Novelty Assessment Report

Paper: SHAPO: Sharpness-Aware Policy Optimization for Safe Exploration

PDF URL: https://openreview.net/pdf?id=7cUxi8LbKD

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-30

Abstract

Safe exploration is a prerequisite for deploying reinforcement learning (RL) agents in safety-critical domains. In this paper, we approach safe exploration through the lens of epistemic uncertainty, where the actor's sensitivity to parameter perturbations serves as a practical proxy for regions of high uncertainty. We propose Sharpness-Aware Policy Optimization (SHAPO), a sharpness-aware policy update rule that evaluates gradients at perturbed parameters, making policy updates pessimistic with respect to the actor's epistemic uncertainty. Analytically we show that this adjustment implicitly reweighs policy gradients, amplifying the influence of rare unsafe actions while tempering contributions from already safe ones, thereby biasing learning toward conservative behavior in under-explored regions. Across several continuous-control tasks, our method consistently improves both safety and task performance over existing baselines, significantly expanding their Pareto frontiers.

Core Task Landscape

This paper addresses: **Safe Exploration in Reinforcement Learning Under Constraints**

A total of **50 papers** were analyzed and organized into a taxonomy with **23 categories**.

Taxonomy Overview

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

- **Constraint Formulation and Problem Frameworks**
- **Algorithm Design for Safe Exploration**
- **Uncertainty Quantification and Epistemic Safety**
- **Constraint Manifold and Geometric Approaches**
- **Statewise and Instantaneous Safety**
- **Multi-Agent and Hierarchical Safe RL**
- **Interactive and Human-in-the-Loop Safe Learning**
- **Application Domains**
- **Exploration Strategy Design**

Complete Taxonomy Tree

- Safe Exploration in Reinforcement Learning Under Constraints Survey Taxonomy
- Constraint Formulation and Problem Frameworks
  - Generalized Safe Exploration Formulations (1 papers)
  - [1] Safe exploration in reinforcement learning: A generalized formulation and algorithms (Wachi Akifumi, 2023) View paper
  - Constrained MDP Foundations (2 papers)
  - [19] Learning with safety constraints: Sample complexity of reinforcement learning for constrained mdps (Aria HasanzadeZonuzy, 2021) View paper
  - [26] A Survey of Safe Reinforcement Learning and Constrained MDPs: A Technical Survey on Single-Agent and Multi-Agent Safety (Kushwaha Ankita, 2025) View paper
  - Constraint Representation Taxonomies (2 papers)
  - [15] Safe exploration techniques for reinforcement learningâ□an overview (Martin PeÄ□ka, 2014) View paper
  - [29] A survey of constraint formulations in safe reinforcement learning (Wachi Akifumi, 2024) View paper
  - Natural Language and Non-Standard Constraints (1 papers)
  - [6] Safe reinforcement learning with natural language constraints (Yang, 2021) View paper
- Algorithm Design for Safe Exploration
  - Safety Layer and Action Correction Methods (3 papers)
  - [4] Safe reinforcement learning via hierarchical adaptive chance-constraint safeguards (Zhaorun Chen, 2024) View paper
  - [10] Safe exploration in continuous action spaces (Dalal, 2018) View paper
  - [25] Safe Reinforcement Learning for Constrained Optimal Control With Provable Guarantees: Applications to Motion Planning (Fei Zhang, 2025) View paper
  - Policy Optimization with Safety Constraints ★ (4 papers)
  - [0] SHAPO: Sharpness-Aware Policy Optimization for Safe Exploration (Anon et al., 2026) View paper
  - [11] Model-based safe deep reinforcement learning via a constrained proximal policy optimization algorithm (Jayant, 2022) View paper
  - [14] Safety optimized reinforcement learning via multi-objective policy optimization (Homayoun Honari, 2024) View paper
  - [30] Feasible actor-critic: Constrained reinforcement learning for ensuring statewise safety (Ma, 2021) View paper
  - Model-Based Safe RL (2 papers)
  - [20] Actsafe: Active exploration with safety constraints for reinforcement learning (As, 2024) View paper

- [41] Safe exploration and optimization of constrained mdps using gaussian processes (Akifumi Wachi, 2018) View paper
- Recovery and Backup Policy Approaches (2 papers)
- [2] Safe exploration for reinforcement learning. (Alexander, 2008) View paper
- [3] Recovery rl: Safe reinforcement learning with learned recovery zones (Brijen Thananjeyan, 2021) View paper
- Conservative and Distributional Safety Critics (3 papers)
- [8] Conservative safety critics for exploration (Bharadhwaj, 2020) View paper
- [12] Safety-constrained reinforcement learning with a distributional safety critic (Qisong Yang, 2023) View paper
- [50] Off-Policy Conservative Distributional Reinforcement Learning With Safety Constraints (Heng-Rui Zhang, 2024) View paper
- Safe Exploration with Prior Knowledge (5 papers)
- [9] Reinforcement learning by guided safe exploration (Qisong Yang, 2023) View paper
- [13] Safety-constrained reinforcement learning for MDPs (Sebastian Junges, 2016) View paper
- [33] Near-optimal Conservative Exploration in Reinforcement Learning under Episode-wise Constraints (Li Donghao, 2023) View paper
- [34] Enhance Exploration in Safe Reinforcement Learning with Contrastive Representation Learning (Duc Kien Doan, 2025) View paper
- [36] Reinforcement learning with safe exploration for network security (Canhuang Dai, 2019) View paper
- Zero-Violation and Always-Safe Methods (3 papers)
- [22] Safe reinforcement learning in constrained markov decision processes (Akifumi Wachi, 2020) View paper
- [31] Alwayssafe: Reinforcement learning without safety constraint violations during training (T. D. SimÃ£o, 2021) View paper
- [38] Learn zero-constraint-violation safe policy in model-free constrained reinforcement learning (Haitong Ma, 2024) View paper
- Task-Agnostic and Reward-Free Safe Exploration (2 papers)
- [24] Cem: Constrained entropy maximization for task-agnostic safe exploration (Qisong Yang, 2023) View paper
- [45] Safe Reinforcement Learning via Episodic Control (Zhuo Li, 2025) View paper

- Uncertainty Quantification and Epistemic Safety
  - Epistemic Uncertainty for Safe Exploration (2 papers)
  - [40] Safe Exploration in Reinforcement Learning for Learning from Human Experts (Jorge RamÃrez, 2023) View paper
  - [46] Safe Reinforcement Learning in Autonomous Driving With Epistemic Uncertainty Estimation (Zheng Zhang, 2024) View paper
  - Probabilistic Constraints and Chance-Constrained RL (1 papers)
  - [43] Probabilistic constraint for safety-critical reinforcement learning (Wei-Qin Chen, 2024) View paper
- Constraint Manifold and Geometric Approaches
  - Control Barrier Functions and Lyapunov Methods (4 papers)
  - [7] Safe reinforcement learning on the constraint manifold: Theory and applications (Puze Liu, 2025) View paper
  - [16] Robot reinforcement learning on the constraint manifold (Puze Liu, 2022) View paper
  - [21] Adaptive Safety-Certified Reinforcement Learning for Constrained Optimal Control of Autonomous Robots With Uncertainties (Fei Zhang, 2025) View paper
  - [39] Reinforcement learning with safety and stability guarantees during exploration for linear systems (Zahra Marvi, 2022) View paper
  - Constraint Manifold Projection (1 papers)
  - [5] Safe exploration in reinforcement learning: Theory and applications in robotics (Berkenkamp, 2019) View paper
- Statewise and Instantaneous Safety
  - Hard Instantaneous Constraints (2 papers)
  - [17] Safe Exploration for Constrained Reinforcement Learning with Provable Guarantees (Archana Bura, 2021) View paper
  - [27] Safe reinforcement learning with instantaneous constraints: the role of aggressive exploration (Wei, 2024) View paper
- Multi-Agent and Hierarchical Safe RL (1 papers)
  - [23] Constrained meta-reinforcement learning for adaptable safety guarantee with differentiable convex programming (Cho MinJae, 2024) View paper
- Interactive and Human-in-the-Loop Safe Learning (1 papers)
  - [28] Safe exploration for interactive machine learning (Turchetta, 2019) View paper
- Application Domains
  - Autonomous Driving and Motion Planning (2 papers)
  - [18] Constraints driven safe reinforcement learning for autonomous driving decision-making (Fei Gao, 2024) View paper
  - [35] Enhancing Autonomous Lane-Changing Safety: Deep Reinforcement Learning via Pre-Exploration in Parallel Imaginary Environments (Zhiqun Hu, 2024) View paper
  - Energy and Power Systems (4 papers)
  - [37] Constrained Reinforcement Learning for Safe Heat Pump Control (Zhang Bao-he, 2024) View paper
  - [42] Rethinking Safe Policy Learning for Complex Constraints Satisfaction: A Glimpse in Real-Time Security Constrained Economic Dispatch Integrating Energy Storage Units (Jianxiong Hu, 2025) View paper
  - [47] Real-Time Sequential Security-Constrained Optimal Power Flow: A Hybrid Knowledge-Data-Driven Reinforcement Learning Approach (Zhongkai Yi, 2024) View paper
  - [49] Safe Reinforcement Learning for Energy Management of Electrified Vehicle With Novel Physics-Informed Exploration Strategy (Atriya Biswas, 2024) View paper
  - Disturbance and Uncertainty Handling (1 papers)
  - [48] Safe Exploration Method for Reinforcement Learning under Existence of Disturbance (Okawa, 2022) View paper
- Exploration Strategy Design (2 papers)
  - [32] Explicit Explore, Exploit, or Escape (): near-optimal safety-constrained reinforcement learning in polynomial time (DM Bossens, 2023) View paper
  - [44] Safe Reinforcement Learning with Constraints: A Survey (Zhengyu Chen, 2025) View paper

Narrative

Core task: safe exploration in reinforcement learning under constraints. The field addresses how agents can learn effective policies while respecting safety requirements throughout the learning process. The taxonomy reveals a rich structure organized around several complementary perspectives. Constraint Formulation and Problem Frameworks establish the mathematical foundations, defining how safety requirements are encoded—ranging from hard constraints in Safety Constrained MDPs[13] to probabilistic formulations like Hierarchical Chance Constraints[4]. Algorithm Design for Safe Exploration encompasses policy optimization methods that directly

integrate safety into learning, while Uncertainty Quantification and Epistemic Safety focuses on managing unknown risks through conservative estimation, as seen in Conservative Safety Critics[8] and Epistemic Uncertainty[46]. Geometric perspectives emerge in Constraint Manifold approaches like Constraint Manifold[7] and Robot Constraint Manifold[16], which exploit the structure of feasible state spaces. The taxonomy also distinguishes between Statewise and Instantaneous Safety (ensuring safety at every step versus over trajectories), includes Multi-Agent and Hierarchical settings, and recognizes Interactive and Human-in-the-Loop methods such as Natural Language Constraints[6]. Application Domains and Exploration Strategy Design round out the landscape, connecting theory to practice.

Several active research directions reveal key trade-offs between exploration efficiency and safety guarantees. Works emphasizing provable safety like Provable Guarantees[17] and Zero Constraint Violation[38] contrast with methods that balance constraint satisfaction with learning speed, such as Recovery RL[3] which allows temporary violations with recovery mechanisms. SHAPO[0] sits within the Policy Optimization with Safety Constraints branch alongside Constrained PPO[11] and Feasible Actor Critic[30], sharing their focus on integrating constraint handling directly into policy gradient methods. Compared to Multi Objective Safety[14], which frames safety as one objective among many, SHAPO emphasizes constraint satisfaction as a hard requirement rather than a preference to be traded off. The positioning reflects a broader tension in the field: whether to pursue conservative approaches that guarantee safety from the outset or to enable more aggressive exploration with corrective mechanisms, a question that remains central as methods scale to complex real-world domains.

Related Works in Same Category

The following **3 sibling papers** share the same taxonomy leaf node with the original paper:

1. Model-based safe deep reinforcement learning via a constrained proximal policy optimization algorithm

Authors: Jayant, Ashish Kumar, Bhatnagar, Shalabh, Ashish Kumar Jayant, et al. (6 authors total) | **Year/Venue**: 2022 | **URL**: View paper

Abstract

During initial iterations of training in most Reinforcement Learning (RL) algorithms, agents perform a significant number of random exploratory steps. In the real world, this can limit the practicality of these algorithms as it can lead to potentially dangerous behavior. Hence safe exploration is a critical issue in applying RL algorithms in the real world. This problem has been recently well studied under the Constrained Markov Decision Process (CMDP) Framework, where in addition to single-stag...

Relationship Analysis

Both papers belong to the Policy Optimization with Safety Constraints category, using policy gradient methods that integrate safety constraints into optimization. They overlap in addressing safe exploration through constrained policy optimization frameworks, with both employing Lagrangian relaxation approaches combined with PPO-style updates. The key difference is that SHAPO focuses on epistemic uncertainty through sharpness-aware parameter perturbations to bias learning toward conservative behavior, while the candidate paper emphasizes model-based learning with ensemble dynamics models to improve sample efficiency and reduce cumulative constraint violations during training.

---

2. Safety optimized reinforcement learning via multi-objective policy optimization

Authors: Homayoun Honari, Mehran Ghafarian Tamizi, Homayoun Najjaran | **Year/Venue**: 2024 | **URL**: View paper

Abstract

Safe reinforcement learning (Safe RL) refers to a class of techniques that aim to prevent RL algorithms from violating constraints in the process of decision-making and exploration during trial and error. In this paper, a novel model-free Safe RL algorithm, formulated based on the multi-objective policy optimization framework is introduced where the policy is optimized towards optimality and safety, simultaneously. The optimality is achieved by the environment reward function that is subsequentl...

Relationship Analysis

Both papers belong to the Policy Optimization with Safety Constraints category, employing policy gradient methods that integrate safety considerations into the optimization process. They overlap in addressing safe exploration through policy updates that balance task performance with constraint satisfaction, using critic-based safety evaluation mechanisms. The key difference is that SHAPO uses sharpness-aware optimization to handle epistemic uncertainty via parameter perturbations, while SORL formulates safety as a multi-objective optimization problem with reward shaping based on a safety critic, introducing an aggressiveness parameter to tune the safety-performance tradeoff.

---

3. Feasible actor-critic: Constrained reinforcement learning for ensuring statewise safety

Authors: Ma, Haitong, Haitong Ma, Guan Yang, Yang Guan, et al. (18 authors total) | **Year/Venue**: 2021 | **URL**: View paper

Abstract

The safety constraints commonly used by existing safe reinforcement learning (RL) methods are defined only on expectation of initial states, but allow each certain state to be unsafe, which is unsatisfying for real-world safety-critical tasks. In this paper, we introduce the feasible actor-critic (FAC) algorithm, which is the first model-free constrained RL method that considers statewise safety, e.g, safety for each initial state. We claim that some states are inherently unsafe no matter what p...

Relationship Analysis

Both papers belong to the Policy Optimization with Safety Constraints category, integrating safety considerations directly into policy gradient methods. While SHAPO addresses safe exploration through epistemic uncertainty and sharpness-aware updates that bias learning toward conservative behavior in under-explored regions, FAC focuses on statewise safety constraints ensuring safety for each feasible initial state rather than just expected safety. The key difference is that SHAPO uses parameter perturbation sensitivity as a proxy for uncertainty to guide safe exploration, whereas FAC employs a multiplier network to enforce and indicate feasibility at the state level, providing guarantees for individual states rather than trajectory expectations.

Contributions Analysis

Overall novelty summary. The paper proposes Sharpness-Aware Policy Optimization (SHAPO), which uses parameter perturbations as a proxy for epistemic uncertainty to guide safe exploration. It resides in the 'Policy Optimization with Safety Constraints' leaf, which contains four papers including the original work. This leaf sits within the broader 'Algorithm Design for Safe Exploration' branch, indicating a moderately populated research direction focused on integrating safety directly into policy gradient updates rather than using post-hoc corrections or model-based planning.

The taxonomy reveals neighboring approaches that handle safety through different mechanisms. The 'Safety Layer and Action Correction Methods' leaf (three papers) applies analytical filters after policy decisions, while 'Model-Based Safe RL' (two papers) leverages dynamics models for predictive safety. The 'Uncertainty Quantification and Epistemic Safety' branch addresses similar concerns about unknown risks but through conservative critics and probabilistic constraints rather than sharpness-aware updates. SHAPO's epistemic uncertainty framing connects conceptually to this branch, though it operationalizes uncertainty differently via parameter sensitivity rather than explicit distributional modeling.

Among thirteen candidates examined across three contributions, none were identified as clearly refuting the work. The core SHAPO method examined five candidates with zero refutations, while the analytical gradient reweighting characterization examined eight candidates, also with zero refutations. The reinterpretation of Fisher-SAM as epistemic pessimism examined no candidates. This limited search scope—thirteen papers from semantic retrieval—suggests the analysis captures closely related policy optimization methods but may not cover the full breadth of uncertainty-driven safe exploration approaches or sharpness-aware techniques from adjacent fields.

Given the search scale, the work appears to occupy a relatively distinct position within policy optimization methods, combining sharpness awareness with safety constraints in a novel way. However, the analysis does not exhaustively cover connections to broader sharpness-aware learning literature or alternative epistemic uncertainty quantification methods outside the top-thirteen semantic matches. The taxonomy structure suggests this is an active but not overcrowded research direction, with room for differentiation among the four sibling papers in the same leaf.

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

Contribution 1: Sharpness-Aware Policy Optimization (SHAPO) method

Description: SHAPO is a novel policy update method that computes gradients at perturbed parameters to incorporate the actor's epistemic uncertainty. This approach makes policy updates pessimistic by evaluating the gradient at an adjusted parameter that minimizes expected return within a trust region, thereby promoting conservative behavior in under-explored regions.

This contribution was assessed against **5 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Domain-inspired sharpness-aware minimization under domain shifts

URL: View paper

Brief Assessment

Domain Sharpness Aware[52] focuses on domain generalization in supervised learning with domain shifts, not reinforcement learning policy optimization with epistemic uncertainty.

---

2. Momentum-sam: Sharpness aware minimization without computational overhead

URL: View paper

Brief Assessment

Momentum SAM[51] focuses on supervised learning optimization for neural networks, not reinforcement learning policy optimization. The candidate addresses computational efficiency in sharpness-aware minimization for classification tasks, while the original proposes a novel policy update method for safe exploration in RL using epistemic uncertainty.

---

3. Generalizable Prompt Learning via Gradient Constrained Sharpness-Aware Minimization

URL: View paper

Brief Assessment

Gradient Constrained SAM[54] addresses prompt learning for vision-language models with a focus on maintaining performance across seen and unseen classes, while SHAPO targets safe exploration in reinforcement learning by computing gradients at perturbed parameters to handle epistemic uncertainty in policy optimization. These are fundamentally different application domains and technical objectives.

---

4. Improving generalization of robot locomotion policies via Sharpness-Aware Reinforcement Learning

URL: View paper

Brief Assessment

Robot Locomotion SAM[53] applies sharpness-aware optimization to robot locomotion policies for improved generalization to environmental variations, not for safe exploration through epistemic uncertainty. The candidate focuses on sim-to-real transfer robustness in contact-rich environments, while the original addresses safe exploration in constrained MDPs by evaluating gradients at perturbed parameters to handle epistemic uncertainty.

---

5. Distribution-Free Uncertainty Quantification for Kernel Methods by Gradient Perturbations

URL: View paper

Brief Assessment

Gradient Perturbations[55] focuses on uncertainty quantification for kernel methods in supervised learning by perturbing residuals in gradient computations. SHAPO addresses safe exploration in reinforcement learning by computing policy gradients at perturbed parameters to incorporate epistemic uncertainty, which is a fundamentally different problem domain and methodology.

Contribution 2: Reinterpretation of Fisher-SAM as pessimism under epistemic uncertainty

Description: The authors provide a theoretical reinterpretation showing that the sharpness-aware parameter perturbation corresponds to optimizing under epistemic uncertainty about policy parameters. They demonstrate that the adjusted parameter can be viewed as the most likely parameter falling in the lower tail of the uncertainty distribution, thereby formalizing the pessimistic bias.

This contribution was assessed against **0 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

Contribution 3: Analytical characterization of gradient reweighting for rare actions

Description: Through analysis in a simplified Gaussian policy setting, the authors show that SHAPO's gradient modification assigns greater weight to rare unsafe actions (negative advantage) while downweighting rare safe actions (positive advantage). This reweighting mechanism explains how SHAPO promotes safe exploration by treating unsafe rare events more seriously during policy updates.

This contribution was assessed against **8 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Primal-Dual Policy Gradient and Augmented MDP Approaches

URL: View paper

Brief Assessment

Primal Dual Policy[61] focuses on Lp risk-constrained RL with primal-dual policy gradient methods for handling nonlinear risk constraints, not on gradient reweighting mechanisms for rare unsafe actions in safe exploration contexts.

---

2. Probabilistic constraint for safety-critical reinforcement learning

URL: View paper

Brief Assessment

Probabilistic Constraint[43] focuses on gradient expressions for probabilistic safety constraints in RL, not on gradient reweighting mechanisms that amplify rare unsafe actions as in SHAPO's analytical framework.

---

3. Scalable Safe Reinforcement Learning via Neural Approximations of Control-Theoretic Regulators

URL: View paper

Brief Assessment

Neural Approximations Regulators[57] focuses on control-theoretic regulators for safe RL, not on gradient reweighting mechanisms for rare unsafe actions. The candidate's approach uses neural approximations of control-theoretic methods rather than analyzing policy gradient modifications.

---

4. Model-based safe deep reinforcement learning via a constrained proximal policy optimization algorithm

URL: View paper

Brief Assessment

Constrained PPO[11] focuses on model-based safe RL using Lagrangian relaxation with PPO, without analyzing gradient reweighting mechanisms for rare actions. The paper does not provide analytical characterization of how policy gradients treat rare unsafe versus safe actions differently.

---

5. Catastrophic-risk-aware reinforcement learning with extreme-value-theory-based policy gradients

URL: View paper

Brief Assessment

Extreme Value Theory[59] focuses on catastrophic risk minimization using EVT-based policy gradients for tail risk estimation, not on gradient reweighting mechanisms for rare unsafe versus safe actions during policy updates.

---

6. Deep Reinforcement Learning From Demonstrations to Assist Service Restoration in Islanded Microgrids

URL: View paper

Brief Assessment

Service Restoration Microgrids[60] focuses on safe exploration in microgrid service restoration using imitation learning and action clipping, not on analytical characterization of policy gradient reweighting mechanisms for rare unsafe actions.

---

7. Task-Oriented Learning from Positive-Unlabeled Data: Addressing Imbalance, Bias, and Uncertainty

URL: View paper

Brief Assessment

Positive Unlabeled Learning[58] addresses task-oriented learning from imbalanced data through loss reweighting, not policy gradient modifications for safe RL exploration with rare unsafe actions.

---

8. Safe Reinforcement Learning through Phasic Safety-Oriented Policy Optimization.

URL: View paper

Brief Assessment

Phasic Safety Optimization[56] focuses on periodic safety updates through auxiliary policies and safety shields in model-free environments, not on analytical characterization of gradient reweighting mechanisms for rare unsafe actions.

Appendix: Text Similarity Detection

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

References

- [0] SHAPO: Sharpness-Aware Policy Optimization for Safe Exploration View paper
- [1] Safe exploration in reinforcement learning: A generalized formulation and algorithms View paper
- [2] Safe exploration for reinforcement learning. View paper
- [3] Recovery rl: Safe reinforcement learning with learned recovery zones View paper
- [4] Safe reinforcement learning via hierarchical adaptive chance-constraint safeguards View paper
- [5] Safe exploration in reinforcement learning: Theory and applications in robotics View paper
- [6] Safe reinforcement learning with natural language constraints View paper
- [7] Safe reinforcement learning on the constraint manifold: Theory and applications View paper
- [8] Conservative safety critics for exploration View paper
- [9] Reinforcement learning by guided safe exploration View paper
- [10] Safe exploration in continuous action spaces View paper
- [11] Model-based safe deep reinforcement learning via a constrained proximal policy optimization algorithm View paper
- [12] Safety-constrained reinforcement learning with a distributional safety critic View paper
- [13] Safety-constrained reinforcement learning for MDPs View paper
- [14] Safety optimized reinforcement learning via multi-objective policy optimization View paper
- [15] Safe exploration techniques for reinforcement learningâ□□an overview View paper
- [16] Robot reinforcement learning on the constraint manifold View paper
- [17] Safe Exploration for Constrained Reinforcement Learning with Provable Guarantees View paper
- [18] Constraints driven safe reinforcement learning for autonomous driving decision-making View paper
- [19] Learning with safety constraints: Sample complexity of reinforcement learning for constrained mdps View paper
- [20] Actsafe: Active exploration with safety constraints for reinforcement learning View paper
- [21] Adaptive Safety-Certified Reinforcement Learning for Constrained Optimal Control of Autonomous Robots With Uncertainties View paper
- [22] Safe reinforcement learning in constrained markov decision processes View paper
- [23] Constrained meta-reinforcement learning for adaptable safety guarantee with differentiable convex programming View paper
- [24] Cem: Constrained entropy maximization for task-agnostic safe exploration View paper
- [25] Safe Reinforcement Learning for Constrained Optimal Control With Provable Guarantees: Applications to Motion Planning View paper

- [26] A Survey of Safe Reinforcement Learning and Constrained MDPs: A Technical Survey on Single-Agent and Multi-Agent Safety View paper
- [27] Safe reinforcement learning with instantaneous constraints: the role of aggressive exploration View paper
- [28] Safe exploration for interactive machine learning View paper
- [29] A survey of constraint formulations in safe reinforcement learning View paper
- [30] Feasible actor-critic: Constrained reinforcement learning for ensuring statewise safety View paper
- [31] Alwayssafe: Reinforcement learning without safety constraint violations during training View paper
- [32] Explicit Explore, Exploit, or Escape (): near-optimal safety-constrained reinforcement learning in polynomial time View paper
- [33] Near-optimal Conservative Exploration in Reinforcement Learning under Episode-wise Constraints View paper
- [34] Enhance Exploration in Safe Reinforcement Learning with Contrastive Representation Learning View paper
- [35] Enhancing Autonomous Lane-Changing Safety: Deep Reinforcement Learning via Pre-Exploration in Parallel Imaginary Environments View paper
- [36] Reinforcement learning with safe exploration for network security View paper
- [37] Constrained Reinforcement Learning for Safe Heat Pump Control View paper
- [38] Learn zero-constraint-violation safe policy in model-free constrained reinforcement learning View paper
- [39] Reinforcement learning with safety and stability guarantees during exploration for linear systems View paper
- [40] Safe Exploration in Reinforcement Learning for Learning from Human Experts View paper
- [41] Safe exploration and optimization of constrained mdps using gaussian processes View paper
- [42] Rethinking Safe Policy Learning for Complex Constraints Satisfaction: A Glimpse in Real-Time Security Constrained Economic Dispatch Integrating Energy Storage Units View paper
- [43] Probabilistic constraint for safety-critical reinforcement learning View paper
- [44] Safe Reinforcement Learning with Constraints: A Survey View paper
- [45] Safe Reinforcement Learning via Episodic Control View paper
- [46] Safe Reinforcement Learning in Autonomous Driving With Epistemic Uncertainty Estimation View paper
- [47] Real-Time Sequential Security-Constrained Optimal Power Flow: A Hybrid Knowledge-Data-Driven Reinforcement Learning Approach View paper
- [48] Safe Exploration Method for Reinforcement Learning under Existence of Disturbance View paper
- [49] Safe Reinforcement Learning for Energy Management of Electrified Vehicle With Novel Physics-Informed Exploration Strategy View paper
- [50] Off-Policy Conservative Distributional Reinforcement Learning With Safety Constraints View paper
- [51] Momentum-sam: Sharpness aware minimization without computational overhead View paper
- [52] Domain-inspired sharpness-aware minimization under domain shifts View paper
- [53] Improving generalization of robot locomotion policies via Sharpness-Aware Reinforcement Learning View paper
- [54] Generalizable Prompt Learning via Gradient Constrained Sharpness-Aware Minimization View paper
- [55] Distribution-Free Uncertainty Quantification for Kernel Methods by Gradient Perturbations View paper
- [56] Safe Reinforcement Learning through Phasic Safety-Oriented Policy Optimization. View paper
- [57] Scalable Safe Reinforcement Learning via Neural Approximations of Control-Theoretic Regulators View paper
- [58] Task-Oriented Learning from Positive-Unlabeled Data: Addressing Imbalance, Bias, and Uncertainty View paper
- [59] Catastrophic-risk-aware reinforcement learning with extreme-value-theory-based policy gradients View paper
- [60] Deep Reinforcement Learning From Demonstrations to Assist Service Restoration in Islanded Microgrids View paper
- [61] Primal-Dual Policy Gradient and Augmented MDP Approaches View paper