ICR-RL: Deep Reinforcement Learning via In-Context-Regression

A Gradient-Free Framework Leveraging Foundational In Context Regression Models for Reinforcement Learning

David Schiff, Ofir Lindenbaum, Yonathan Efroni

01

The Paradigm Shift

Exploring a simpler alternative to Reinforcement Learning via a reduction to regression.

The Deep RL Bottleneck

Optimization Complexity

Standard Deep RL relies on backpropagation, making it memory-intensive and slow to adapt to new environments.

Hyperparameter Sensitivity

Algorithms like PPO and DQN are brittle, requiring precise tuning of clipping ratios and update frequencies to avoid divergence.

Goal: A Foundation Model (FM) capable of zero-shot adaptation without per-task backpropagation.

Core Idea: Regression Oracle

We treat the Reinforcement Learning problem as a pure **In-Context Regression (ICR)** task.

Using **TabPFN/TabICL**, a transformer pre-trained on synthetic tabular data, we approximate the optimal Q-function directly from transition context.

The ICR-RL Framework

ICR-FQI

An inference-only procedure to estimate value functions by iteratively refitting data without weight updates.

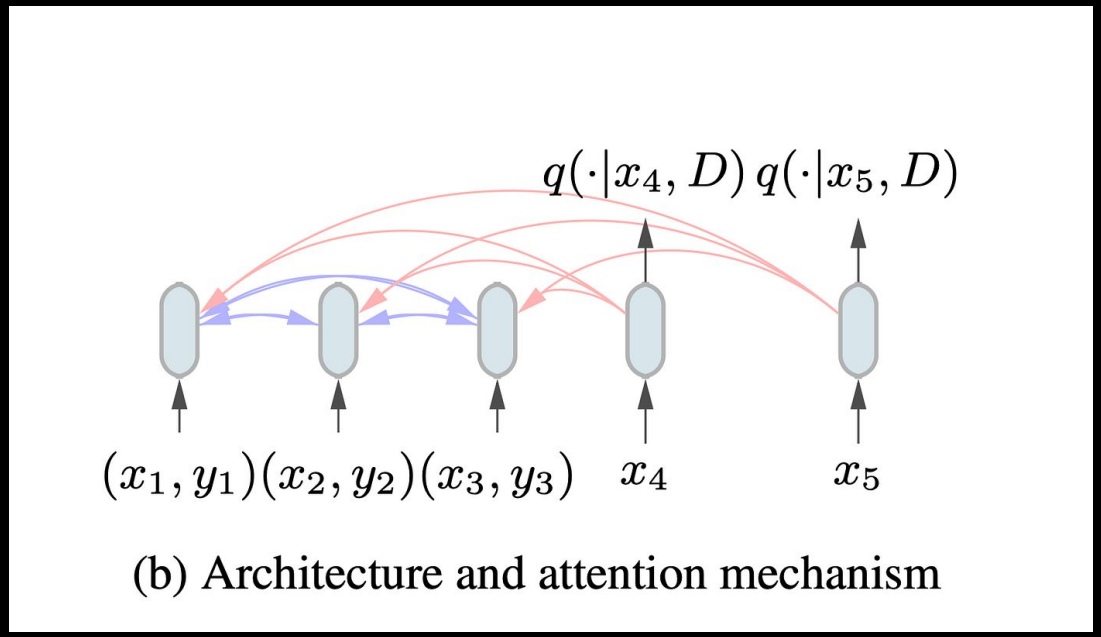
In Context Regression Backbone

A transformer pre-trained on millions of synthetic tasks serving as a general-purpose regression oracle.

Context Mgmt

Sophisticated engineering to retain high-reward transitions and prune redundant data for efficiency.

Gradient-Free Results



(b) Architecture and attention mechanism

0

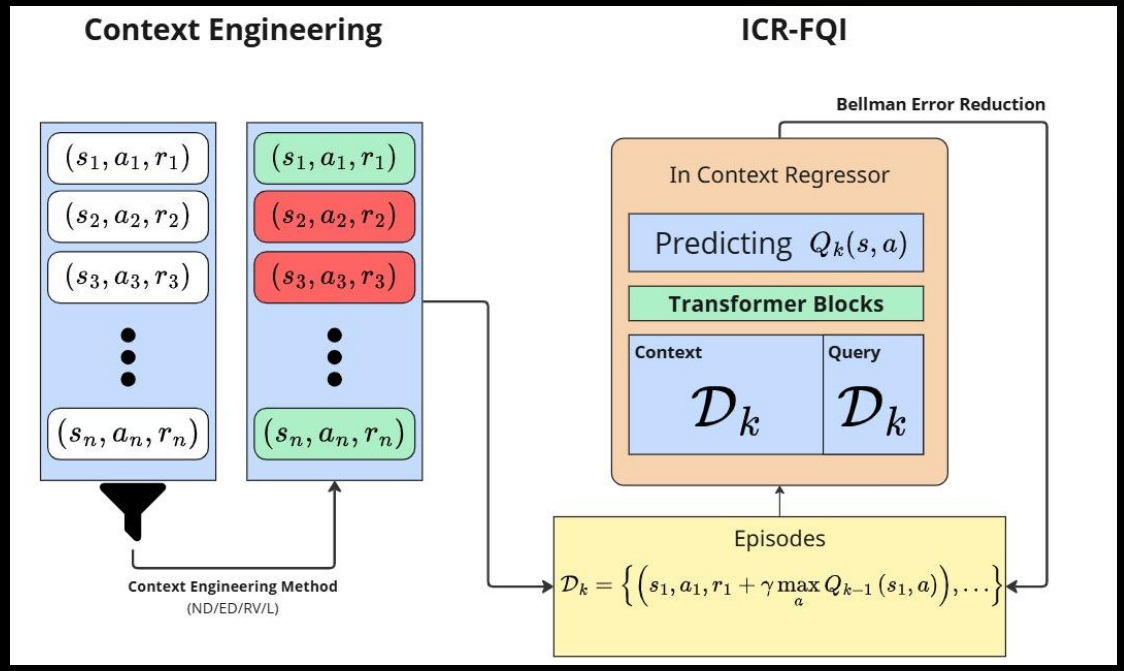
Zero Gradient Updates

ICR-RL achieves competitive performance in classic control environments (CartPole, MountainCar, Acrobot) without a single backpropagation step.

This proves that **In-Context Learning** is a viable path for Reinforcement Learning foundation models.

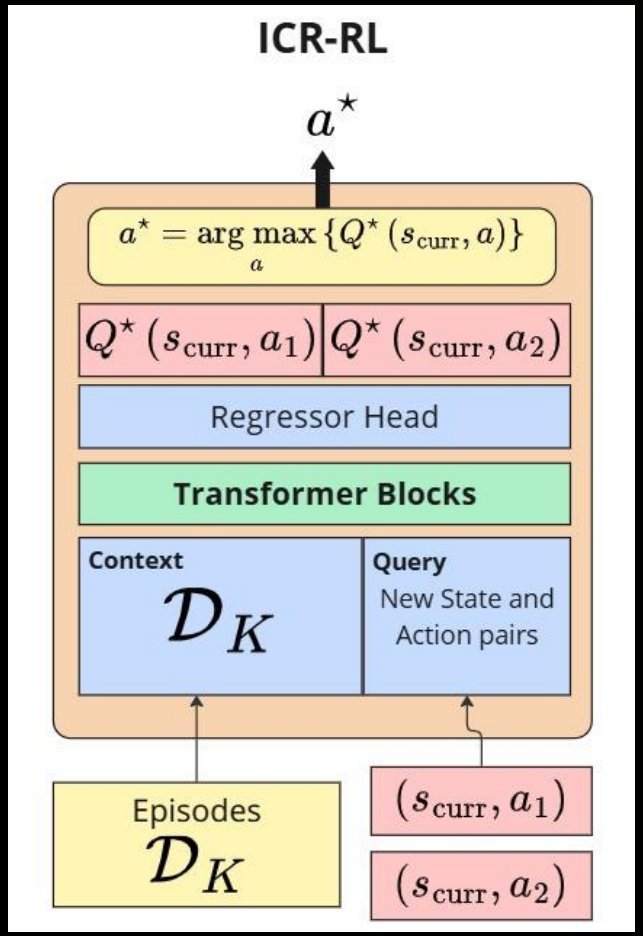
How ICR-RL Works

- **Context Buildup:** We collect trajectories of interactions of the agent with the environment
- **Context Engineering:** We use various techniques to reduce uninformative episodes and free up “learning space” in the context
- **In Context Learning FQI:** We run our in context regression iteratively on our context data to predict the Q values and feed back the outputs into the inputs until we converge

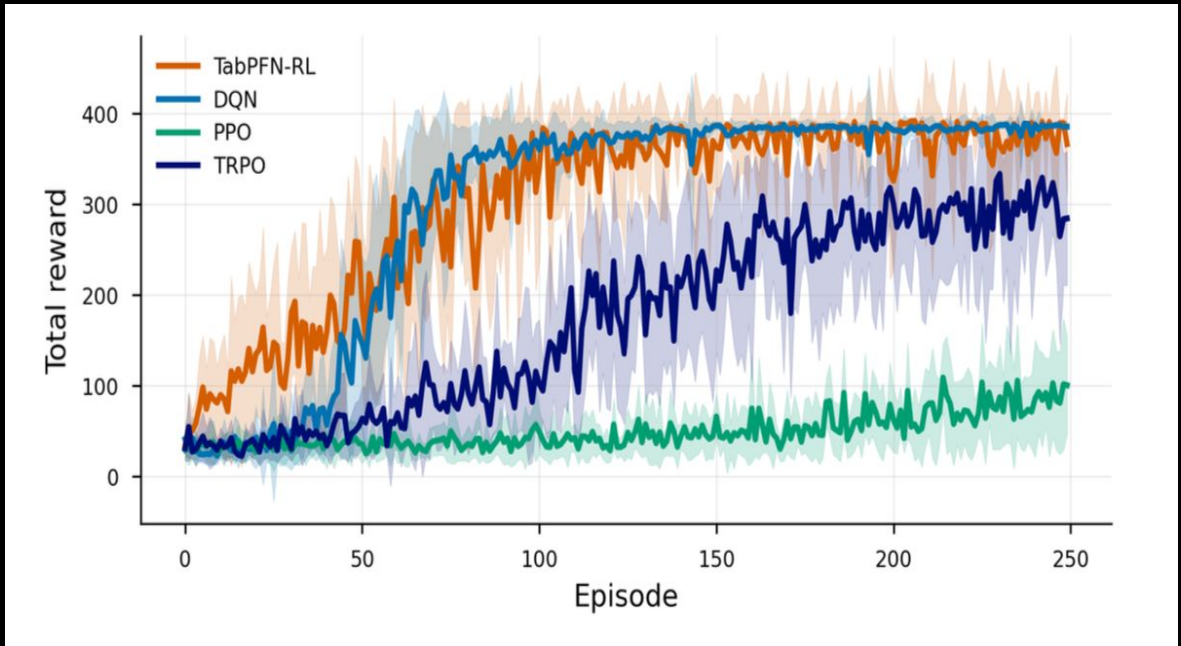


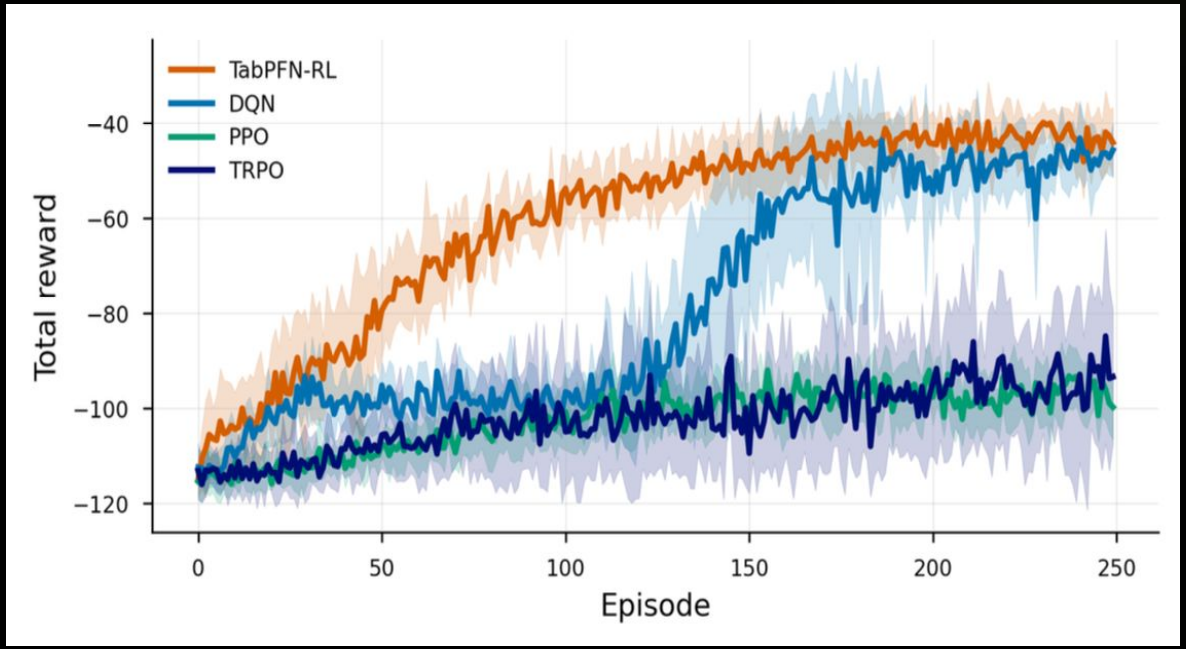
How ICR-RL Works

- **Action Selection:** When running the ICR-RL agent, we select an action by plugging into our context the next possible states and actions and selecting the state, action pair with the highest Q value
- **Repeat:** The data collected from the trajectory will be further used to improve our agent.



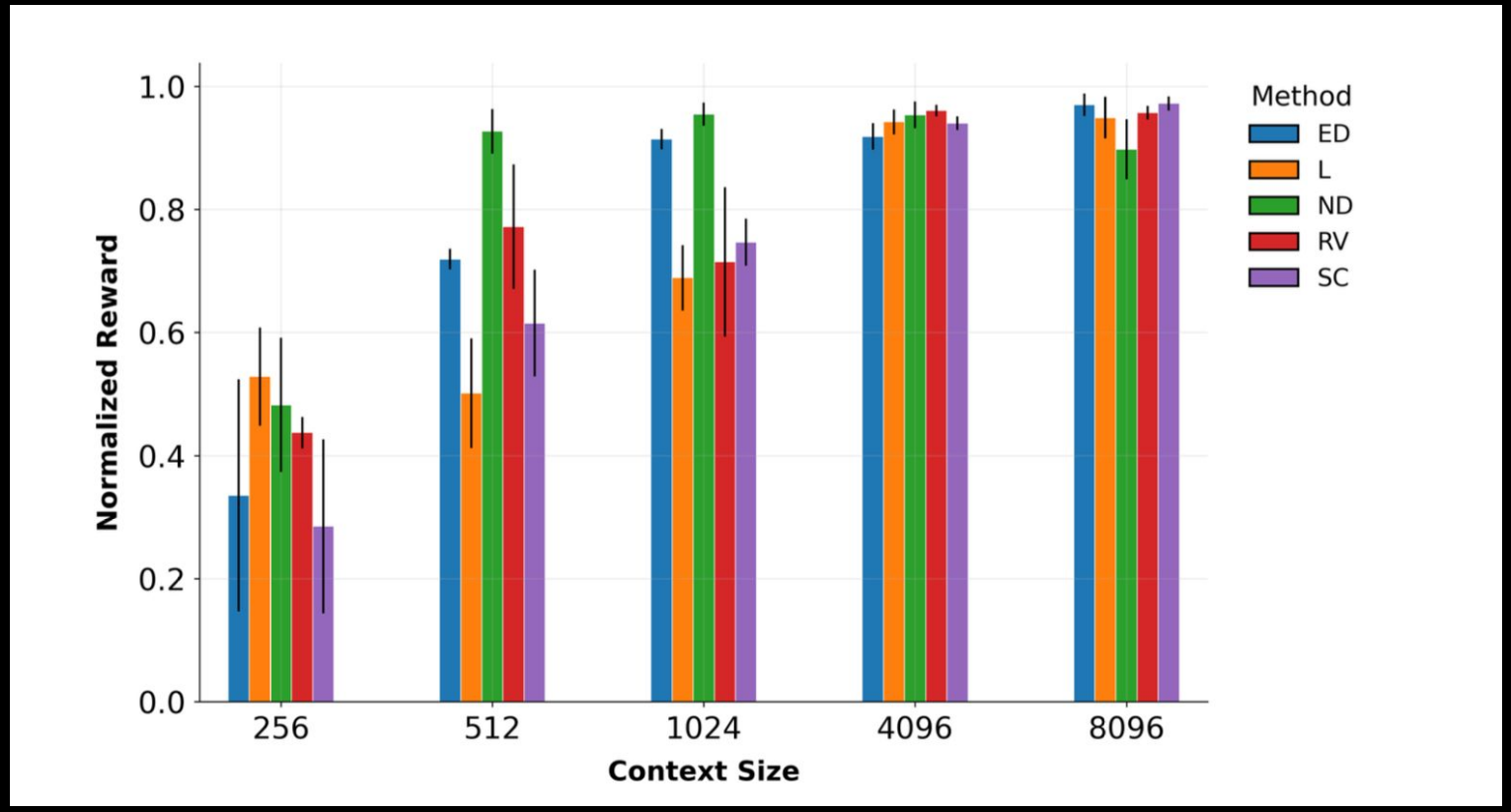
Performance Benchmark





Results indicate that ICR-RL can match or exceed established baselines in classic control benchmarks.

Context Engineering Trends



De-duplication (ND): Identifies near-duplicate pairs to keep the context informative yet compact.

Efficiency: Reduced execution time by ~40% while maintaining performance peaks.

Strategic Comparison

Method	Gradient-Free	Zero-Shot	No RL Pretraining	Continuous States	Online Interaction
DQN [12] / PPO [17] / TRPO [16]	✗	✗	✓	✓	✓
Decision Transformer [3]	✗	✗	✗	✓	✗
Trajectory Transformer [9]	✗	✗	✗	✓	✗
RL^2 [6]	✗	✗	✗	✓	✓
Algorithm Distillation [10]	✗	✗	✗	✓	✓
ICQL [22]	✗	✗	✗	✓	✗
OmniRL [21]	✓	✓	✗	✗	✓
ICR-RL (ours)	✓	✓	✓	✓	✓

Conclusion

"Improvements in general-purpose Regression FMs can directly translate to improvements in Reinforcement Learning."

Questions & Discussion

Thank you for your time. For more details, explore the ICR-RL framework and TabPFN backbone.

davidschiff100@gmail.com