Novelty Assessment Report

Paper: Sample Lottery: Unsupervised Discovery of Critical Instances for LLM Reasoning
PDF URL: https://openreview.net/pdf?id=76OZBE4Rb6
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-07

Abstract

Reinforcement Learning with Verifiable Reward (RLVR) has equipped large language models (LLMs) with the capability of reasoning over complicated logical problems through policy optimization. However, conventional methods require complete annotation of the entire dataset and allocate computation uniformly over all samples. We articulate the lottery sample hypothesis in policy optimization of LLMs: a large training set contains a small subset that, when trained alone, yields performance comparable to that of the full dataset. This paper therefore explores the following question: How can we identify these lottery-winning samples from the original dataset without access to answers? Unlike prior efforts that analyze the effect of different samples in the training set with complete annotation, this paper focuses on the unsupervised discovery of critical instances for LLM reasoning and proposes a novel framework termed Complementary Conformal Selection (CONST). Specifically, CONST evaluates the importance of samples by considering two complementary components: procedural volatility and outcome volatility. Procedural volatility measures the potential variations during the LLM's reasoning process, while outcome volatility captures inconsistencies in the final answer. Subsequently, conformal prediction is used to obtain a prediction set whose cardinality serves as the criterion for selecting the lottery-winning samples for annotation. We also provide a theoretical analysis, showing that CONST can effectively approximate the optimal policy. Extensive experiments on various LLMs across different datasets demonstrate the effectiveness of CONST.

Core Task Landscape

This paper addresses: **Unsupervised Discovery of Critical Instances for LLM Reasoning**

A total of **20 papers** were analyzed and organized into a taxonomy with **12 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Instance Selection and Sample Efficiency**
- **Knowledge Integration and Retrieval**
- **Internal Representation Analysis and Manipulation**
- **Unsupervised Structure and Pattern Extraction**
- **Domain-Specific Unsupervised Applications**

Complete Taxonomy Tree

- Unsupervised Discovery of Critical Instances for LLM Reasoning Survey Taxonomy
- Instance Selection and Sample Efficiency
  - Critical Instance Discovery for Policy Optimization ★ (2 papers)
  - [0] Sample Lottery: Unsupervised Discovery of Critical Instances for LLM Reasoning (Anon et al., 2026) View paper
  - [15] Reasoning under Uncertainty: Efficient LLM Inference via Unsupervised Confidence Dilution and Convergent Adaptive Sampling (Zhenning Shi, 2025) View paper
  - In-Context Example Mining (1 papers)
  - [17] Effective Self-Mining of In-Context Examples for Unsupervised Machine Translation with LLMs (Abdul-Mageed, 2024) View paper
- Knowledge Integration and Retrieval
  - Retrieval-Augmented Inference (1 papers)
  - [1] Rethinking with Retrieval: Faithful Large Language Model Inference (He, 2023) View paper
  - Model Selection and Routing (2 papers)
  - [2] Universal model routing for efficient llm inference (Jitkrittum, 2025) View paper
  - [9] Leveraging LLMs for Unsupervised Dense Retriever Ranking (Khramtsova, 2024) View paper
- Internal Representation Analysis and Manipulation
  - Activation-Based Intervention (2 papers)
  - [4] Inference-time intervention: Eliciting truthful answers from a language model (Li Kenneth, 2023) View paper
  - [10] Discovering bias in latent space: An unsupervised debiasing approach (Adila, 2024) View paper
  - Unsupervised Reasoning Behavior Discovery (2 papers)
  - [14] Challenges with unsupervised LLM knowledge discovery (Farquhar, 2023) View paper
  - [16] Fantastic Reasoning Behaviors and Where to Find Them: Unsupervised Discovery of the Reasoning Process (Zhenyu Zhang, 2025) View paper
- Unsupervised Structure and Pattern Extraction
  - Log Parsing and Template Inference (3 papers)
  - [7] No More Labelled Examples? An Unsupervised Log Parser with LLMs (Junjie Huang, 2025) View paper
  - [12] LUNAR: Unsupervised LLM-based log parsing (Huang JunJie, 2024) View paper

- [20] Benchmarking Extraction of Structured Data from Templatized Documents (M Hasan, 2025) View paper
- Relation and Grammar Inference (3 papers)
- [5] Let's Discover More API Relations: A Large Language Model-Based AI Chain for Unsupervised API Relation Inference (Qing Huang, 2024) View paper
- [11] CAM: A Large Language Model-based Creative Analogy Mining Framework (Bhavya Bhavya, 2023) View paper
- [13] A logical word embedding for learning grammar (Deyo, 2023) View paper

- Domain-Specific Unsupervised Applications
  - Educational Content Generation (1 papers)
  - [3] Unsupervised distractor generation via large language model distilling and counterfactual contrastive decoding (Qu, 2024) View paper
  - Social and Behavioral Inference (2 papers)
  - [6] Evaluating large language models for user stance detection on X (Twitter) (Margherita Gambini, 2024) View paper
  - [19] HiCoTraj: Zero-Shot Demographic Reasoning via Hierarchical Chain-of-Thought Prompting from Trajectory (Xie, 2025) View paper
  - Multimodal Disambiguation (1 papers)
  - [18] Quantum Visual Word Sense Disambiguation: Unraveling Ambiguities Through Quantum Inference Model (Wenbo Qiao, 2025) View paper
  - Policy Learning from Language (1 papers)
  - [8] RLZero: Direct Policy Inference from Language Without In-Domain Supervision (Sikchi, 2024) View paper

Narrative

Core task: unsupervised discovery of critical instances for LLM reasoning. The field organizes around five main branches that reflect different strategies for improving LLM performance without extensive labeled data. Instance Selection and Sample Efficiency focuses on identifying high-value training or inference examples, often through methods that prioritize informative samples or optimize policy learning. Knowledge Integration and Retrieval emphasizes augmenting models with external information sources, such as retrieval-based approaches like Rethinking with Retrieval[1] that dynamically incorporate relevant context. Internal Representation Analysis and Manipulation examines how models encode and process information internally, enabling interventions such as Inference Time Intervention[4] or probing latent structures as in Bias in Latent Space[10]. Unsupervised Structure and Pattern Extraction targets the automatic discovery of regularities, relations, or templates from unstructured data, exemplified by works like API Relations Discovery[5] and Unsupervised Log Parser[7]. Finally, Domain-Specific Unsupervised Applications adapt these techniques to specialized settings, ranging from creative tasks like Creative Analogy Mining[11] to trajectory analysis in HiCoTraj[19].

A particularly active line of work within Instance Selection explores how to identify critical training instances that disproportionately influence model behavior, balancing sample efficiency with policy optimization. Sample Lottery[0] sits squarely in this space, proposing mechanisms to discover instances that act as pivotal learning signals for reasoning tasks. This emphasis contrasts with nearby efforts like Confidence Dilution[15], which examines how instance selection interacts with model uncertainty, and RLZero[8], which integrates reinforcement learning to refine instance prioritization. Meanwhile, works in Unsupervised Structure and Pattern Extraction, such as Unsupervised Distractor Generation[3], highlight complementary challenges in mining structural patterns without supervision. The central tension across these branches revolves around whether to focus on curating better training data, leveraging external knowledge, or directly manipulating internal model states—each offering distinct trade-offs in interpretability, scalability, and domain adaptability.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. Reasoning under Uncertainty: Efficient LLM Inference via Unsupervised Confidence Dilution and Convergent Adaptive Sampling

Authors: Zhenning Shi, Yi-jia Zhu, Yi Xie, Yijia Zhu, Junhan Shi, et al. (10 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

â¦ Comprehensive experiments across three reasoning datasets demonstrate that our approach â¦ provides a scalable, unsupervised solution for reliable and efficient LLM reasoning. â¦

Relationship Analysis

Both papers belong to the Critical Instance Discovery for Policy Optimization category, focusing on unsupervised identification of high-value training samples for LLM reasoning. The original paper (Sample Lottery) discovers critical instances for annotation and training by measuring procedural and outcome volatility in reasoning paths, then uses conformal prediction to select lottery-winning samples for RLVR optimization. The candidate paper addresses a different problem: it calibrates confidence scores during inference through diversity-aware dilution and uses adaptive sampling to reduce computational cost at test time, rather than selecting training instances.

Contributions Analysis

Overall novelty summary. The paper introduces the lottery sample hypothesis for reinforcement learning with verifiable reward in LLMs, proposing that a small subset of training instances can yield performance comparable to the full dataset. It resides in the 'Critical Instance Discovery for Policy Optimization' leaf, which contains only two papers total. This leaf sits within the broader 'Instance Selection and Sample Efficiency' branch, indicating a relatively sparse research direction focused specifically on unsupervised identification of high-value training samples through volatility or uncertainty measures. The small population suggests this is an emerging rather than saturated area.

The taxonomy reveals that neighboring work primarily explores alternative strategies for improving LLM efficiency. The sibling leaf 'In-Context Example Mining' addresses demonstration selection for few-shot learning, while adjacent branches tackle knowledge integration through retrieval-augmented inference and internal representation manipulation via activation-based intervention. The scope note explicitly excludes supervised selection methods and general active learning, positioning this work at the intersection of policy optimization and unsupervised sample prioritization. This boundary clarifies that the contribution targets training-time instance discovery rather than inference-time retrieval or supervised curation.

Among the three contributions analyzed, the lottery sample hypothesis examined ten candidates and found one potentially refutable prior work, suggesting some overlap in the conceptual foundation. The CONST framework examined only two candidates with no clear refutations, indicating limited direct precedent within the search scope. The theoretical analysis component examined ten candidates without refutations. These statistics reflect a search of twenty-two total candidates, not an exhaustive literature review. The framework and theoretical components appear more novel within this limited examination, while the core hypothesis shows measurable prior exploration.

Based on the top-22 semantic matches examined, the work appears to occupy a sparsely populated research direction with modest prior overlap on its foundational hypothesis but less precedent for its specific framework. The limited search scope means potentially relevant work outside these candidates remains unexamined. The taxonomy structure suggests the paper addresses an underexplored intersection

of policy optimization and unsupervised instance selection, though the single refutable finding indicates the core motivation has some established grounding.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Lottery sample hypothesis for RLVR on LLMs

Description: The authors formalize a hypothesis stating that a small subset of training samples can achieve performance comparable to the full dataset when used alone for reinforcement learning with verifiable reward on large language models. This hypothesis motivates the unsupervised discovery of critical instances.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. GENEREIT: generating multi-talented reinforcement learning agents

URL: View paper

Brief Assessment

GENEREIT[26] focuses on meta-reinforcement learning for generating multi-talented agents across different game environments, not on sample selection or subset discovery for language model training with verifiable rewards.

---

2. Sample trajectory selection method based on large language model in reinforcement learning

URL: View paper

Brief Assessment

Trajectory Selection[21] focuses on using LLMs to select high-quality trajectories in traditional RL environments (Gym, RLCard), not on subset selection for RLVR training of LLMs themselves. The domains and problem formulations are fundamentally different.

---

3. Reinforcement learning for reasoning in large language models with one training example

URL: View paper

Prior Art Analysis

One Training Example[25] demonstrates that reinforcement learning with verifiable reward can be effective using as few as one training example, achieving performance comparable to training on thousands of examples. This directly refutes the novelty of the lottery sample hypothesis by showing empirically that a single example (or very few examples) can match full dataset performance in RLVR for LLMs. The candidate paper provides extensive experimental evidence across multiple models and benchmarks, demonstrating that '1-shot rlvr is sufficient to trigger substantial improvements in reasoning tasks, even matching the performance of rlvr with thousands of examples.' This prior work establishes the same core finding that the original paper articulates as their hypothesis.

Evidence

Evidence 1 - **Rationale**: Both papers articulate the same fundamental finding: that a very small subset (even a single example) from a large training set can achieve performance comparable to the full dataset in RLVR for LLMs. The candidate paper provides concrete experimental validation of this phenomenon before the original paper's hypothesis was formalized. - **Original**: we articulate the lottery sample hypothesis in policy optimization of llms: a large training set contains a small subset that, when trained alone, yields performance comparable to that of the full dataset. - **Candidate**: we show that reinforcement learning with verifiable reward using one training example (1-shot rlvr) is effective in incentivizing the mathematical reasoning capabilities of large language models (llms). applying rlvr to the base model qwen2.5-math-1.5b, we identify a single example that elevates mod...

Evidence 2 - **Rationale**: The candidate paper demonstrates the practical implications of using very few examples in RLVR, showing that full annotation is not required and that computation can be concentrated on minimal instances - the exact benefits claimed by the original paper's hypothesis. - **Original**: with this hypothesis, it is possible to break conventional approaches from two aspects (as illustrated in figure 1): (i) full annotation of the dataset is no longer required, and ground truth answers of several lottery-winning samples are sufficient; (ii) computation can be concentrated on several c... - **Candidate**: we find that rlvr with 1 example{π13} (35.7%) performs close to that with 1.2k dsr-sub (35.9%), and rlvr with 2 examples {π1, π13} (36.6%) even performs better than rlvr with dsr-sub and as well as using 7.5k math train dataset (36.7%).

Evidence 3 - **Rationale**: This evidence pair shows that the candidate paper already established through experiments the exact phenomenon that the original paper articulates as their hypothesis - that a small subset (even one example) can match full dataset performance in RLVR. - **Original**: based on these findings, we articulate the lottery sample hypothesis in rlvr of llms: a large training set for rlvr on llms contains a small subset that, when trained alone, can achieve performance comparable to that of the full dataset. - **Candidate**: we find that selecting one specific example as the training dataset can achieve similar downstream performance to that of the 1.2k deepscaler subset (dsr-sub) containing that example. specifically, this improves the qwen2.5-math-1.5b model from 36.0% to 73.6% on math500, and from 17.6% to 35.7% on a...

---

4. Tempera: Test-time prompting via reinforcement learning

URL: View paper

Brief Assessment

Tempera[30] focuses on test-time prompt editing for language models using reinforcement learning, not on sample selection or subset efficiency in RLVR training. The paper addresses a different problem domain (prompt optimization vs. training data selection).

---

5. Improving Data Efficiency for LLM Reinforcement Fine-tuning Through Difficulty-targeted Online Data Selection and Rollout Replay

URL: View paper

Brief Assessment

Difficulty Targeted Selection[28] focuses on adaptive difficulty-based online data selection and rollout replay for computational efficiency, not on identifying a small subset that can achieve comparable performance to the full dataset when trained alone.

---

6. Inference-aware fine-tuning for best-of-n sampling in large language models

URL: View paper

Brief Assessment

Inference Aware Fine Tuning[27] focuses on optimizing inference-time strategies (best-of-n sampling) rather than identifying critical training subsets. The candidate does not address sample selection or subset discovery for training efficiency.

---

7. Prompt optimization with EASE? efficient ordering-aware automated selection of exemplars

URL: View paper

Brief Assessment

EASE[23] focuses on exemplar selection for in-context learning in LLMs, not on sample selection for reinforcement learning with verifiable rewards. The technical domains are fundamentally different.

---

8. Advances in Statistical Inference and Policy Optimization for Reinforcement Learning

URL: View paper

Brief Assessment

Statistical Inference Policy[29] focuses on statistical inference methods and confidence intervals for policy optimization, not on identifying critical training subsets for RLVR on LLMs or the lottery sample hypothesis.

---

9. Not All Rollouts are Useful: Down-Sampling Rollouts in LLM Reinforcement Learning

URL: View paper

Brief Assessment

Down Sampling Rollouts[22] focuses on computational efficiency through strategic rollout subset selection during policy updates, not on identifying critical training samples for annotation. The candidate addresses a different problem (compute/memory asymmetry in updates) rather than unsupervised discovery of lottery-winning samples for minimal annotation.

---

10. Sample efficient preference alignment in LLMs via active exploration

URL: View paper

Brief Assessment

Active Exploration Alignment[24] focuses on preference-based feedback and active selection of prompts/completions for efficient RLHF/DPO training, not on discovering critical subsets within a fixed dataset for RLVR optimization. The technical approaches differ fundamentally: the candidate uses dueling bandits and uncertainty-based acquisition functions, while the original proposes conformal prediction with procedural/outcome volatility for unsupervised sample discovery.

Contribution 2: Complementary Conformal Selection (CONST) framework

Description: The authors propose CONST, a novel framework that identifies critical training instances without requiring ground truth annotations. CONST evaluates sample importance using procedural volatility and outcome volatility, then applies conformal prediction to select lottery-winning samples for annotation and optimization.

This contribution was assessed against **2 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Applying Conformal Prediction for LLM Multi-Label Text Classification

URL: View paper

Brief Assessment

Conformal Prediction Classification[42] applies conformal prediction to multi-label text classification tasks with LLMs, focusing on producing label-wise prediction sets with coverage guarantees. This differs fundamentally from CONST, which uses conformal prediction for unsupervised critical instance selection in RLVR training by evaluating procedural and outcome volatility.

---

2. Financial Time Series: Adaptive Forecasting Frameworks

URL: View paper

Brief Assessment

Adaptive Forecasting Frameworks[41] focuses on financial time series forecasting using regime-switching models and deep learning architectures. It does not address unsupervised critical instance selection for reinforcement learning or conformal prediction for sample selection in LLM training contexts.

Contribution 3: Theoretical analysis of CONST approximating optimal policy

Description: The authors establish a theoretical generalization bound demonstrating that under the lottery sample hypothesis and certain standard assumptions, CONST can effectively approximate the optimal policy parameter setup with sufficiently large question datasets and verified rewards.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Generalized proximal policy optimization with sample reuse

URL: View paper

Brief Assessment

Generalized PPO[32] focuses on off-policy policy optimization with sample reuse in general RL settings, not on sample selection or lottery sample hypothesis for LLM reasoning tasks. The theoretical frameworks address fundamentally different problems.

---

2. RL-Selector: Reinforcement Learning-Guided Data Selection via Redundancy Assessment

URL: View paper

Brief Assessment

RL Selector[33] focuses on data selection for training efficiency via redundancy assessment in supervised learning, not on policy optimization with sample selection in reinforcement learning for LLM reasoning tasks.

---

3. Communication-Efficient Policy Gradient Methods for Distributed Reinforcement Learning

URL: View paper

Brief Assessment

Communication Efficient Policy[37] focuses on distributed policy optimization with communication efficiency in multi-agent/parallel RL settings, not on sample selection or generalization bounds for policy optimization under a lottery sample hypothesis.

---

4. Safety Filtering While Training: Improving the Performance and Sample Efficiency of Reinforcement Learning Agents

URL: View paper

Brief Assessment

Safety Filtering Training[34] focuses on incorporating safety filters into RL training for robotics control tasks, not on theoretical generalization bounds for policy optimization with sample selection in reinforcement learning for LLMs.

---

5. Model-Based Reinforcement Learning for Cavity Filter Tuning

URL: View paper

Brief Assessment

Cavity Filter Tuning[38] focuses on model-based RL for hardware tuning in telecommunications, not on policy optimization with sample selection or generalization bounds for LLM reasoning tasks.

---

6. Reinforcement Learning-Controlled Subspace Ensemble Sampling for Complex Data Structures

URL: View paper

Brief Assessment

Subspace Ensemble Sampling[35] focuses on imbalanced data mining with feature selection and sampling strategies, not on policy optimization with sample selection in reinforcement learning for LLMs or theoretical generalization bounds for such systems.

---

7. Generalizing Off-Policy Learning under Sample Selection Bias

URL: View paper

Brief Assessment

Sample Selection Bias[36] addresses policy generalization under sample selection bias in offline RL settings, focusing on distribution shift between training and target populations. This is fundamentally different from CONST's theoretical analysis of selecting critical training samples for RLVR optimization of LLMs.

---

8. Imbalanced Sample Selection With Deep Reinforcement Learning for Fault Diagnosis

URL: View paper

Brief Assessment

Imbalanced Sample Selection[39] focuses on fault diagnosis in industrial processes using reinforcement learning for sample selection under class imbalance, not on policy optimization with verifiable rewards for LLM reasoning tasks or generalization bounds for such settings.

---

9. Adversarial Defense Mechanisms for Supervised Learning

URL: View paper

Brief Assessment

Adversarial Defense Mechanisms[40] focuses on supervised learning defense strategies using policy gradient methods in reinforcement learning contexts, not on generalization bounds for policy optimization with sample selection in RLVR of LLMs.

---

10. Online difficulty filtering for reasoning oriented reinforcement learning

URL: View paper

Brief Assessment

Online Difficulty Filtering[31] focuses on dynamic difficulty filtering in GRPO for reasoning tasks, not on generalization bounds for sample selection methods like CONST. The theoretical frameworks address different problems.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Sample Lottery: Unsupervised Discovery of Critical Instances for LLM Reasoning View paper
- [1] Rethinking with Retrieval: Faithful Large Language Model Inference View paper
- [2] Universal model routing for efficient llm inference View paper
- [3] Unsupervised distractor generation via large language model distilling and counterfactual contrastive decoding View paper
- [4] Inference-time intervention: Eliciting truthful answers from a language model View paper
- [5] Let's Discover More API Relations: A Large Language Model-Based AI Chain for Unsupervised API Relation Inference View paper
- [6] Evaluating large language models for user stance detection on X (Twitter) View paper
- [7] No More Labelled Examples? An Unsupervised Log Parser with LLMs View paper
- [8] RLZero: Direct Policy Inference from Language Without In-Domain Supervision View paper
- [9] Leveraging LLMs for Unsupervised Dense Retriever Ranking View paper
- [10] Discovering bias in latent space: An unsupervised debiasing approach View paper
- [11] CAM: A Large Language Model-based Creative Analogy Mining Framework View paper
- [12] LUNAR: Unsupervised LLM-based log parsing View paper
- [13] A logical word embedding for learning grammar View paper
- [14] Challenges with unsupervised LLM knowledge discovery View paper
- [15] Reasoning under Uncertainty: Efficient LLM Inference via Unsupervised Confidence Dilution and Convergent Adaptive Sampling View paper
- [16] Fantastic Reasoning Behaviors and Where to Find Them: Unsupervised Discovery of the Reasoning Process View paper
- [17] Effective Self-Mining of In-Context Examples for Unsupervised Machine Translation with LLMs View paper
- [18] Quantum Visual Word Sense Disambiguation: Unraveling Ambiguities Through Quantum Inference Model View paper
- [19] HiCoTraj: Zero-Shot Demographic Reasoning via Hierarchical Chain-of-Thought Prompting from Trajectory View paper
- [20] Benchmarking Extraction of Structured Data from Templatized Documents View paper
- [21] Sample trajectory selection method based on large language model in reinforcement learning View paper
- [22] Not All Rollouts are Useful: Down-Sampling Rollouts in LLM Reinforcement Learning View paper
- [23] Prompt optimization with EASE? efficient ordering-aware automated selection of exemplars View paper
- [24] Sample efficient preference alignment in LLMs via active exploration View paper
- [25] Reinforcement learning for reasoning in large language models with one training example View paper
- [26] GENEREIT: generating multi-talented reinforcement learning agents View paper
- [27] Inference-aware fine-tuning for best-of-n sampling in large language models View paper
- [28] Improving Data Efficiency for LLM Reinforcement Fine-tuning Through Difficulty-targeted Online Data Selection and Rollout Replay View paper
- [29] Advances in Statistical Inference and Policy Optimization for Reinforcement Learning View paper

- [30] Tempera: Test-time prompting via reinforcement learning View paper
- [31] Online difficulty filtering for reasoning oriented reinforcement learning View paper
- [32] Generalized proximal policy optimization with sample reuse View paper
- [33] RL-Selector: Reinforcement Learning-Guided Data Selection via Redundancy Assessment View paper
- [34] Safety Filtering While Training: Improving the Performance and Sample Efficiency of Reinforcement Learning Agents View paper
- [35] Reinforcement Learning-Controlled Subspace Ensemble Sampling for Complex Data Structures View paper
- [36] Generalizing Off-Policy Learning under Sample Selection Bias View paper
- [37] Communication-Efficient Policy Gradient Methods for Distributed Reinforcement Learning View paper
- [38] Model-Based Reinforcement Learning for Cavity Filter Tuning View paper
- [39] Imbalanced Sample Selection With Deep Reinforcement Learning for Fault Diagnosis View paper
- [40] Adversarial Defense Mechanisms for Supervised Learning View paper
- [41] Financial Time Series: Adaptive Forecasting Frameworks View paper
- [42] Applying Conformal Prediction for LLM Multi-Label Text Classification View paper