Novelty Assessment Report

Paper: Reasoning via Test-Time Instance-Level Policy Gradient in Latent Space

PDF URL: https://openreview.net/pdf?id=5ENCXZyQCK

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-07

Abstract

Large Language Models (LLMs) typically reason through explicit, step-by-step natural-language traces. Humans, however, also rely on non-linguistic, unconscious processes, such as the inspirations that emerge during the incubation period. In this work, we introduce LatentSeek, a novel framework designed to enhance the reasoning capabilities of LLMs through Test-Time Instance-level Policy Gradient within the model's latent space—thus complementing explicit natural-language steps. LatentSeek employs policy gradient optimization to iteratively refine latent representations, guided solely by a self-generated reward signal. This allows the model to adapt its reasoning trajectory dynamically on a per-instance basis. Empirical evaluations across diverse benchmarks, GSM8K, MATH-500, and AIME2024 as well as multiple LLM families (e.g., LLaMA, Qwen) demonstrate that LatentSeek outperforms established baselines, including Chain-of-Thought (CoT), Best-of-N (BoN) and training-based methods. Further analysis indicates that LatentSeek is computationally efficient, typically converging within a few optimization iterations for average-level problems. Moreover, the model's performance improves as the number of latent update iterations increases, highlighting the benefits of exploring within the latent space. These findings highlight LatentSeek as a lightweight and effective paradigm for improving the reasoning capabilities of LLMs without changing their parameters.

Core Task Landscape

This paper addresses: **Test-Time Reasoning Enhancement Through Latent Space Policy Gradient Optimization**

A total of **20 papers** were analyzed and organized into a taxonomy with **18 categories**.

Taxonomy Overview

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

- **Test-Time Latent Reasoning Optimization**
- **Training-Based Latent Reasoning Frameworks**
- **Multimodal Latent Visual Reasoning**
- **Latent Reasoning for Domain-Specific Applications**
- **Latent Skill and State Representation Learning**

Complete Taxonomy Tree

- Test-Time Reasoning Enhancement Through Latent Space Policy Gradient Optimization Survey Taxonomy
- Test-Time Latent Reasoning Optimization
  - Policy Gradient-Based Latent Optimization ★ (3 papers)
  - [0] Reasoning via Test-Time Instance-Level Policy Gradient in Latent Space (Anon et al., 2026) View paper
  - [1] Seek in the Dark: Reasoning via Test-Time Instance-Level Policy Gradient in Latent Space (Li Hengli, 2025) View paper
  - [11] Thinking on the Fly: Test-Time Reasoning Enhancement via Latent Thought Policy Optimization (Liang Yan, 2025) View paper
  - Adaptive Compute Allocation in Latent Space (1 papers)
  - [15] Learning to Ponder: Adaptive Reasoning in Latent Space (He Yixin, 2025) View paper
  - Multimodal Latent Reasoning at Test Time (1 papers)
  - [10] MILR: Improving Multimodal Image Generation via Test-Time Latent Reasoning (Mi Ya-peng, 2025) View paper
- Training-Based Latent Reasoning Frameworks
  - Reinforcement Learning for Latent Reasoning (2 papers)
  - [3] Hybrid Latent Reasoning via Reinforcement Learning (Yue, 2025) View paper
  - [17] Reinforcement Learning for Latent-Space Thinking in LLMs (Enes Ãzeren, 2025) View paper
  - Latent State Transition Modeling (1 papers)
  - [6] Ctrls: Chain-of-thought reasoning via latent state-transition (Wu, 2025) View paper
  - Compressed Latent Reasoning (1 papers)
  - [12] Think Silently, Think Fast: Dynamic Latent Compression of LLM Reasoning Chains (Tan Wen-hui, 2025) View paper
  - Latent Reward Signal Analysis (1 papers)
  - [16] Latent Thinking Optimization: Your Latent Reasoning Language Model Secretly Encodes Reward Signals in Its Latent Thoughts (Du, 2025) View paper
- Multimodal Latent Visual Reasoning
  - Autoregressive Visual Latent Reasoning (1 papers)
  - [5] Latent visual reasoning (Li, 2025) View paper
  - Continuous Visual Thought Generation (1 papers)
  - [4] Monet: Reasoning in latent visual space beyond images and language (Qixun Wang, 2025) View paper
  - Dynamic Multimodal Interleaving (1 papers)
  - [18] Reasoning Within the Mind: Dynamic Multimodal Interleaving in Latent Space (Chengzhi Liu, 2025) View paper

- Latent Reasoning for Domain-Specific Applications
  - Sequential Recommendation Systems (1 papers)
  - [2] LARES: Latent Reasoning for Sequential Recommendation (Liu En-Ze, 2025) View paper
  - Combinatorial Optimization (1 papers)
  - [14] Combinatorial Optimization with Policy Adaptation using Latent Space Search (Chalumeau, 2023) View paper
  - Robotic Control Policies (1 papers)
  - [7] Steering Your Diffusion Policy with Latent Space Reinforcement Learning (Wagenmaker, 2025) View paper
- Latent Skill and State Representation Learning
  - Latent Reasoning Skill Abstraction (1 papers)
  - [8] LaRS: Latent reasoning skills for chain-of-thought reasoning (Qi Yanjun, 2024) View paper
  - Causal State Representation for Offline RL (1 papers)
  - [9] Policy-Guided Causal State Representation for Offline Reinforcement Learning Recommendation (Siyu Wang, 2025) View paper
  - Hierarchical Latent Space Policies (1 papers)
  - [20] Latent Space Policies for Hierarchical Reinforcement Learning (Tuomas Haarnoja, 2018) View paper
  - Multi-Agent Value Function Factorization (1 papers)
  - [13] Value Functions Factorization With Latent State Information Sharing in Decentralized Multi-Agent Policy Gradients (Hanhan Zhou, 2023) View paper
  - Safe Driving State Representation (1 papers)
  - [19] Policy-Gradient and Actor-Critic Based State Representation Learning for Safe Driving of Autonomous Vehicles. (Abhishekh Gupta, 2020) View paper

Narrative

Core task: test-time reasoning enhancement through latent space policy gradient optimization. The field centers on improving model reasoning capabilities by optimizing latent representations at inference time, rather than relying solely on pre-trained weights. The taxonomy reveals several major branches: Test-Time Latent Reasoning Optimization focuses on methods that adapt or search within latent spaces during inference, often using policy gradients or search strategies to refine intermediate reasoning steps. Training-Based Latent Reasoning Frameworks emphasize learning structured latent representations during the training phase that facilitate downstream reasoning. Multimodal Latent Visual Reasoning extends these ideas to vision-language settings, where latent codes must bridge modalities. Domain-Specific Applications tailor latent reasoning to particular tasks such as robotics or autonomous driving, while Latent Skill and State Representation Learning explores how to discover and leverage abstract state or skill embeddings. Representative works like Seek in the Dark[1] and LARES[2] illustrate test-time optimization approaches, whereas Monet[4] and Latent Visual Reasoning[5] highlight multimodal extensions.

Within the test-time optimization branch, a particularly active line of work explores policy gradient-based methods for refining latent reasoning trajectories on the fly. Test-Time Policy Gradient[0] sits squarely in this cluster, emphasizing direct policy gradient updates in latent space to enhance reasoning quality without additional training. This contrasts with nearby approaches: Seek in the Dark[1] employs search-based strategies to navigate latent spaces, while Thinking on the Fly[11] and Think Silently Fast[12] investigate adaptive computation budgets and silent reasoning tokens. The central trade-off across these methods involves balancing computational overhead at test time against the gains in reasoning accuracy or robustness. Test-Time Policy Gradient[0] distinguishes itself by leveraging policy gradients to iteratively improve latent representations, positioning it as a gradient-driven alternative to search-heavy or token-based reasoning augmentation strategies.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Seek in the Dark: Reasoning via Test-Time Instance-Level Policy Gradient in Latent Space

Authors: Li Hengli, Li Chenxi, Wu, Tong, Zhu, et al. (14 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Reasoning ability, a core component of human intelligence, continues to pose a significant challenge for Large Language Models (LLMs) in the pursuit of AGI. Although model performance has improved under the training scaling law, significant challenges remain, particularly with respect to training algorithms, such as catastrophic forgetting, and the limited availability of novel training data. As an alternative, test-time scaling enhances reasoning performance by increasing test-time computation ...

⚠ Similarity Notice

These papers appear to be the same work or closely related variants. Both introduce 'LatentSeek' (or 'LATENTSEEK'), a framework for test-time reasoning enhancement through policy gradient optimization in latent space. The titles, core methodology (test-time instance-level policy gradient in latent space), technical approach (using REINFORCE for latent representation optimization), and evaluation benchmarks (GSM8K, MATH-500, AIME2024) are nearly identical, strongly suggesting they are the same paper or different versions of the same submission.

---

2. Thinking on the Fly: Test-Time Reasoning Enhancement via Latent Thought Policy Optimization

Authors: Liang Yan, Shan LianLei | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Recent advancements in Large Language Models (LLMs) have shifted from explicit Chain-of-Thought (CoT) reasoning to more efficient latent reasoning, where intermediate thoughts are represented as vectors rather than text. However, latent reasoning can be brittle on challenging, out-of-distribution tasks where robust reasoning is most critical. To overcome these limitations, we introduce Latent Thought Policy Optimization (LTPO), a parameter-free framework that enhances LLM reasoning entirely at t...

Relationship Analysis

Both papers belong to the same taxonomy category of policy gradient-based latent optimization for test-time reasoning enhancement. They share the core approach of using policy gradient methods to iteratively refine latent representations during inference without model parameter updates, and both employ self-derived reward signals from the frozen LLM. The key differences are: (1) LATENTSEEK optimizes token-wise latent representations independently using REINFORCE with self-generated rewards based on full autoregressive decoding, while LTPO (Thinking on the Fly) treats latent thought vectors as dynamic parameters optimized via a confidence-based reward computed directly from output distributions without text generation during optimization; (2) LTPO introduces special 'latent thought tokens' as placeholders, whereas LATENTSEEK operates on existing reasoning sequence representations; (3) LTPO emphasizes computational efficiency by avoiding text generation in the RL loop, while LATENTSEEK performs full decoding at each iteration.

Contributions Analysis

Overall novelty summary. The paper introduces LatentSeek, a framework that applies policy gradient optimization to refine latent representations at test time for improved reasoning. It resides in the 'Policy Gradient-Based Latent Optimization' leaf, which contains only three papers total, indicating a relatively sparse research direction within the broader taxonomy of test-time latent reasoning. This leaf sits under 'Test-Time Latent Reasoning Optimization,' a branch that contrasts with training-based methods and multimodal approaches, suggesting the work occupies a focused niche exploring gradient-driven test-time adaptation rather than search-based or training-heavy alternatives.

The taxonomy reveals neighboring leaves such as 'Adaptive Compute Allocation in Latent Space' and 'Multimodal Latent Reasoning at Test Time,' which explore dynamic resource allocation and cross-modal reasoning respectively. LatentSeek diverges from these by concentrating on policy gradient updates within a single modality's latent space, rather than multimodal fusion or adaptive compute budgets. The broader 'Training-Based Latent Reasoning Frameworks' branch contains methods like reinforcement learning for latent reasoning and latent state transition modeling, which differ fundamentally by requiring parameter updates during training rather than test-time optimization alone.

Among the three contributions analyzed, the first two—LatentSeek framework and policy gradient optimization method—appear relatively novel within the limited search scope of 29 candidates, with zero refutable candidates found across 19 examined papers. The third contribution, test-time scaling analysis, encountered one refutable candidate among 10 examined, suggesting some prior work exists on analyzing computational scaling in latent reasoning. The statistics indicate that while the core framework and optimization approach show limited overlap with the examined literature, the scaling analysis component has more substantial prior coverage, though the search scope remains modest.

Based on the limited top-K semantic search and citation expansion covering 29 candidates, the work appears to occupy a sparsely populated research direction with minimal direct overlap in its core contributions. However, the analysis does not cover exhaustive literature review, and the single refutable pair for the scaling contribution suggests adjacent work exists. The taxonomy structure confirms this is an emerging area with few sibling papers, though definitive novelty claims would require broader literature coverage beyond the examined candidate set.

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

Contribution 1: LATENTSEEK framework for test-time instance-level policy gradient in latent space

Description: The authors propose LATENTSEEK, a framework that enhances LLM reasoning by performing test-time optimization of latent representations using policy gradient methods. Unlike training-based approaches, it operates on frozen models and dynamically refines reasoning trajectories for each problem instance without 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. Dynamic Prompt Learning via Policy Gradient for Semi-structured Mathematical Reasoning

URL: View paper

Brief Assessment

Dynamic Prompt Learning[29] focuses on learning to select in-context examples for few-shot GPT-3 using policy gradient, not on optimizing latent representations within frozen LLMs for reasoning tasks.

---

2. Difftori: Differentiable trajectory optimization for deep reinforcement and imitation learning

URL: View paper

Brief Assessment

Difftori[32] focuses on differentiable trajectory optimization for reinforcement learning and imitation learning in robotics, not test-time policy gradient optimization for language model reasoning in latent space.

---

3. MILR: Improving Multimodal Image Generation via Test-Time Latent Reasoning

URL: View paper

Brief Assessment

MILR[10] focuses on multimodal image generation via test-time latent reasoning, not general language model reasoning. The candidate optimizes unified image-text latent representations for image synthesis, while the original addresses LLM reasoning enhancement through latent-space policy gradients.

---

4. Rl of thoughts: Navigating llm reasoning with inference-time reinforcement learning

URL: View paper

Brief Assessment

RL of Thoughts[31] focuses on training a navigator model to select and combine predefined logic blocks for constructing task-specific reasoning structures, not on optimizing latent representations through policy gradient methods in the latent space of frozen LLMs.

---

5. Combinatorial Optimization with Policy Adaptation using Latent Space Search

URL: View paper

Brief Assessment

Latent Space Search[14] focuses on combinatorial optimization problems (TSP, CVRP, JSSP) using a latent space of diverse policies for search, while LATENTSEEK addresses LLM reasoning enhancement through test-time policy gradient optimization in latent representations without parameter updates. The problem domains, objectives, and technical approaches are fundamentally different.

---

6. Seek in the Dark: Reasoning via Test-Time Instance-Level Policy Gradient in Latent Space

URL: View paper

Brief Assessment

Seek in the Dark[1] focuses on test-time instance-level adaptation (TTIA) within latent space using policy gradient methods. While both papers explore test-time optimization in latent space, the candidate does not demonstrate that this specific framework existed prior to the original work.

---

7. Retroformer: Retrospective Large Language Agents with Policy Gradient Optimization

URL: View paper

Brief Assessment

Retroformer[33] focuses on optimizing a retrospective model that refines prompts through verbal feedback across multiple trials, not on test-time optimization of latent representations within a single instance using policy gradients in latent space.

8. Latent visual reasoning

URL: View paper

Brief Assessment

Latent Visual Reasoning[5] focuses on multimodal visual reasoning by reconstructing visual tokens in a joint semantic space, not on test-time policy gradient optimization for general language model reasoning tasks.

9. Step-Aware Policy Optimization for Reasoning in Diffusion Large Language Models

URL: View paper

Brief Assessment

Step-Aware Policy[30] focuses on training-time policy optimization for diffusion language models using process-based rewards to structure reasoning hierarchies, not test-time optimization of frozen models in latent space as LATENTSEEK does.

10. Hybrid Latent Reasoning via Reinforcement Learning

URL: View paper

Brief Assessment

Hybrid Latent Reasoning[3] focuses on training-time RL optimization with a gating mechanism to blend discrete tokens and continuous representations, whereas the original paper performs test-time optimization on frozen models without parameter updates.

Contribution 2: Policy gradient optimization method for latent representations

Description: The authors develop a policy gradient-based optimization procedure that iteratively updates token-wise latent representations guided by self-generated reward signals. This method treats latent representations as independent variables and uses REINFORCE to perform gradient ascent in the latent space.

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. Meta-reinforcement learning algorithm based on reward and dynamic inference

URL: View paper

Brief Assessment

Meta-RL Reward Inference[21] focuses on task latent variable inference in meta-reinforcement learning contexts, not on optimizing token-wise latent representations for reasoning tasks using self-generated rewards as in the original paper.

2. Kalman Filter Enhanced GRPO for Reinforcement Learning-Based Language Model Reasoning

URL: View paper

Brief Assessment

Kalman Enhanced GRPO[27] focuses on improving advantage estimation in policy gradient methods for language model outputs using Kalman filtering, not on optimizing latent representations. The original paper optimizes token-wise latent representations in the model's hidden space, while this candidate addresses advantage function estimation in the output space.

3. Steering Your Diffusion Policy with Latent Space Reinforcement Learning

URL: View paper

Brief Assessment

Steering Diffusion Policy[7] optimizes latent-noise space for diffusion policies in robotic control, not token-wise latent representations for language model reasoning. The technical domains and optimization targets differ fundamentally.

4. Seek in the Dark: Reasoning via Test-Time Instance-Level Policy Gradient in Latent Space

URL: View paper

Brief Assessment

Seek in the Dark[1] employs policy gradient to iteratively update latent representations guided by self-generated rewards. However, the candidate does not provide evidence that this specific approach to optimizing latent representations using policy gradients for LLM reasoning was previously established.

5. Latent Safety-Constrained Policy Approach for Safe Offline Reinforcement Learning

URL: View paper

Brief Assessment

Latent Safety-Constrained[25] focuses on safe offline RL using CVAEs to model safety constraints in latent space, not on policy gradient optimization of latent representations for reasoning tasks. The candidate addresses safety-constrained policy learning, while the original work optimizes latent representations for test-time reasoning enhancement.

6. Interpretable multi-agent reinforcement learning via multi-head variational autoencoders

URL: View paper

Brief Assessment

Multi-Head VAE[24] focuses on multi-agent reinforcement learning with variational autoencoders for interpretability in agent coordination, not on optimizing latent representations for reasoning tasks using policy gradients and self-generated rewards as in the original paper.

7. Test-Time Policy Adaptation for Enhanced Multi-Turn Interactions with LLMs

URL: View paper

Brief Assessment

Test-Time Policy Adaptation[22] focuses on adapting model parameters during multi-turn interactions using user feedback as rewards, not on optimizing latent representations token-by-token using REINFORCE in latent space as the original paper does.

8. Reasoning with latent diffusion in offline reinforcement learning

URL: View paper

Brief Assessment

Latent Diffusion Reasoning[28] focuses on diffusion models for offline RL with latent trajectory representations, not policy gradient optimization of latent representations using reward signals for inference as described in the original paper.

9. Visual reinforcement learning with imagined goals

URL: View paper

Brief Assessment

Visual Imagined Goals[23] focuses on goal-conditioned reinforcement learning for robotic manipulation using VAE-based latent representations, not on policy gradient optimization of latent representations themselves for inference-time reasoning.

Contribution 3: Test-time scaling analysis in latent space

Description: The authors demonstrate that reasoning performance improves as the number of latent-space optimization iterations increases, establishing a complementary scaling dimension beyond token generation. This reveals that exploration within the latent space offers a promising direction for test-time scaling.

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. REFINER: Reasoning Feedback on Intermediate Representations

URL: View paper

Brief Assessment

REFINER[36] focuses on iterative refinement of intermediate reasoning steps through structured feedback from a critic model, not on test-time scaling through latent-space optimization iterations. The candidate's framework involves a generator-critic interaction loop where feedback improves reasoning quality, which is fundamentally different from exploring latent space via policy gradient optimization to achieve test-time scaling.

---

2. Scaling up Test-Time Compute with Latent Reasoning: A Recurrent Depth Approach

URL: View paper

Prior Art Analysis

Recurrent Depth Reasoning[34] demonstrates that reasoning performance improves as the number of latent-space recurrent iterations increases at test-time, establishing test-time scaling through iterative optimization in latent space. The candidate paper shows that by iterating a recurrent block at test-time (up to 64 iterations), the model can improve performance on reasoning benchmarks like GSM8K and ARC Challenge. This directly demonstrates prior work on test-time scaling through latent-space optimization iterations, refuting the novelty claim that the original authors were first to explore this dimension.

Evidence

Evidence 1 - **Rationale**: Both papers describe test-time performance improvements through increased iterations in latent space, showing that this scaling dimension was explored in prior work. - **Original**: the model's performance improves as the number of latent update iterations increases, highlighting the benefits of exploring within the latent space - **Candidate**: we train a 3.5b parameter language model with depth recurrence. at test time, the model can iterate longer to use more compute and improve its performance. instead of scaling test-time reasoning by "verbalizing" in long chains-of-thought, the model improves entirely by reasoning in latent space.

Evidence 2 - **Rationale**: Both papers explicitly analyze and demonstrate test-time scaling through increased latent-space iterations, showing this was explored in prior work published before the original submission. - **Original**: we conduct a scaling analysis, revealing that performance at test time improves with an increased number of update iterations, highlighting the potential of test-time scaling in the latent space - **Candidate**: as we increase compute, the performance on these benchmarks increases. hellaswag only needs 8 recurrences to achieve near peak performance while other benchmarks make use of more compute... performance on gsm8k cot (strict match and flexible match), hellaswag (acc norm.), and humaneval (pass@1). as ...

---

3. MILR: Improving Multimodal Image Generation via Test-Time Latent Reasoning

URL: View paper

Brief Assessment

MILR[10] demonstrates test-time scaling for image generation tasks by increasing optimization steps, but this is applied to multimodal image synthesis rather than general reasoning performance in language models.

---

4. Self-Refine: Iterative Refinement with Self-Feedback

URL: View paper

Brief Assessment

Self-Refine[35] focuses on iterative refinement through natural language feedback in the output space, not latent-space optimization. The candidate operates on explicit text outputs rather than continuous latent representations, representing a fundamentally different approach to test-time improvement.

---

5. Inference-time alignment in continuous space

URL: View paper

Brief Assessment

Inference-Time Alignment[41] focuses on inference-time alignment through energy-based optimization in continuous logit space for safety/truthfulness tasks, not on test-time scaling for reasoning performance through iterative latent-space optimization as demonstrated in the original paper.

---

6. A survey of scaling in large language model reasoning

URL: View paper

Brief Assessment

Scaling LLM Reasoning[39] is a survey paper that reviews existing work on test-time scaling but does not present original empirical demonstrations of latent-space optimization improving with iteration count. The candidate discusses latent-space reasoning as one dimension among many scaling strategies, whereas the original paper presents novel experimental evidence of performance gains through iterative latent optimization.

---

7. Think before recommend: Unleashing the latent reasoning power for sequential recommendation

URL: View paper

Brief Assessment

Think Before Recommend[42] focuses on sequential recommendation systems using multi-step reasoning during inference, not on general test-time scaling through latent-space optimization for reasoning performance in language models.

---

8. MAgICoRe: Multi-Agent, Iterative, Coarse-to-Fine Refinement for Reasoning

URL: View paper

Brief Assessment

MAgICoRe[40] focuses on test-time scaling through iterative multi-agent refinement of explicit reasoning chains, not latent-space optimization. The candidate operates on natural language outputs with external reward models, whereas the original explores optimization within the model's continuous latent representations.

---

9. Iterative Reasoning Preference Optimization

URL: View paper

Brief Assessment

Iterative Reasoning Preference[38] focuses on iterative preference optimization during training through multiple model iterations (m1, m2, m3, m4), not test-time optimization in latent space. The candidate's iterations involve retraining model parameters with preference pairs, fundamentally different from the original's test-time latent representation updates without parameter changes.

---

10. Investigating inference-time scaling for chain of multi-modal thought: A preliminary study

URL: View paper

Brief Assessment

Multi-Modal Thought Scaling[37] focuses on inference-time scaling through multi-modal (text+image) thought chains in vision-language models, not latent-space optimization. The candidate explores scaling via sampling and tree search methods over explicit multi-modal reasoning steps, whereas the original paper optimizes continuous latent representations via policy gradient.

Appendix: Text Similarity Detection

Textual similarity detection checked 29 papers and found 1 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. MILR: Improving Multimodal Image Generation via Test-Time Latent Reasoning

Detected in: Contribution: contribution_1, Contribution: contribution_3

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] Reasoning via Test-Time Instance-Level Policy Gradient in Latent Space View paper
- [1] Seek in the Dark: Reasoning via Test-Time Instance-Level Policy Gradient in Latent Space View paper
- [2] LARES: Latent Reasoning for Sequential Recommendation View paper
- [3] Hybrid Latent Reasoning via Reinforcement Learning View paper
- [4] Monet: Reasoning in latent visual space beyond images and language View paper
- [5] Latent visual reasoning View paper
- [6] Ctrls: Chain-of-thought reasoning via latent state-transition View paper
- [7] Steering Your Diffusion Policy with Latent Space Reinforcement Learning View paper
- [8] LaRS: Latent reasoning skills for chain-of-thought reasoning View paper
- [9] Policy-Guided Causal State Representation for Offline Reinforcement Learning Recommendation View paper
- [10] MILR: Improving Multimodal Image Generation via Test-Time Latent Reasoning View paper
- [11] Thinking on the Fly: Test-Time Reasoning Enhancement via Latent Thought Policy Optimization View paper
- [12] Think Silently, Think Fast: Dynamic Latent Compression of LLM Reasoning Chains View paper
- [13] Value Functions Factorization With Latent State Information Sharing in Decentralized Multi-Agent Policy Gradients View paper
- [14] Combinatorial Optimization with Policy Adaptation using Latent Space Search View paper
- [15] Learning to Ponder: Adaptive Reasoning in Latent Space View paper
- [16] Latent Thinking Optimization: Your Latent Reasoning Language Model Secretly Encodes Reward Signals in Its Latent Thoughts View paper
- [17] Reinforcement Learning for Latent-Space Thinking in LLMs View paper
- [18] Reasoning Within the Mind: Dynamic Multimodal Interleaving in Latent Space View paper
- [19] Policy-Gradient and Actor-Critic Based State Representation Learning for Safe Driving of Autonomous Vehicles. View paper
- [20] Latent Space Policies for Hierarchical Reinforcement Learning View paper
- [21] Meta-reinforcement learning algorithm based on reward and dynamic inference View paper
- [22] Test-Time Policy Adaptation for Enhanced Multi-Turn Interactions with LLMs View paper
- [23] Visual reinforcement learning with imagined goals View paper
- [24] Interpretable multi-agent reinforcement learning via multi-head variational autoencoders View paper
- [25] Latent Safety-Constrained Policy Approach for Safe Offline Reinforcement Learning View paper
- [26] Rethinking action spaces for reinforcement learning in end-to-end dialog agents with latent variable models View paper
- [27] Kalman Filter Enhanced GRPO for Reinforcement Learning-Based Language Model Reasoning View paper
- [28] Reasoning with latent diffusion in offline reinforcement learning View paper
- [29] Dynamic Prompt Learning via Policy Gradient for Semi-structured Mathematical Reasoning View paper
- [30] Step-Aware Policy Optimization for Reasoning in Diffusion Large Language Models View paper
- [31] Rl of thoughts: Navigating llm reasoning with inference-time reinforcement learning View paper
- [32] Difftori: Differentiable trajectory optimization for deep reinforcement and imitation learning View paper
- [33] Retroformer: Retrospective Large Language Agents with Policy Gradient Optimization View paper
- [34] Scaling up Test-Time Compute with Latent Reasoning: A Recurrent Depth Approach View paper
- [35] Self-Refine: Iterative Refinement with Self-Feedback View paper
- [36] REFINER: Reasoning Feedback on Intermediate Representations View paper
- [37] Investigating inference-time scaling for chain of multi-modal thought: A preliminary study View paper
- [38] Iterative Reasoning Preference Optimization View paper
- [39] A survey of scaling in large language model reasoning View paper

- [40] MAgICoRe: Multi-Agent, Iterative, Coarse-to-Fine Refinement for Reasoning View paper
- [41] Inference-time alignment in continuous space View paper
- [42] Think before recommend: Unleashing the latent reasoning power for sequential recommendation View paper