International Platform of Registered Systematic Review and Meta-analysis Protocols

INPLASY

The Exploration Bottleneck: A Review of Strategies in Deep Reinforcement Learning for Sparse Reward Environments

INPLASY202590003
doi: 10.37766/inplasy2025.9.0003
Received: 2 September 2025
Published: 2 September 2025

Hosu, I; Rebedea, T; Trăușan-Matu, Ș.

Corresponding author:
Traian Rebedea

traian.rebedea@upb.ro

Author Affiliation:
Universitatea Națională de Știință și Tehnologie POLITEHNICA București.

ADMINISTRATIVE INFORMATION

Support - Not applicable.

Review Stage at time of this submission - Completed but not published.

Conflicts of interest - Author Traian Rebedea was employed by the company NVIDIA. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

INPLASY registration number: INPLASY202590003

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 2 September 2025 and was last updated on 2 September 2025.

INTRODUCTION

Review question / Objective This review has four main objectives. First, we aim to identify and categorize the major exploration strategies proposed for sparse-reward deep reinforcement learning (DRL), including approaches based on count-based novelty, curiosity and prediction error, pseudo-counts, episodic or impact-driven signals, evolutionary and parameter-space noise, hierarchical goal structures, imitation learning, and model-based exploration.

Second, we examine how these strategies perform on hard-exploration games in the Atari Learning Environment under current evaluation protocols.

Third, we analyze common architectural and algorithmic patterns that interact with exploration, including distributional value functions, recurrent memory, world models, and goal-conditioning.

Finally, we aim to highlight the key open challenges in exploration for sparse-reward environments, including issues of generalization, sample efficiency, and reproducibility, as well as limitations in current measurement practices.

Condition being studied No conditions.

METHODS

Search strategy Representative queries (Boolean OR across synonyms) that we used are:
- ("deep reinforcement learning" OR "RL") AND (exploration OR "sparse reward" OR "hard-exploration")
- (count OR pseudo-count OR density) AND (exploration) AND (Atari)

- (curiosity OR "intrinsic motivation" OR "prediction error" OR ICM OR RND OR RIDE)
- (episodic curiosity OR reachability OR impact-driven)
- (Go-Explore OR NGU OR Agent57 OR "Bootstrapped DQN" OR NoisyNet OR "parameter noise")
- (imitation OR demonstration OR "from demonstrations" OR "human preferences")
- (model-based OR "world model" OR MuZero OR Dreamer) AND (exploration)

We searched the following sources for the period 2015--2025:
- Digital libraries and indexes: arXiv, OpenReview (ICLR), NeurIPS Proceedings, ICML Proceedings, AAAI Proceedings, JMLR/JAIR, Nature/Science. These include the major venues (conferences, journals, and self archiving services) used by the DRL community.
- Secondary discovery: backward/forward snowballing from included papers.

Participant or population No human oarticipants, review is on reinforcement learning in video games.

Intervention Not applicable.

Comparator Not applicable.

Study designs to be included Not applicable.

Eligibility criteria Inclusion criteria. We included primary research articles published between 2015 and 2025 that focus on exploration strategies in deep reinforcement learning (DRL) for high-dimensional or sparse-reward tasks, such as Atari, Procgen, and 3D control environments. Eligible studies had to explicitly target exploration through mechanisms like novelty or curiosity bonuses, episodic memory, goal generation, bootstrapped uncertainty, parameter-space noise, imitation or demonstration-based exploration, or model-based planning for exploration. Additionally, we included papers that proposed strong baselines or state-of-the-art agents with substantial exploration components—such as NGU, Agent57, Go-Explore, or Rainbow variants—or that primarily analyzed the effects of exploration mechanisms.

Exclusion criteria. We excluded studies that focused exclusively on dense-reward continuous control tasks without any explicit exploration components, as well as papers in pure optimization or systems engineering that lacked direct relevance to exploration. We also excluded robotics-only papers unless the exploration of sparse-reward environments was clearly the central contribution.

Information sources We searched the following sources for the period 2015--2025:
- Digital libraries and indexes: arXiv, OpenReview (ICLR), NeurIPS Proceedings, ICML Proceedings, AAAI Proceedings, JMLR/JAIR, Nature/Science. These include the major venues (conferences, journals, and self archiving services) used by the DRL community.
- Secondary discovery: backward/forward snowballing from included papers.

Main outcome(s) Not applicable.

Additional outcome(s) Not applicable.

Data management Not applicable.

Quality assessment / Risk of bias analysis Not applicable.

Strategy of data synthesis Not applicable.

Subgroup analysis Not applicable.

Sensitivity analysis Not applicable.

Language restriction Not applicable.

Country(ies) involved Romania - Universitatea Națională de Știință și Tehnologie POLITEHNICA București.

Keywords deep reinforcement learning; sparse rewards; intrinsic motivation; imitation learning; representation learning; hard-exploration environments.

Dissemination plans Not applicable.

Contributions of each author
Author 1 - Ionel-Alexandru Hosu Hosu - Authot 1 drafted the manuscript.
Email: ionel.hosu@stud.acs.pub.ro
Author 2 - Traian Rebedea - Author 2 provided review, editing and supervision.
Email: traian.rebedea@upb.ro
Author 3 - Ștefan Trăușan-Matu - Author 3 provided review, editing and supervision.
Email: stefan.trausan@upb.ro