Novelty Assessment Report

Paper: Reinforcement Learning for Machine Learning Engineering Agents
PDF URL: https://openreview.net/pdf?id=mfIbSouoaZ
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-04

Abstract

Machine learning engineering (MLE) has a clear objective: Given an MLE task and a verifier (e.g., performance on some held-out data), what is the most effective way to utilize compute to achieve the best performance for the given task? Existing language model (LM) agents rely on prompting frontier LMs and accumulating experience non-parametrically by storing and retrieving experience through agent scaffolds and test-time compute. In this paper, we show that in environments such as MLE where a good verifier is available, adapting the LM parameters through gradient updates can be more effective in utilizing compute and agent's experience. Specifically, we show that agents backed by weaker models that improve via reinforcement learning (RL) can eventually outperform agents backed by much larger, but static models for a given MLE task. We identify two major challenges with RL in this setting. First, actions can take a variable amount of time (e.g., executing code for different solutions), which leads to asynchronous policy gradient updates that favor faster but suboptimal solutions. We propose duration-aware gradient updates in a distributed asynchronous RL framework to amplify high-cost but high-reward actions. Second, using performance on the held-out data as a reward for MLE provides limited feedback. A program that's nearly correct is treated the same as one that fails entirely (e.g., during data loading). We propose environment instrumentation to offer verifiable partial credit, using a separate, static language model to insert print statement to an existing program. Our experiments suggest that a small LM (Qwen2.5-3B) adapted with RL, when given enough compute, can solve an MLE task better than prompting a frontier model (Claude-3.5-Sonnet) with the state-of-the-art agent scaffold (AIDE) by an average of 22% across 12 Kaggle tasks.

Core Task Landscape

This paper addresses: **Adapting language models for machine learning engineering tasks using reinforcement learning**
A total of **50 papers** were analyzed and organized into a taxonomy with **19 categories**.

Taxonomy Overview

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

- **RL-Based LLM Alignment and Reasoning Enhancement**
- **LLM-Guided RL for Interactive Decision-Making**
- **Domain-Specific LLM Adaptation with RL**
- **RL Algorithms and Training Innovations for LLMs**
- **Surveys and Taxonomies of RL-LLM Integration**

Complete Taxonomy Tree

- Adapting language models for machine learning engineering tasks using reinforcement learning Survey Taxonomy
- RL-Based LLM Alignment and Reasoning Enhancement
  - General Alignment via Human Feedback (4 papers)
  - [4] Fine-tuning language models with reward learning on policy (Huang Fei, 2024) View paper
  - [8] Okapi: Instruction-tuned large language models in multiple languages with reinforcement learning from human feedback (Viet Lai, 2023) View paper
  - [10] Training language models to follow instructions with human feedback (Ouyang Long, 2022) View paper
  - [30] Fine-Tuning Language Models from Human Preferences (Ziegler, 2019) View paper
  - Reasoning and Complex Task Optimization (4 papers)
  - [1] Teaching large language models to reason with reinforcement learning (Havrilla, 2024) View paper
  - [18] Advancing language model reasoning through reinforcement learning and inference scaling (Hou Zhenyu, 2025) View paper
  - [23] Improving large language models via fine-grained reinforcement learning with minimum editing constraint (Chen Zhipeng, 2024) View paper
  - [48] GPG: A Simple and Strong Reinforcement Learning Baseline for Model Reasoning (Chu, 2025) View paper
  - Fine-Tuning Methodologies and Training Dynamics (5 papers)
  - [20] Rl fine-tuning heals ood forgetting in sft (Hangzhan Jin, 2025) View paper
  - [28] Supervised fine tuning on curated data is reinforcement learning (and can be improved) (Qin, 2025) View paper
  - [29] Inverse reinforcement learning meets large language model post-training: Basics, advances, and opportunities (Sun Hao, 2025) View paper
  - [33] All roads lead to likelihood: The value of reinforcement learning in fine-tuning (Swamy, 2025) View paper
  - [46] Rl is neither a panacea nor a mirage: Understanding supervised vs. reinforcement learning fine-tuning for llms (Wu, 2025) View paper
- LLM-Guided RL for Interactive Decision-Making
  - Pre-Trained LLMs as Policy Initializers (4 papers)
  - [3] Pre-trained language models for interactive decision-making (Li Shuang, 2022) View paper

- [11] Real-time integration of fine-tuned large language model for improved decision-making in reinforcement learning (Xiancai Xiang, 2024) View paper
- [14] Unleashing the power of pre-trained language models for offline reinforcement learning (Shi, 2023) View paper
- [21] Rladapter: Bridging large language models to reinforcement learning in open worlds (Zhang Wan-peng, 2023) View paper
- LLM-Based Reward Shaping and Exploration (4 papers)
- [7] LAPP: Large Language Model Feedback for Preference-Driven Reinforcement Learning (Wei Xiao, 2025) View paper
- [9] Self-refined large language model as automated reward function designer for deep reinforcement learning in robotics (Song JiaYang, 2023) View paper
- [15] Guiding pretraining in reinforcement learning with large language models (Du Yuqing, 2023) View paper
- [22] Natural language reinforcement learning (Feng, 2024) View paper
- Robotics and Embodied Control (3 papers)
- [5] Fine-tuning large vision-language models as decision-making agents via reinforcement learning (Hao Bai, 2024) View paper
- [37] Improving Vision-Language-Action Model with Online Reinforcement Learning (Yanjiang Guo, 2025) View paper
- [38] Plan-seq-learn: Language model guided rl for solving long horizon robotics tasks (Dalal, 2024) View paper
- Multi-Agent and Collaborative Systems (3 papers)
- [13] Learning to deliberate: Meta-policy collaboration for agentic llms with multi-agent reinforcement learning (Yang Wei, 2025) View paper
- [24] Pilotrl: Training language model agents via global planning-guided progressive reinforcement learning (Chen Chong, 2025) View paper
- [39] Maporl: Multi-agent post-co-training for collaborative large language models with reinforcement learning (Guo, 2025) View paper
- Game-Theoretic and Theoretical Frameworks (1 papers)
- [40] Large Language Models as Agents in Two-Player Games (Liu Yang, 2024) View paper

- Domain-Specific LLM Adaptation with RL
- Software Engineering and Code Generation (3 papers)
- [2] SEGym: optimizing large language model assisted software engineering agents with reinforcement learning (Gerhard Stenzel, 2024) View paper
- [26] Leveraging reinforcement learning and large language models for code optimization (Shukai Duan, 2023) View paper
- [35] SWE-RL: Advancing LLM Reasoning via Reinforcement Learning on Open Software Evolution (Wei, 2025) View paper
- Machine Learning Engineering Automation ★ (2 papers)
- [0] Reinforcement Learning for Machine Learning Engineering Agents (Anon et al., 2026) View paper
- [42] ML-Agent: Reinforcing LLM Agents for Autonomous Machine Learning Engineering (Liu Ze-xi, 2025) View paper
- Healthcare and Clinical Applications (1 papers)
- [12] Adapting Open-Source Large Language Models for Cost-Effective, Expert-Level Clinical Note Generation with On-Policy Reinforcement Learning (Wang Han-yin, 2024) View paper
- Specialized Task Domains (8 papers)
- [6] Integrating large language models, reinforcement learning, and machine learning for intelligent indoor thermal comfort regulation (Deli Liu, 2025) View paper
- [16] Exploiting large language model with reinforcement learning for generative job recommendations (Zhi Zheng, 2026) View paper
- [17] Rl2: Reinforce large language model to assist safe reinforcement learning for energy management of active distribution networks (Xu Yang, 2025) View paper
- [25] Fine-tuning a large language model with reinforcement learning for educational question generation (Salima Lamsiyah, 2024) View paper
- [27] Industrial internet of things with large language models (llms): an intelligence-based reinforcement learning approach (Yuzheng Ren, 2024) View paper
- [36] Decision-Making Large Language Model for Wireless Communication: A Comprehensive Survey on Key Techniques (Ning Yang, 2025) View paper
- [44] Large language model-enhanced reinforcement learning for low-altitude economy networking (Cai LingYi, 2025) View paper
- [47] RLMolLM: Reinforcement Learning-Enhanced Language Model Framework for Inverse Molecular Design (Xiaobo Lin, 2025) View paper
- Domain-General Fine-Tuning Frameworks (1 papers)
- [19] A fine-tuned large language model for domain-specific with reinforcement learning (Amelia Ritahani Ismail, 2024) View paper

- RL Algorithms and Training Innovations for LLMs
- Policy Optimization and Gradient Methods (1 papers)
- [41] Remax: A simple, effective, and efficient reinforcement learning method for aligning large language models (Li, 2023) View paper
- Value-Based and Q-Learning Approaches (1 papers)
- [32] Q-sft: Q-learning for language models via supervised fine-tuning (Hong, 2024) View paper
- Risk-Aware and Safety-Oriented Training (1 papers)
- [45] Risk-averse fine-tuning of large language models (Chaudhary, 2024) View paper
- Self-Adaptive and Meta-Learning Systems (2 papers)
- [34] Self-Adapting Language Models (Pari, 2025) View paper
- [43] Adaptive Learning Machines: A Framework for Dynamic and Real-Time ML Applications (Shethiya, 2024) View paper
- Multimodal and Vision-Language RL (1 papers)
- [50] Reason-rft: Reinforcement fine-tuning for visual reasoning of vision language models (Tan Huajie, 2025) View paper

- Surveys and Taxonomies of RL-LLM Integration (2 papers)
- [31] Reinforcement learning in large language models (llms): The rise of ai language giants (Baihan Lin, 2024) View paper
- [49] The rl/llm taxonomy tree: Reviewing synergies between reinforcement learning and large language models (Singh, 2024) View paper

Narrative

Core task: Adapting language models for machine learning engineering tasks using reinforcement learning. The field has crystallized around several complementary directions. RL-Based LLM Alignment and Reasoning Enhancement focuses on improving model outputs through human feedback and reasoning capabilities, exemplified by foundational work like InstructGPT[10] and newer reasoning

methods such as Teaching LLMs Reasoning[1]. LLM-Guided RL for Interactive Decision-Making explores how pretrained language models can inform sequential decision problems, with approaches like Pretrained LMs Decisions[3] bridging natural language understanding and policy learning. Domain-Specific LLM Adaptation with RL targets specialized applications—ranging from software engineering environments like SEGym[2] and SWE-RL[35] to molecular design in RLMolLM[47] and clinical documentation in Clinical Note Generation[12]—where RL fine-tunes models for narrow, high-stakes tasks. Meanwhile, RL Algorithms and Training Innovations for LLMs advances the underlying optimization machinery, investigating techniques like Supervised Fine-tuning RL[28] and risk-aware methods such as Risk-averse Fine-tuning[45]. Finally, Surveys and Taxonomies of RL-LLM Integration, including RL LLM Taxonomy[49], provide structured overviews of this rapidly evolving landscape.

A particularly active line of work centers on automating complex engineering workflows. ML Engineering Agents[0] sits squarely within the Domain-Specific LLM Adaptation branch, specifically targeting machine learning engineering automation—a niche that also includes ML-Agent[42], which similarly applies RL to streamline ML development pipelines. Compared to broader software engineering agents like SWE-RL[35], ML Engineering Agents[0] narrows its scope to the iterative, experiment-heavy nature of model training and hyperparameter tuning. This contrasts with more general-purpose interactive agents such as Pretrained LMs Decisions[3], which emphasize flexible decision-making across diverse environments. The central tension across these branches involves balancing domain specialization—where tight coupling to task structure yields strong performance—against the generality and sample efficiency that pretrained models promise. ML Engineering Agents[0] exemplifies this trade-off by leveraging RL to adapt language models specifically for the feedback-rich, code-and-data-centric loops characteristic of ML engineering.

Related Works in Same Category

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

1. ML-Agent: Reinforcing LLM Agents for Autonomous Machine Learning Engineering

Authors: Liu Ze-xi, Chai, Jingyi, Zhu Xinyu, Tang Shuo, et al. (10 authors total) | **Year/Venue**: 2025 • arXiv.org | **URL**: View paper

Abstract

The emergence of large language model (LLM)-based agents has significantly advanced the development of autonomous machine learning (ML) engineering. However, most existing approaches rely heavily on manual prompt engineering, failing to adapt and optimize based on diverse experimental experiences. Focusing on this, for the first time, we explore the paradigm of learning-based agentic ML, where an LLM agent learns through interactive experimentation on ML tasks using online reinforcement learning...

Relationship Analysis

Both papers belong to the Machine Learning Engineering Automation category, focusing on using RL to train LLM agents for automating ML pipeline tasks. They share substantial overlap in their core approach of applying reinforcement learning to adapt small language models (Qwen2.5-3B vs Qwen-2.5-7B) for autonomous ML engineering tasks, both addressing challenges like sparse rewards and action execution variability. The key differences are that the original paper emphasizes duration-aware gradient updates for variable-time actions and environment instrumentation for partial credit, while the candidate paper focuses on exploration-enriched fine-tuning, step-wise RL training, and a unified reward module for diverse ML feedback signals.

Contributions Analysis

Overall novelty summary. The paper proposes adapting language models via reinforcement learning for machine learning engineering tasks, introducing duration-aware gradient updates and environment instrumentation for partial credit. It resides in the 'Machine Learning Engineering Automation' leaf under 'Domain-Specific LLM Adaptation with RL', which contains only two papers total (including this one). This represents a relatively sparse research direction within the broader taxonomy of 50 papers, suggesting the specific focus on RL-based automation of ML engineering workflows is still emerging compared to more crowded areas like general alignment or software engineering applications.

The taxonomy reveals neighboring leaves in software engineering (SEGym, SWE-RL) and other specialized domains (healthcare, molecular design), but these target different task structures. The closest conceptual relatives appear in 'RL-Based LLM Alignment and Reasoning Enhancement', particularly reasoning optimization methods, and in 'LLM-Guided RL for Interactive Decision-Making', which explores policy learning with pretrained models. However, the scope notes clarify that this work's emphasis on automating iterative ML development pipelines—with verifiable feedback loops and code execution—distinguishes it from both general-purpose interactive agents and broader software engineering automation.

Among 23 candidates examined across three contributions, none were flagged as clearly refuting the work. The duration-aware gradient updates contribution examined 10 candidates with zero refutable matches; environment instrumentation examined 4 with none refutable; the demonstration that RL-adapted small models outperform static frontier models examined 9 with none refutable. This suggests that within the limited search scope—focused on top semantic matches and citation expansion—the specific combination of asynchronous RL training dynamics, partial credit mechanisms, and empirical comparisons on ML engineering tasks appears relatively unexplored in prior literature.

The analysis reflects a targeted literature search rather than exhaustive coverage, and the sparse taxonomy leaf (two papers) indicates this research direction is nascent. While no direct overlaps emerged among examined candidates, the limited scope means potentially relevant work in adjacent areas—such as asynchronous RL methods outside the LLM context or ML automation without RL—may not have been captured. The novelty assessment is thus conditional on the search boundaries and the specific framing of ML engineering as an RL problem.

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

Contribution 1: Duration-aware gradient updates for distributed asynchronous RL

Description: The authors introduce a method to reweight policy gradient updates by action execution duration in distributed RL settings. This addresses the problem where asynchronous training favors faster actions, ensuring that slower but potentially higher-reward actions receive fair consideration during parameter updates.

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. Stellaris: Staleness-Aware Distributed Reinforcement Learning with Serverless Computing

URL: View paper

Brief Assessment

Stellaris[60] focuses on staleness-aware gradient aggregation in serverless computing environments, not on reweighting gradients by action execution duration in agentic settings where actions have variable execution times.

---

2. Asynchronous stochastic gradient descent for extreme-scale recommender systems

URL: View paper

Brief Assessment

Extreme-scale Recommender Systems[58] addresses asynchronous SGD for recommender systems with staleness normalization based on execution time, but does not focus on reinforcement learning or action execution duration in agentic settings. The candidate's staleness normalization is designed for parameter server architectures in supervised learning, not for RL policy gradient updates where actions have variable execution times.

---

3. Asynchronous Federated Reinforcement Learning with Policy Gradient Updates: Algorithm Design and Convergence Analysis

URL: View paper

Brief Assessment

Asynchronous Federated RL[61] addresses delay-adaptive updates in federated RL settings where multiple agents collaborate, not the single-agent distributed RL framework with variable action execution times described in the original paper.

---

4. Staleness-aware async-sgd for distributed deep learning

URL: View paper

Brief Assessment

Staleness-aware Async-SGD[55] addresses gradient staleness in distributed deep learning by modulating learning rates based on how outdated gradients are (time since parameter update). The original paper addresses action execution duration in RL environments where different actions take different amounts of wall-clock time to execute. These are fundamentally different problems: one concerns parameter staleness in distributed optimization, the other concerns variable-duration environment interactions in RL.

---

5. Addressing stale gradients in scalable federated deep reinforcement learning

URL: View paper

Brief Assessment

Stale Gradients Federated[56] addresses stale gradients in federated RL across heterogeneous devices in Atari games, not duration-aware reweighting for variable-time action execution in ML engineering tasks.

---

6. TransAL-CC: An Asynchronous Reinforcement Learning Approach for Multipath Transmission Congestion Control in Power IoT

URL: View paper

Brief Assessment

TransAL-CC[59] focuses on multipath transmission congestion control in Power IoT networks. The available text fragments mention 'stable gradient flow throughout asynchronous actor' but do not provide sufficient detail about duration-aware gradient reweighting mechanisms for variable-time action execution in distributed RL settings.

---

7. Addressing stale gradients in asynchronous federated deep reinforcement learning

URL: View paper

Brief Assessment

Stale Gradients Asynchronous[62] addresses stale gradients in federated RL where workers have variable offline intervals, not variable action execution durations. The original paper's duration-aware updates reweight by action execution time (e.g., training different ML models), while the candidate focuses on handling delayed gradient updates from asynchronous workers in federated settings.

---

8. FedStaleWeight: Buffered Asynchronous Federated Learning with Fair Aggregation via Staleness Reweighting

URL: View paper

Brief Assessment

FedStaleWeight[63] addresses staleness reweighting in federated learning for distributed machine learning training, not reinforcement learning. The original paper's contribution focuses specifically on RL agents with variable action execution times in MLE tasks, while the candidate addresses federated learning aggregation across heterogeneous compute devices.

---

9. Communication-Constrained Distributed Learning: TSI-Aided Asynchronous Optimization with Stale Gradient

URL: View paper

Brief Assessment

TSI-Aided Asynchronous[64] addresses gradient staleness in distributed learning through timing side information and worker selection, but focuses on communication-constrained federated learning settings rather than reinforcement learning with variable-duration action execution in agentic environments.

---

10. Accelerating distributed reinforcement learning with in-switch computing

URL: View paper

Brief Assessment

In-switch Computing[57] addresses staleness of gradients in distributed RL due to asynchrony, but does not specifically address duration-aware reweighting based on action execution time. The original paper's contribution focuses on reweighting gradients by action duration (Δt) to prevent bias toward faster actions in variable-time execution environments.

Contribution 2: Environment instrumentation for verifiable partial credit

Description: The authors propose using a static copy of the language model to instrument agent-generated code by inserting print statements. This provides intermediate feedback and partial credit for completing high-level procedures (e.g., loading data, training models), mitigating the sparse reward problem in MLE tasks.

This contribution was assessed against **4 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. A model for interaction of agents and environments

URL: View paper

Brief Assessment

Agent Environment Interaction[53] focuses on abstract models of agent-environment interaction using insertion functions and transition systems. It does not address code instrumentation, print statement insertion, or partial credit mechanisms for machine learning tasks.

---

2. Improved Methods based on Too Many Cooks

URL: View paper

Brief Assessment

Too Many Cooks[54] focuses on multi-agent coordination in cooking tasks using Bayesian delegation and DQN, not on code instrumentation or partial credit mechanisms for machine learning engineering tasks.

---

3. Design and implementation of a fully transparent partial abort support for software transactional memory

URL: View paper

Brief Assessment

Transparent Partial Abort[52] focuses on software transactional memory systems for concurrent programming, using static binary instrumentation to undo transaction operations. The original paper's environment instrumentation inserts print statements into agent-generated code for ML tasks to provide partial credit feedback, which is a fundamentally different domain and purpose.

---

4. CREW-WILDFIRE: Benchmarking Agentic Multi-Agent Collaborations at Scale

URL: View paper

Brief Assessment

CREW-WILDFIRE[51] focuses on multi-agent wildfire response coordination benchmarks with spatial reasoning and team communication, not on code instrumentation or partial credit mechanisms for machine learning engineering tasks.

Contribution 3: Demonstration that RL-adapted small models outperform prompting frontier models

Description: The authors demonstrate empirically that a small language model (Qwen2.5-3B) adapted through RL can solve MLE tasks better than prompting a frontier model (Claude-3.5-Sonnet) with state-of-the-art agent scaffolds, achieving an average 22% improvement across 12 Kaggle tasks.

This contribution was assessed against **9 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Hawkeye: Model Collaboration for Efficient Reasoning

URL: View paper

Brief Assessment

Hawkeye[72] focuses on reducing redundancy in chain-of-thought reasoning through model collaboration (large model generates concise instructions, small model expands responses), not on demonstrating that RL-adapted small models outperform prompting frontier models on MLE tasks. The original paper's contribution is about RL training for machine learning engineering agents, while Hawkeye[72] addresses reasoning efficiency through a different architectural approach.

---

2. Seed-Prover 1.5: Mastering Undergraduate-Level Theorem Proving via Learning from Experience

URL: View paper

Brief Assessment

Seed-Prover[73] focuses on formal theorem proving in Lean, not machine learning engineering tasks. The domains, evaluation metrics, and technical approaches are fundamentally different.

---

3. Kevin: Multi-Turn RL for Generating CUDA Kernels

URL: View paper

Brief Assessment

Kevin[70] focuses on CUDA kernel generation tasks with multi-turn RL, not general machine learning engineering tasks across diverse domains (vision, language, tabular data). The technical domains and task types are fundamentally different.

---

4. R2Vul: Learning to Reason about Software Vulnerabilities with Reinforcement Learning and Structured Reasoning Distillation

URL: View paper

Brief Assessment

R2Vul[67] focuses on software vulnerability detection using RLAIF with structured reasoning distillation, not on machine learning engineering tasks. The candidate demonstrates a 1.5B model outperforming larger models in vulnerability detection, but this is a different domain and methodology from the original paper's MLE agent framework with duration-aware gradients and environment instrumentation.

---

5. LIMR: Less is More for RL Scaling

URL: View paper

Brief Assessment

LIMR[68] focuses on data efficiency in RL training for reasoning tasks (math problems), demonstrating that a small subset of training samples can match full dataset performance. The original paper addresses a different problem: showing that RL-adapted small models can outperform prompting large models on machine learning engineering tasks, not reasoning tasks.

---

6. Simulating fish autonomous swimming behaviours using deep reinforcement learning based on Kolmogorovâ□□Arnold Networks

URL: View paper

Brief Assessment

Fish Swimming Behaviors[65] focuses on comparing different neural network architectures (KANs vs LSTMs/MLPs) for fish swimming simulation in fluid dynamics environments, not on comparing small RL-adapted models against large language models for machine learning engineering tasks.

---

7. VerIPO: Cultivating Long Reasoning in Video-LLMs via Verifier-Gudied Iterative Policy Optimization

URL: View paper

Brief Assessment

VerIPO[74] focuses on applying RL to video-LLMs for video reasoning tasks, not on machine learning engineering (MLE) tasks. The candidate does not address the specific claim about small models solving MLE tasks better than frontier models with agent scaffolds.

---

8. Ring-lite: Scalable Reasoning via C3PO-Stabilized Reinforcement Learning for LLMs

URL: View paper

Brief Assessment

Ring-lite[71] focuses on MoE-based model optimization via RL for reasoning tasks (AIME, GPQA-Diamond), not on machine learning engineering tasks. The candidate does not address MLE agent scaffolds or Kaggle competitions.

---

9. Ghostnetv3: Exploring the training strategies for compact models

URL: View paper

Brief Assessment

GhostNetV3[69] focuses on training strategies for compact CNN models on image classification tasks, not on reinforcement learning adaptation of language models for machine learning engineering tasks.

Appendix: Text Similarity Detection

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

References

- [0] Reinforcement Learning for Machine Learning Engineering Agents View paper
- [1] Teaching large language models to reason with reinforcement learning View paper
- [2] SEGym: optimizing large language model assisted software engineering agents with reinforcement learning View paper
- [3] Pre-trained language models for interactive decision-making View paper
- [4] Fine-tuning language models with reward learning on policy View paper
- [5] Fine-tuning large vision-language models as decision-making agents via reinforcement learning View paper
- [6] Integrating large language models, reinforcement learning, and machine learning for intelligent indoor thermal comfort regulation View paper
- [7] LAPP: Large Language Model Feedback for Preference-Driven Reinforcement Learning View paper
- [8] Okapi: Instruction-tuned large language models in multiple languages with reinforcement learning from human feedback View paper
- [9] Self-refined large language model as automated reward function designer for deep reinforcement learning in robotics View paper
- [10] Training language models to follow instructions with human feedback View paper
- [11] Real-time integration of fine-tuned large language model for improved decision-making in reinforcement learning View paper
- [12] Adapting Open-Source Large Language Models for Cost-Effective, Expert-Level Clinical Note Generation with On-Policy Reinforcement Learning View paper
- [13] Learning to deliberate: Meta-policy collaboration for agentic llms with multi-agent reinforcement learning View paper
- [14] Unleashing the power of pre-trained language models for offline reinforcement learning View paper
- [15] Guiding pretraining in reinforcement learning with large language models View paper
- [16] Exploiting large language model with reinforcement learning for generative job recommendations View paper
- [17] Rl2: Reinforce large language model to assist safe reinforcement learning for energy management of active distribution networks View paper
- [18] Advancing language model reasoning through reinforcement learning and inference scaling View paper
- [19] A fine-tuned large language model for domain-specific with reinforcement learning View paper
- [20] Rl fine-tuning heals ood forgetting in sft View paper
- [21] Rladapter: Bridging large language models to reinforcement learning in open worlds View paper
- [22] Natural language reinforcement learning View paper
- [23] Improving large language models via fine-grained reinforcement learning with minimum editing constraint View paper
- [24] Pilotrl: Training language model agents via global planning-guided progressive reinforcement learning View paper
- [25] Fine-tuning a large language model with reinforcement learning for educational question generation View paper
- [26] Leveraging reinforcement learning and large language models for code optimization View paper
- [27] Industrial internet of things with large language models (llms): an intelligence-based reinforcement learning approach View paper
- [28] Supervised fine tuning on curated data is reinforcement learning (and can be improved) View paper
- [29] Inverse reinforcement learning meets large language model post-training: Basics, advances, and opportunities View paper
- [30] Fine-Tuning Language Models from Human Preferences View paper
- [31] Reinforcement learning in large language models (llms): The rise of ai language giants View paper
- [32] Q-sft: Q-learning for language models via supervised fine-tuning View paper
- [33] All roads lead to likelihood: The value of reinforcement learning in fine-tuning View paper
- [34] Self-Adapting Language Models View paper
- [35] SWE-RL: Advancing LLM Reasoning via Reinforcement Learning on Open Software Evolution View paper
- [36] Decision-Making Large Language Model for Wireless Communication: A Comprehensive Survey on Key Techniques View paper
- [37] Improving Vision-Language-Action Model with Online Reinforcement Learning View paper
- [38] Plan-seq-learn: Language model guided rl for solving long horizon robotics tasks View paper
- [39] Maporl: Multi-agent post-co-training for collaborative large language models with reinforcement learning View paper
- [40] Large Language Models as Agents in Two-Player Games View paper
- [41] Remax: A simple, effective, and efficient reinforcement learning method for aligning large language models View paper
- [42] ML-Agent: Reinforcing LLM Agents for Autonomous Machine Learning Engineering View paper
- [43] Adaptive Learning Machines: A Framework for Dynamic and Real-Time ML Applications View paper
- [44] Large language model-enhanced reinforcement learning for low-altitude economy networking View paper
- [45] Risk-averse fine-tuning of large language models View paper
- [46] Rl is neither a panacea nor a mirage: Understanding supervised vs. reinforcement learning fine-tuning for llms View paper
- [47] RLMolLM: Reinforcement Learning-Enhanced Language Model Framework for Inverse Molecular Design View paper
- [48] GPG: A Simple and Strong Reinforcement Learning Baseline for Model Reasoning View paper
- [49] The rl/llm taxonomy tree: Reviewing synergies between reinforcement learning and large language models View paper
- [50] Reason-rft: Reinforcement fine-tuning for visual reasoning of vision language models View paper
- [51] CREW-WILDFIRE: Benchmarking Agentic Multi-Agent Collaborations at Scale View paper
- [52] Design and implementation of a fully transparent partial abort support for software transactional memory View paper
- [53] A model for interaction of agents and environments View paper
- [54] Improved Methods based on Too Many Cooks View paper
- [55] Staleness-aware async-sgd for distributed deep learning View paper

- [56] Addressing stale gradients in scalable federated deep reinforcement learning View paper
- [57] Accelerating distributed reinforcement learning with in-switch computing View paper
- [58] Asynchronous stochastic gradient descent for extreme-scale recommender systems View paper
- [59] TransAL-CC: An Asynchronous Reinforcement Learning Approach for Multipath Transmission Congestion Control in Power IoT View paper
- [60] Stellaris: Staleness-Aware Distributed Reinforcement Learning with Serverless Computing View paper
- [61] Asynchronous Federated Reinforcement Learning with Policy Gradient Updates: Algorithm Design and Convergence Analysis View paper
- [62] Addressing stale gradients in asynchronous federated deep reinforcement learning View paper
- [63] FedStaleWeight: Buffered Asynchronous Federated Learning with Fair Aggregation via Staleness Reweighting View paper
- [64] Communication-Constrained Distributed Learning: TSI-Aided Asynchronous Optimization with Stale Gradient View paper
- [65] Simulating fish autonomous swimming behaviours using deep reinforcement learning based on Kolmogorovâ□□Arnold Networks View paper
- [66] Revolutionizing reinforcement learning framework for diffusion large language models View paper
- [67] R2Vul: Learning to Reason about Software Vulnerabilities with Reinforcement Learning and Structured Reasoning Distillation View paper
- [68] LIMR: Less is More for RL Scaling View paper
- [69] Ghostnetv3: Exploring the training strategies for compact models View paper
- [70] Kevin: Multi-Turn RL for Generating CUDA Kernels View paper
- [71] Ring-lite: Scalable Reasoning via C3PO-Stabilized Reinforcement Learning for LLMs View paper
- [72] Hawkeye: Model Collaboration for Efficient Reasoning View paper
- [73] Seed-Prover 1.5: Mastering Undergraduate-Level Theorem Proving via Learning from Experience View paper
- [74] VerIPO: Cultivating Long Reasoning in Video-LLMs via Verifier-Gudied Iterative Policy Optimization View paper