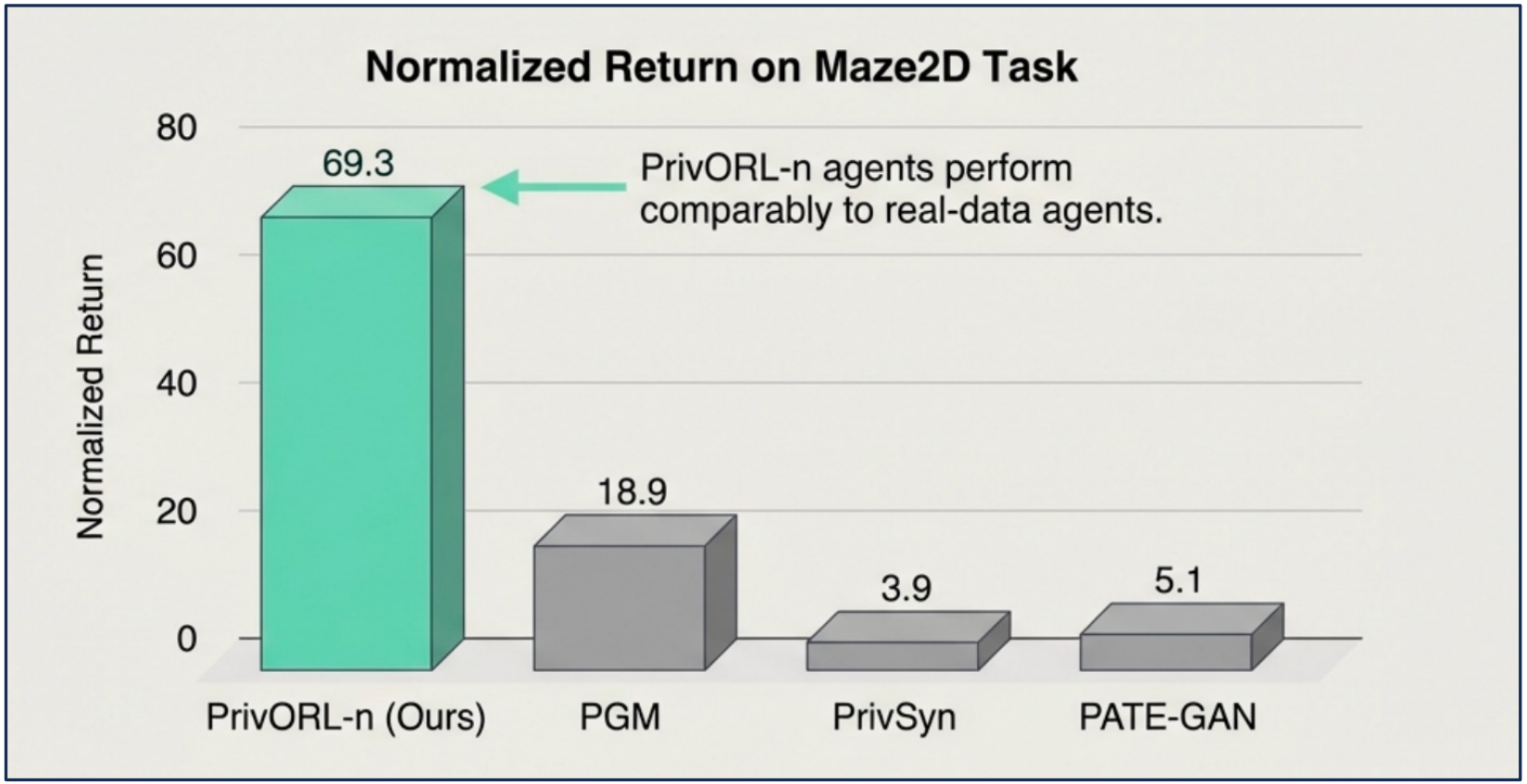
Evaluation Metrics: Averaged Return

$$\frac{1}{|\mathcal{T}|}\sum_{\tau\in\mathcal{T}} R(\tau) \qquad R(\tau)=\sum_{i=0}^{|\tau|}\gamma_i r_i$$

$$\tau : (\langle s_0, a_0, r_0\rangle, \langle s_1, a_1, r_1\rangle, \cdots, \langle s_{|\tau|}, a_{|\tau|}, r_{|\tau|}\rangle)$$

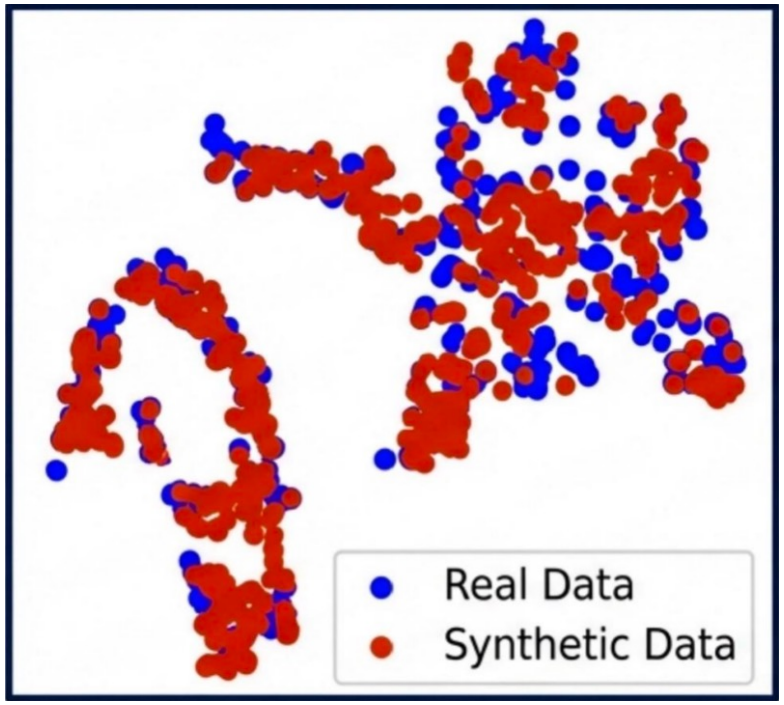
Utility Results: Superior Performance

In Maze2D tasks (epsilon=10), PrivORL significantly outperforms state-of-the-art baselines.

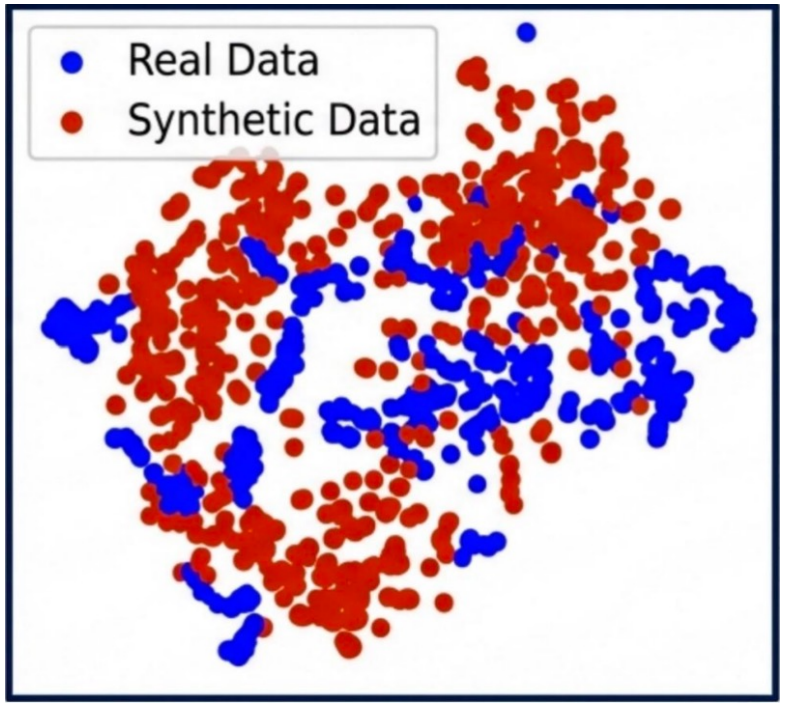


Fidelity Results: Mirroring the Real Distribution

PrivORL vs Real



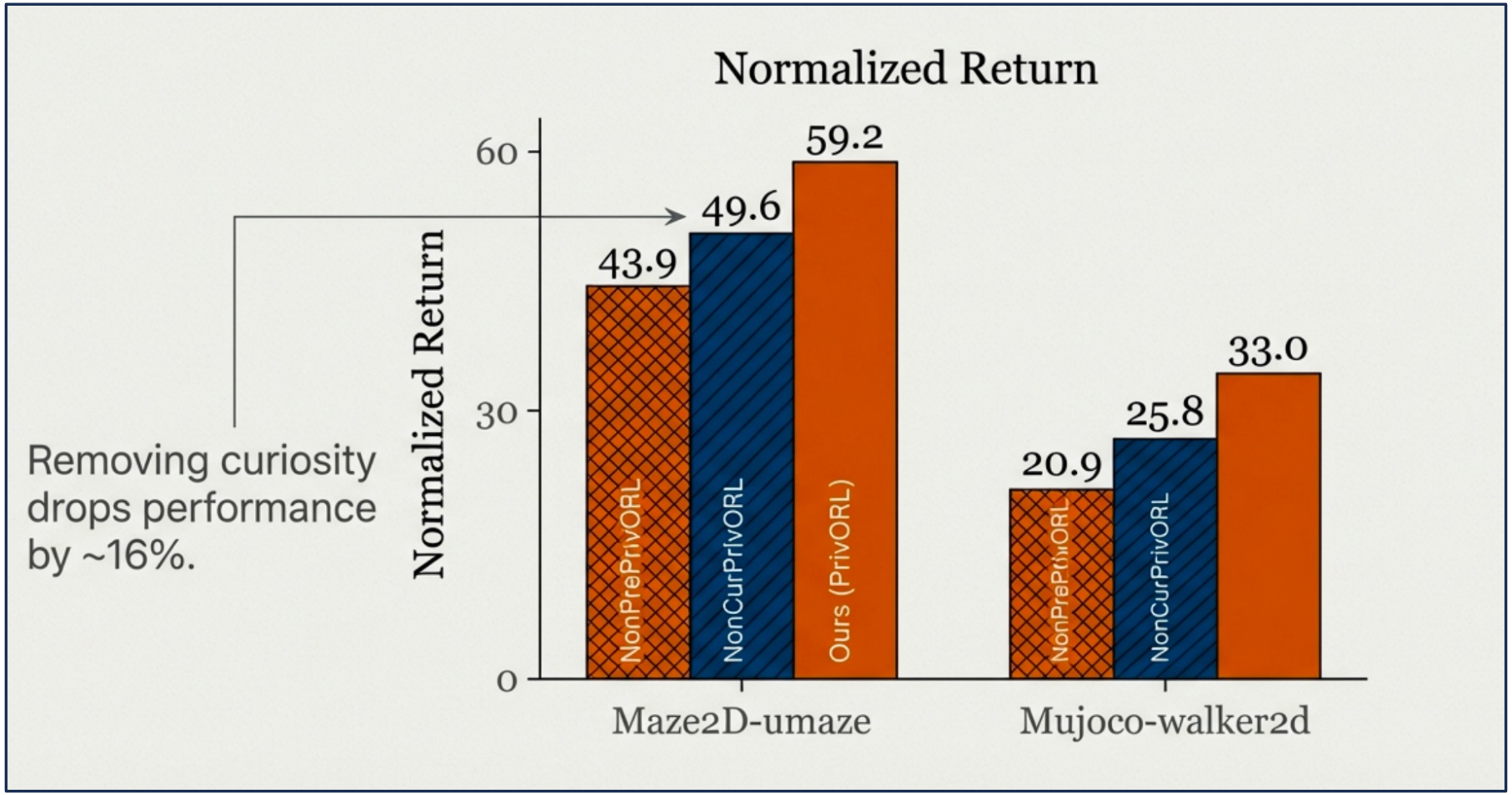
Baseline (PATE-GAN) vs Real



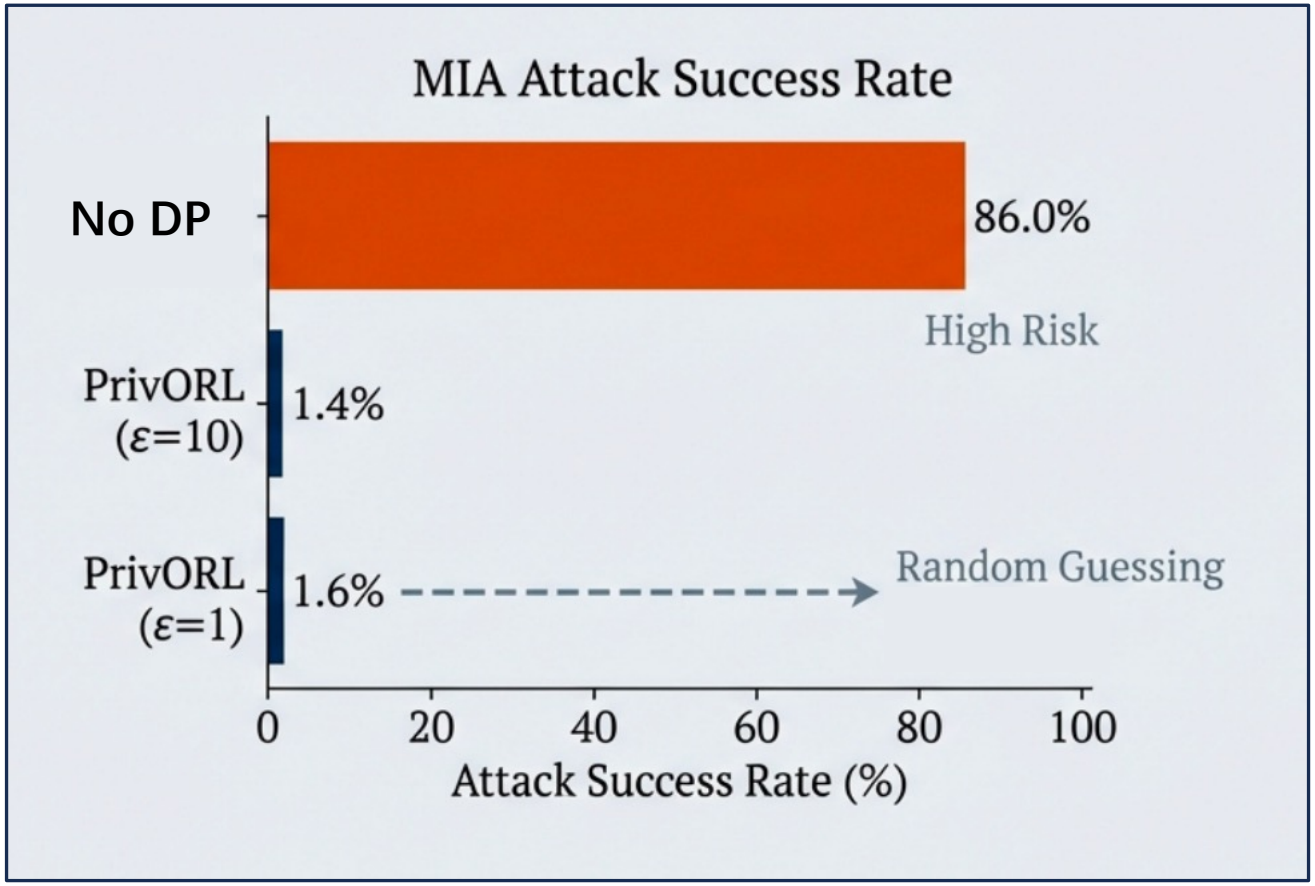
Marginal Statistics: >0.92 (High similarity in single variables).

Correlation: >0.97 (Preserves relationships between state, action, and reward).

Curiosity Creates Better Synthetic Datasets



Empirical Privacy Attac



Threat Model: White-box Membership Inference Attack (MIA).

Result: Even under a relaxed privacy budget of $\varepsilon = 10$,the attack success rate is reduced to near-random guessing levels.

Impact: PrivORL effectively severs the link between synthetic data and the original user records, preventing risks of private information.

This paper proposed a uniform DP offline RL dataset (including both transition and trajectory) synthesis method.

Hope it inspires!

Questions are welcome 😊!

chengong@virginia.edu

Paper

Artifact