Novelty Assessment Report

Paper: Reasoning without Training: Your Base Model is Smarter Than You Think

PDF URL: https://openreview.net/pdf?id=Vsgq2ldr4K

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-27

Abstract

Frontier reasoning models have exhibited incredible capabilities across a wide array of disciplines, driven by posttraining large language models (LLMs) with reinforcement learning (RL). However, despite the widespread success of this paradigm, much of the literature has been devoted to disentangling truly novel behaviors that emerge during RL but are not present in the base models. In our work, we approach this question from a different angle, instead asking whether comparable reasoning capabilities can be elicited from base models at inference time by pure sampling, without any additional training. Inspired by Markov chain Monte Carlo (MCMC) techniques for sampling from sharpened distributions, we propose a simple iterative sampling algorithm leveraging the base models' own likelihoods. Over different base models, we show that our algorithm offers substantial boosts in reasoning that nearly match and even outperform those from RL on a wide variety of single-shot tasks, including MATH500, HumanEval, and GPQA. Moreover, our sampler avoids the collapse in diversity over multiple samples that is characteristic of RL-posttraining. Crucially, our method does not require training, curated datasets, or a verifier, suggesting broad applicability beyond easily verifiable domains.

Core Task Landscape

This paper addresses: **eliciting reasoning capabilities from base language models through inference-time sampling**

A total of **50 papers** were analyzed and organized into a taxonomy with **22 categories**.

Taxonomy Overview

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

- **Inference-Time Sampling and Search Strategies**
- **Training-Aware and Hybrid Approaches**
- **Efficiency and Acceleration Techniques**
- **Domain-Specific Applications**
- **Theoretical Foundations and Analysis**
- **Auxiliary Techniques and Mechanisms**

Complete Taxonomy Tree

- eliciting reasoning capabilities from base language models through inference-time sampling Survey Taxonomy
- Inference-Time Sampling and Search Strategies
  - Pure Sampling-Based Methods ★ (4 papers)
  - [0] Reasoning without Training: Your Base Model is Smarter Than You Think (Anon et al., 2026) View paper
  - [14] Reasoning with sampling: Your base model is smarter than you think (Karan, 2025) View paper
  - [18] Flaming-hot Initiation with Regular Execution Sampling for Large Language Models (Chen, 2024) View paper
  - [21] Rollout Roulette: A Probabilistic Inference Approach to Inference-Time Scaling of LLMs using Particle-Based Monte Carlo Methods (Puri, 2025) View paper
  - Verification-Guided Sampling (4 papers)
  - [1] Scaling LLM test-time compute optimally can be more effective than scaling parameters for reasoning (CV Snell, 2025) View paper
  - [5] Sample, scrutinize and scale: Effective inference-time search by scaling verification (Zhao, 2025) View paper
  - [12] GenPRM: Scaling Test-Time Compute of Process Reward Models via Generative Reasoning (Zhao Jian, 2025) View paper
  - [49] Probabilistic Optimality for Inference-time Scaling (Wang Youkang, 2025) View paper
  - Structured Search and Tree-Based Exploration (4 papers)
  - [4] Rl of thoughts: Navigating llm reasoning with inference-time reinforcement learning (Li, 2025) View paper
  - [32] Parathinker: Native parallel thinking as a new paradigm to scale llm test-time compute (Wen Hao, 2025) View paper
  - [38] Forest-of-Thought: Scaling Test-Time Compute for Enhancing LLM Reasoning (Han, 2024) View paper
  - [40] Wider or Deeper? Scaling LLM Inference-Time Compute with Adaptive Branching Tree Search (Inoue Yuichi, 2025) View paper
  - Iterative Refinement and Multi-Round Reasoning (3 papers)
  - [25] Think twice: Enhancing llm reasoning by scaling multi-round test-time thinking (Tian Xiao-yu, 2025) View paper
  - [31] Self-rewarding correction for mathematical reasoning (Xiong Wei, 2025) View paper
  - [36] A Survey of Reinforcement Learning in Large Language Models: From Data Generation to Test-Time Inference (Zichuan Guo, 2025) View paper
- Training-Aware and Hybrid Approaches
  - Inference-Aware Fine-Tuning (2 papers)
  - [8] Inference-aware fine-tuning for best-of-n sampling in large language models (Chow, 2024) View paper
  - [16] Teaching Large Language Models to Reason through Learning and Forgetting (T Ni, 2025) View paper
  - Reinforcement Learning for Reasoning (2 papers)

- [6] Towards large reasoning models: A survey of reinforced reasoning with large language models (Xu Fengli, 2025) View paper
    - [24] Enhancing LLM Reasoning via Critique Models with Test-Time and Training-Time Supervision (Xi, 2024) View paper
    - Test-Time Training and Adaptation (3 papers)
    - [15] The Surprising Effectiveness of Test-Time Training for Abstract Reasoning (Ekin AkyÃ¼rek, 2024) View paper
    - [22] MedAdapter: Efficient Test-Time Adaptation of Large Language Models Towards Medical Reasoning (Shi, 2024) View paper
    - [35] The Surprising Effectiveness of Test-Time Training for Few-Shot Learning (AkyÃ¼rek, 2024) View paper
- Efficiency and Acceleration Techniques
    - Speculative and Accelerated Decoding (2 papers)
    - [9] Specreason: Fast and accurate inference-time compute via speculative reasoning (Pan Rui, 2025) View paper
    - [43] Accelerated Test-Time Scaling with Model-Free Speculative Sampling (Dingliwal, 2025) View paper
    - Energy-Aware Scaling (1 papers)
    - [29] The Energy Cost of Reasoning: Analyzing Energy Usage in LLMs with Test-time Compute (Jin, 2025) View paper
    - Reasoning Chain Optimization (2 papers)
    - [20] Does Thinking More always Help? Mirage of Test-Time Scaling in Reasoning Models (Ghosal, 2025) View paper
    - [46] Don't Overthink it. Preferring Shorter Thinking Chains for Improved LLM Reasoning (Hassid, 2025) View paper
- Domain-Specific Applications
    - Medical and Scientific Reasoning (4 papers)
    - [10] Webthinker: Empowering large reasoning models with deep research capability (Li Xiaoxi, 2025) View paper
    - [11] m1: Unleash the Potential of Test-Time Scaling for Medical Reasoning with Large Language Models (Huang Xiaoke, 2025) View paper
    - [23] Aviary: training language agents on challenging scientific tasks (Narayanan, 2024) View paper
    - [26] Llama-3-meditron: An open-weight suite of medical llms based on llama-3.1 (A Sallinen, 2025) View paper
    - Coding and Mathematics (5 papers)
    - [7] Fuzzy-Assisted Contrastive Decoding Improving Code Generation of Large Language Models (Shuai Wang, 2025) View paper
    - [17] Reasoning-as-Logic-Units: Scaling Test-Time Reasoning in Large Language Models Through Logic Unit Alignment (Xu Tianyuan, 2025) View paper
    - [27] ScaleRTL: Scaling LLMs with Reasoning Data and Test-Time Compute for Accurate RTL Code Generation (Chenhui Deng, 2025) View paper
    - [41] M1: Towards scalable test-time compute with mamba reasoning models (Wang Junxiong, 2025) View paper
    - [44] MindStar: Enhancing Math Reasoning in Pre-trained LLMs at Inference Time (Kang, 2024) View paper
    - Legal Reasoning (1 papers)
    - [42] Evaluating Test-Time Scaling LLMs for Legal Reasoning: OpenAI o1, DeepSeek-R1, and Beyond (Ying-Hao Hu, 2025) View paper
    - Multimodal and Spatial Reasoning (3 papers)
    - [28] Reflect-DiT: Inference-Time Scaling for Text-to-Image Diffusion Transformers via In-Context Reflection (Li Shu-Fan, 2025) View paper
    - [45] MindJourney: Test-Time Scaling with World Models for Spatial Reasoning (Yang Yuncong, 2025) View paper
    - [50] Aurelia: Test-time Reasoning Distillation in Audio-Visual LLMs (Chowdhury, 2025) View paper
    - Knowledge-Intensive and Retrieval-Augmented Tasks (1 papers)
    - [37] Inference Scaled GraphRAG: Improving Multi Hop Question Answering on Knowledge Graphs (Thompson, 2025) View paper
- Theoretical Foundations and Analysis
    - Comparative Effectiveness Studies (1 papers)
    - [13] Think deep, think fast: Investigating efficiency of verifier-free inference-time-scaling methods (Wang Junlin, 2025) View paper
    - Scaling Laws and Optimization Principles (1 papers)
    - [47] s1: Simple test-time scaling (Niklas Muennighoff, 2025) View paper
    - Surveys and Taxonomies (3 papers)
    - [3] A survey of slow thinking-based reasoning llms using reinforced learning and inference-time scaling law (Ding, 2025) View paper
    - [33] A Survey of Test-Time Compute: From Intuitive Inference to Deliberate Reasoning (Ji Yixin, 2025) View paper
    - [34] A Survey on Test-Time Scaling in Large Language Models: What, How, Where, and How Well? (Zhang, 2025) View paper
- Auxiliary Techniques and Mechanisms
    - Activation and Representation Intervention (1 papers)
    - [2] Inference-time intervention: Eliciting truthful answers from a language model (Li Kenneth, 2023) View paper
    - Decoding Strategies and Output Calibration (2 papers)
    - [19] Deal: Decoding-time alignment for large language models (James Y. Huang, 2025) View paper
    - [39] Improving open-ended text generation via adaptive decoding (Zhu Wen-hong, 2024) View paper
    - Collaborative and Multi-Model Generation (1 papers)
    - [30] Learning to decode collaboratively with multiple language models (Lang, 2024) View paper
    - Security and Robustness (1 papers)
    - [48] SecInfer: Preventing Prompt Injection via Inference-time Scaling (Liu Yu-pei, 2025) View paper

Narrative

Core task: eliciting reasoning capabilities from base language models through inference-time sampling. The field has organized itself around several complementary branches. Inference-Time Sampling and Search Strategies explore pure sampling-based methods that generate multiple candidate solutions and select among them, often using techniques like majority voting or reward-guided selection. Training-Aware and Hybrid Approaches combine inference-time computation with model fine-tuning or reinforcement learning, bridging the gap between static model capabilities and dynamic reasoning. Efficiency and Acceleration Techniques address the computational cost of extended inference, proposing methods to reduce latency while preserving reasoning quality. Domain-Specific Applications tailor these strategies to specialized areas such as medicine, law, and web navigation, while Theoretical Foundations and Analysis provide formal understanding of scaling laws and optimality conditions. Auxiliary Techniques and Mechanisms encompass supporting tools like process reward models, critique mechanisms, and adaptive decoding strategies that enhance the core sampling paradigm.

Recent work has concentrated on understanding how test-time compute scales with performance, as surveyed in Slow Thinking Survey[3] and Test-Time Compute Survey[33], revealing trade-offs between sample diversity, verification accuracy, and computational budget. Within the pure sampling branch, Reasoning Without Training[0] emphasizes extracting reasoning purely at inference time without additional model updates, positioning itself alongside works like Reasoning with Sampling[14] and FIRE Sampling[18] that similarly rely on generating and filtering multiple reasoning paths. This contrasts with hybrid methods such as RL of Thoughts[4] or Inference-Aware Fine-Tuning[8], which interleave sampling with learning signals. A key open question is whether pure sampling can match the performance of training-augmented approaches when both are given comparable computational resources, and how to best allocate that budget across breadth of exploration versus depth of verification.

Related Works in Same Category

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

1. Reasoning with sampling: Your base model is smarter than you think

Authors: Karan, Aayush, Du, Yilun | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Frontier reasoning models have exhibited incredible capabilities across a wide array of disciplines, driven by posttraining large language models (LLMs) with reinforcement learning (RL). However, despite the widespread success of this paradigm, much of the literature has been devoted to disentangling truly novel behaviors that emerge during RL but are not present in the base models. In our work, we approach this question from a different angle, instead asking whether comparable reasoning capabil...

⚠ Similarity Notice

These papers appear to be the same work or very closely related variants. Both papers share nearly identical titles ('Reasoning without Training' vs 'Reasoning with Sampling'), identical abstracts describing MCMC-based sampling from power distributions to elicit reasoning from base models, and the same experimental setup (MATH500, HumanEval, GPQA benchmarks with Qwen2.5-Math-7B and other models). The core technical contribution—using Metropolis-Hastings to sample from p^α without training—is identical in both papers.

---

2. Flaming-hot Initiation with Regular Execution Sampling for Large Language Models

Authors: Chen, Weizhe, Dun Chen, Jin Xing, Liu, et al. (13 authors total) | **Year/Venue**: 2024 • North American Chapter of the Association for Computational Linguistics | **URL**: View paper

Abstract

Since the release of ChatGPT, large language models (LLMs) have demonstrated remarkable capabilities across various domains. A key challenge in developing these general capabilities is efficiently sourcing diverse, high-quality data. This becomes especially critical in reasoning-related tasks with sandbox checkers, such as math or code, where the goal is to generate correct solutions to specific problems with higher probability. In this work, we introduce Flaming-hot Initiation with Regular Exec...

Relationship Analysis

Both papers belong to the Pure Sampling-Based Methods category, focusing on improving reasoning through inference-time sampling without external verifiers. While the original paper uses MCMC-based iterative resampling from power distributions to sharpen the base model's output distribution, FIRE sampling employs a different strategy of high-temperature initialization followed by regular execution sampling to promote diversity and find high-quality responses. The key difference is that the original paper targets sampling from an explicitly sharpened distribution (p^α) using Metropolis-Hastings, whereas FIRE uses temperature variation across different positions within responses to balance exploration and exploitation.

---

3. Rollout Roulette: A Probabilistic Inference Approach to Inference-Time Scaling of LLMs using Particle-Based Monte Carlo Methods

Authors: Puri, Isha, Sudalairaj, Shivchander, Xu, et al. (9 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Large language models (LLMs) have achieved significant performance gains via scaling up model sizes and/or data. However, recent evidence suggests diminishing returns from such approaches, motivating scaling the computation spent at inference time. Existing inference-time scaling methods, usually with reward models, cast the task as a search problem, which tends to be vulnerable to reward hacking as a consequence of approximation errors in reward models. In this paper, we instead cast inference-...

Relationship Analysis

Both papers belong to the Pure Sampling-Based Methods category, focusing on eliciting reasoning capabilities from base models through inference-time sampling without external verifiers. While the original paper proposes an MCMC-based approach targeting power distributions (p^α) to sharpen the base model distribution, the candidate paper frames inference-time scaling as probabilistic inference using particle-based Monte Carlo methods with state-space models and approximate likelihoods. The key difference is that the original paper uses Metropolis-Hastings to sample from sharpened distributions of the base model itself, whereas the candidate employs particle filtering techniques to explore typical sets of state distributions rather than optimizing for modes.

Contributions Analysis

Overall novelty summary. The paper proposes an MCMC-inspired iterative sampling algorithm that uses base models' own likelihoods to sample from sharpened distributions, aiming to elicit reasoning capabilities without training. It resides in the 'Pure Sampling-Based Methods' leaf, which contains only four papers total, indicating a relatively sparse research direction within the broader taxonomy of 50 papers across 22 leaf nodes. This leaf explicitly excludes methods using verifiers or tree search, focusing instead on repeated sampling and likelihood-based selection—precisely the approach this work adopts.

The taxonomy reveals neighboring leaves with distinct strategies: 'Verification-Guided Sampling' employs external verifiers or reward models to select among candidates, while 'Structured Search and Tree-Based Exploration' uses systematic tree search methods. The paper's approach diverges from these by relying solely on the base model's likelihood without external verification or structured exploration. The broader 'Inference-Time Sampling and Search Strategies' branch contains 16 papers, suggesting moderate activity in inference-time methods overall, though the pure sampling subcategory remains less crowded than verification-guided or structured search alternatives.

Among 26 candidates examined, the contribution-level analysis shows mixed novelty signals. The power distribution sampling target (Contribution A) examined 6 candidates with 1 refutable match, suggesting some prior exploration of sharpened distributions. The MCMC-based algorithm (Contribution B) examined 10 candidates with none refutable, indicating stronger technical novelty in the specific algorithmic approach. The empirical claim of matching RL-posttraining (Contribution C) examined 10 candidates with 1 refutable, suggesting that demonstrating parity with training-based methods has been explored before, though perhaps not with this exact sampling technique.

Given the limited search scope of 26 candidates from semantic search, this assessment captures the most directly relevant prior work but cannot claim exhaustive coverage. The paper appears to occupy a moderately novel position within a sparse subcategory, with its core algorithmic contribution (MCMC-based power sampling) showing stronger novelty signals than its conceptual framing or empirical claims about matching RL performance.

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

Contribution 1: Power distribution as a sampling target for reasoning tasks

Description: The authors propose using the power distribution (p raised to power α) as an explicit target for sampling from base language models to enhance reasoning capabilities. This distribution sharpens the base model distribution by upweighting high-likelihood sequences without requiring any training.

This contribution was assessed against **6 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Automating High Energy Physics Data Analysis with LLM-Powered Agents

URL: View paper

Brief Assessment

High Energy Physics[74] focuses on automating HEP data analysis using LLM agents with workflow managers, not on power distribution sampling for reasoning tasks. The candidate does not address sampling distributions or reasoning enhancement techniques.

2. Enhancing Large Language Models with Graph-Based Node Sampling for Fault Attribution in Power Distribution Networks

URL: View paper

Brief Assessment

Graph-Based Node Sampling[70] focuses on graph-based node sampling for fault attribution in power distribution networks (electrical grids), not power distribution sampling targets for language model reasoning tasks. These are entirely different domains with no overlap.

3. Power-Flow: Unlocking LLMs with -Power Distribution Fine-Tuning

URL: View paper

Prior Art Analysis

Power-Flow[75] demonstrates that the concept of using power distribution (p raised to power α) as a sampling target for enhancing reasoning capabilities was already present in prior work. The candidate paper explicitly states that advanced skills like reasoning can be activated by sampling from the power distribution pbase(x)^α, which directly matches the original paper's core contribution. This shows that the original authors were not the first to propose using power distributions as an explicit target for sampling from base language models to enhance reasoning.

Evidence

Evidence 1 - **Rationale**: Both papers explicitly propose using the power distribution (p^α) as a sampling target to enhance reasoning capabilities in language models. The candidate paper demonstrates that this idea of sampling from power distributions to activate reasoning skills was already established in prior work, directly refuting the novelty claim that the original authors were first to introduce this concept. - **Original**: we introduce the power distribution as a useful sampling target for reasoning tasks. since it can be explicitly specified with a base llm, no additional training is required. - **Candidate**: we posit that many advanced skills, such as reasoning and creativity, are already latent within the base model and can be activated by sampling from its power distribution, pbase(x)î±.

Evidence 2 - **Rationale**: The original paper presents the mathematical formulation and reasoning for why power distributions sharpen base model distributions. The candidate paper's reference to the same power distribution formulation (pbase(x)^α) for activating reasoning capabilities shows this approach was already known, challenging the claim of being the first to propose this specific sampling target for reasoning tasks. - **Original**: one natural way to sharpen a distribution p is to sample from the power distribution pα. since p(x) > p(x′) =⇒ p(x)α p(x′)α > p(x) p(x′) (α ∈ [1, ∞]), it follows that exponentiating p increases the relative weight on higher likelihood sequences - **Candidate**: we posit that many advanced skills, such as reasoning and creativity, are already latent within the base model and can be activated by sampling from its power distribution, pbase(x)î±.

4. Build a Multimodal Interaction and Multi-Agent Collaborative Decision-Making Mechanism Enhanced by Large Models in the Intelligent Decision-Making System for â¦

URL: View paper

Brief Assessment

Multimodal Multi-Agent[73] focuses on multimodal interaction and multi-agent collaboration in power distribution systems, not on power distribution as a probability sampling target for language model reasoning.

5. Training a Reasoning Large Language Model for Improving Power Flow Convergence

URL: View paper

Brief Assessment

Power Flow Convergence[71] focuses on training language models for power system control tasks, not on using power distributions (p^α) as sampling targets for reasoning enhancement without training.

6. Uncertainty-Driven Adaptive Sampling for Resource-Efficient Language Model Inference

URL: View paper

Brief Assessment

Uncertainty-Driven Sampling[72] focuses on adaptive sampling based on uncertainty metrics for resource efficiency, not on power distribution (p^α) as an explicit sampling target for reasoning enhancement.

Contribution 2: MCMC-based power sampling algorithm for autoregressive models

Description: The authors develop a training-free sampling algorithm (Algorithm 1) that uses Metropolis-Hastings MCMC with random resampling proposals to approximately sample from the power distribution. The algorithm progressively samples from intermediate distributions in blocks to avoid exponential mixing time issues.

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. Eliciting language model behaviors using reverse language models

URL: View paper

Brief Assessment

Reverse Language Models[60] focuses on reverse language modeling for behavioral elicitation (finding prefixes that elicit specific suffixes), not on MCMC sampling from power distributions for reasoning tasks. The candidate uses Metropolis-Hastings for a different purpose (reversing LM dynamics) rather than sampling from p^α to enhance reasoning capabilities.

2. Sequential Monte Carlo Methods in the nimble and nimbleSMC R Packages

URL: View paper

Brief Assessment

Sequential Monte Carlo[67] focuses on particle filtering methods for state-space models with fixed or unknown parameters, not on MCMC-based power sampling for autoregressive language models. The candidate addresses sequential inference in time-series contexts using importance sampling and resampling, while the original contribution targets sampling from power distributions of language models using Metropolis-Hastings with random resampling proposals.

---

3. Designing proteins with language models

URL: View paper

Brief Assessment

Designing Proteins[64] focuses on protein sequence design using language models with MCMC for optimization of existing sequences, not on developing MCMC sampling algorithms for general autoregressive models to sample from power distributions.

---

4. Principled gradient-based MCMC for conditional sampling of text

URL: View paper

Brief Assessment

Gradient-Based MCMC[65] focuses on conditional sampling from energy-based text models using gradient information in continuous relaxation, while the original paper samples from power distributions of base LLMs for reasoning tasks without external constraints or training.

---

5. Bayesian estimation of an autoregressive model using Markov chain Monte Carlo

URL: View paper

Brief Assessment

Bayesian Autoregressive MCMC[69] focuses on Bayesian parameter estimation for autoregressive time series models, not sampling from power distributions for language model reasoning tasks.

---

6. Amortizing intractable inference in diffusion models for vision, language, and control

URL: View paper

Brief Assessment

Amortizing Intractable Inference[62] focuses on diffusion models for continuous and discrete spaces with trajectory balance objectives, not on MCMC power sampling for autoregressive language models as described in the original paper's Algorithm 1.

---

7. Posterior sampling via autoregressive generation

URL: View paper

Brief Assessment

Posterior Sampling Autoregressive[68] uses Metropolis-Hastings MCMC for bandit decision-making by sampling missing outcomes, not for sampling from power distributions of language models. The candidate focuses on imputing unobserved rewards in recommendation systems, while the original develops an algorithm to sample from p^α distributions for reasoning tasks.

---

8. Mix and match: Learning-free controllable text generation using energy language models

URL: View paper

Brief Assessment

Mix and Match[63] uses Metropolis-Hastings MCMC for energy-based controllable text generation with external discriminators, not for sampling from power distributions of base model likelihoods. The candidate's proposal distribution and acceptance criteria differ fundamentally from the original's power sampling approach.

---

9. Toward automated story generation with markov chain monte carlo methods and deep neural networks

URL: View paper

Brief Assessment

Automated Story Generation[66] uses Metropolis-Hastings MCMC for story generation with acceptance criteria based on neural networks, not for sampling from power distributions of autoregressive language models. The technical objectives and sampling targets are fundamentally different.

---

10. A Bayesian mixture model for Poisson network autoregression

URL: View paper

Brief Assessment

Bayesian Mixture Poisson[61] focuses on Bayesian mixture models for Poisson network autoregression with MCMC for parameter inference, not on power distribution sampling for language models. The MCMC algorithm targets posterior distributions of autoregressive coefficients in count time series, which is a fundamentally different application domain and technical approach.

Contribution 3: Empirical demonstration matching RL-posttraining performance without training

Description: The authors show that their training-free power sampling algorithm achieves single-shot reasoning performance comparable to or exceeding GRPO (a state-of-the-art RL method) across multiple base models and reasoning benchmarks, while maintaining better sample diversity and pass@k performance.

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. Reflexion: language agents with verbal reinforcement learning

URL: View paper

Brief Assessment

Reflexion[53] focuses on verbal reinforcement learning through linguistic feedback and episodic memory for language agents, not on training-free sampling algorithms that match RL performance through power distributions and MCMC methods.

---

2. Seg-zero: Reasoning-chain guided segmentation via cognitive reinforcement

URL: View paper

Brief Assessment

Seg-Zero[52] uses reinforcement learning (specifically GRPO) to train a segmentation model, whereas the original paper achieves comparable performance through pure sampling without any training. These are fundamentally different approaches to the same goal.

3. VEM: Environment-Free Exploration for Training GUI Agent with Value Environment Model

URL: View paper

Brief Assessment

VEM[56] focuses on GUI agent training with value models for environment-free RL, not training-free inference methods for reasoning tasks. The candidate requires pretraining a value environment model, whereas the original paper proposes pure sampling without any training.

4. Better zero-shot reasoning with self-adaptive prompting

URL: View paper

Brief Assessment

Self-Adaptive Prompting[57] focuses on improving zero-shot reasoning through prompt engineering and self-consistency mechanisms, not on matching RL-posttraining performance. The candidate addresses a different problem space (prompt design for reasoning) rather than demonstrating training-free methods that achieve RL-level performance on reasoning benchmarks.

5. ThinkLess: A Training-Free Inference-Efficient Method for Reducing Reasoning Redundancy

URL: View paper

Brief Assessment

ThinkLess[54] focuses on reducing reasoning token redundancy in already-trained models through early termination, not on matching RL performance through inference-time sampling algorithms. The original paper's power sampling algorithm targets sampling from sharpened distributions to match RL capabilities, while ThinkLess addresses computational efficiency by truncating reasoning sequences.

6. Training-free Generation of Temporally Consistent Rewards from VLMs

URL: View paper

Brief Assessment

Training-Free Rewards[59] focuses on VLM-based reward generation for robot manipulation tasks, not on training-free sampling methods for reasoning benchmarks. The candidate addresses reward signal generation from vision-language models, while the original paper addresses inference-time sampling algorithms for language model reasoning.

7. Seg-R1: Segmentation Can Be Surprisingly Simple with Reinforcement Learning

URL: View paper

Brief Assessment

Seg-R1[58] uses RL training (GRPO) for segmentation tasks, not a training-free method. The original paper's contribution is about achieving RL-level performance without any training through pure sampling algorithms.

8. RFG: Test-Time Scaling for Diffusion Large Language Model Reasoning with Reward-Free Guidance

URL: View paper

Brief Assessment

RFG[55] focuses on diffusion large language models (DLLMs) with any-order generation, while the original paper addresses autoregressive base models. RFG[55] requires an enhanced model (RL or SFT) as input, whereas the original claims pure sampling from base models alone.

9. Reasoning with sampling: Your base model is smarter than you think

URL: View paper

Prior Art Analysis

Reasoning with Sampling[14] demonstrates that their training-free power sampling algorithm achieves single-shot reasoning performance comparable to or exceeding GRPO (a state-of-the-art RL method) across multiple base models and reasoning benchmarks. The candidate paper presents nearly identical claims, methodology, and experimental results, including the same benchmarks (math500, humaneval, gpqa), the same base models (qwen2.5-math-7b, qwen2.5-7b, phi-3.5-mini-instruct), and the same RL baseline (GRPO). The candidate paper's abstract and results sections contain text that is nearly word-for-word identical to the original paper, demonstrating that this work was published prior to the original submission.

Evidence

Evidence 1 - **Rationale**: These quotes are identical, demonstrating that Reasoning with Sampling[14] published the exact same claim about matching RL-posttraining performance without training before the original paper's submission. - **Original**: we show that our algorithm offers substantial boosts in reasoning that nearly match and even outperform those from rl on a wide variety of single-shot tasks, including math500, humaneval, and gpqa. moreover, our sampler avoids the collapse in diversity over multiple samples that is characteristic of... - **Candidate**: we show that our algorithm offers substantial boosts in reasoning that nearly match and even outperform those from rl on a wide variety of single-shot tasks, including math500, humaneval, and gpqa. moreover, our sampler avoids the collapse in diversity over multiple samples that is characteristic of...

Evidence 2 - **Rationale**: The figure captions are nearly identical, showing that Reasoning with Sampling[14] presented the same empirical demonstration of matching and outperforming RL-posttraining across the same benchmarks. - **Original**: our sampling algorithm can match and outperform rl-posttraining.left: we compare our sampling algorithm (ours) against the base model (base) and rl-posttraining (grpo) on threeverifiable reasoning tasks (math500, humaneval, gpqa). right: we compare them on an unverifiable general task (alpacaeval2.0... - **Candidate**: our sampling algorithm can match and outperform rl-posttraining.left: we compare our sampling algorithm (ours) against the base model (base) and rl-posttraining (grpo) on threeverifiable reasoning tasks (math500, humaneval, gpqa). right: we compare them on anunverifiable general task(alpacaeval2.0)....

Evidence 3 - **Rationale**: The results sections contain identical text describing the same empirical findings with the same specific performance numbers, demonstrating that Reasoning with Sampling[14] published these results first. - **Original**: across base models of different families, our sampling algorithm achieves massive, near-universal boosts in single-shot accuracies and scores over different reasoning and evaluation tasks that reach, e.g., up to +51.9% on humaneval with phi-3.5-mini and +25.2% on math500 with qwen2.5-math. in partic... - **Candidate**: across base models of different families, our sampling algorithm achieves massive, near-universal boosts in single-shot accuracies and scores over different reasoning and evaluation tasks that reach, e.g., up to+51.9%on humaneval with phi-3.5-mini and+25.2%on math500 with qwen2.5-math. in particular...

Evidence 4 - **Rationale**: Both papers use identical language to describe their training-free approach and comparison against GRPO, showing that Reasoning with Sampling[14] established this contribution prior to the original paper. - **Original**: we benchmark specifically against group relative policy optimization (grpo), which is the standard rl algorithm for enhancing llm reasoning (shao et al., 2024).

crucially, our algorithm is training-free, dataset-free, and verifier-free, avoiding some of the inherent weaknesses of rl methods - **Candidate**: we benchmark specifically against group relative policy optimization (grpo), which is the standard rl algorithm for enhancing llm reasoning [shao et al., 2024]. crucially, our algorithm istraining-free,dataset-free, andverifier-free, avoiding some of the inherent weaknesses of rl methods

10. Absolute zero: Reinforced self-play reasoning with zero data

URL: View paper

Brief Assessment

Absolute Zero[51] uses reinforcement learning with self-play on self-proposed tasks, not a training-free inference-time sampling method. The original paper achieves comparable performance through pure sampling at inference without any training, while Absolute Zero[51] explicitly trains models using RL.

Appendix: Text Similarity Detection

Textual similarity detection checked 28 papers and found 3 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. Reasoning with sampling: Your base model is smarter than you think

Detected in: Core Task (sibling), 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 without Training: Your Base Model is Smarter Than You Think View paper
- [1] Scaling LLM test-time compute optimally can be more effective than scaling parameters for reasoning View paper
- [2] Inference-time intervention: Eliciting truthful answers from a language model View paper
- [3] A survey of slow thinking-based reasoning llms using reinforced learning and inference-time scaling law View paper
- [4] Rl of thoughts: Navigating llm reasoning with inference-time reinforcement learning View paper
- [5] Sample, scrutinize and scale: Effective inference-time search by scaling verification View paper
- [6] Towards large reasoning models: A survey of reinforced reasoning with large language models View paper
- [7] Fuzzy-Assisted Contrastive Decoding Improving Code Generation of Large Language Models View paper
- [8] Inference-aware fine-tuning for best-of-n sampling in large language models View paper
- [9] Specreason: Fast and accurate inference-time compute via speculative reasoning View paper
- [10] Webthinker: Empowering large reasoning models with deep research capability View paper
- [11] m1: Unleash the Potential of Test-Time Scaling for Medical Reasoning with Large Language Models View paper
- [12] GenPRM: Scaling Test-Time Compute of Process Reward Models via Generative Reasoning View paper
- [13] Think deep, think fast: Investigating efficiency of verifier-free inference-time-scaling methods View paper
- [14] Reasoning with sampling: Your base model is smarter than you think View paper
- [15] The Surprising Effectiveness of Test-Time Training for Abstract Reasoning View paper
- [16] Teaching Large Language Models to Reason through Learning and Forgetting View paper
- [17] Reasoning-as-Logic-Units: Scaling Test-Time Reasoning in Large Language Models Through Logic Unit Alignment View paper
- [18] Flaming-hot Initiation with Regular Execution Sampling for Large Language Models View paper
- [19] Deal: Decoding-time alignment for large language models View paper
- [20] Does Thinking More always Help? Mirage of Test-Time Scaling in Reasoning Models View paper
- [21] Rollout Roulette: A Probabilistic Inference Approach to Inference-Time Scaling of LLMs using Particle-Based Monte Carlo Methods View paper
- [22] MedAdapter: Efficient Test-Time Adaptation of Large Language Models Towards Medical Reasoning View paper
- [23] Aviary: training language agents on challenging scientific tasks View paper
- [24] Enhancing LLM Reasoning via Critique Models with Test-Time and Training-Time Supervision View paper
- [25] Think twice: Enhancing llm reasoning by scaling multi-round test-time thinking View paper
- [26] Llama-3-meditron: An open-weight suite of medical llms based on llama-3.1 View paper
- [27] ScaleRTL: Scaling LLMs with Reasoning Data and Test-Time Compute for Accurate RTL Code Generation View paper
- [28] Reflect-DiT: Inference-Time Scaling for Text-to-Image Diffusion Transformers via In-Context Reflection View paper
- [29] The Energy Cost of Reasoning: Analyzing Energy Usage in LLMs with Test-time Compute View paper
- [30] Learning to decode collaboratively with multiple language models View paper
- [31] Self-rewarding correction for mathematical reasoning View paper
- [32] Parathinker: Native parallel thinking as a new paradigm to scale llm test-time compute View paper
- [33] A Survey of Test-Time Compute: From Intuitive Inference to Deliberate Reasoning View paper
- [34] A Survey on Test-Time Scaling in Large Language Models: What, How, Where, and How Well? View paper
- [35] The Surprising Effectiveness of Test-Time Training for Few-Shot Learning View paper
- [36] A Survey of Reinforcement Learning in Large Language Models: From Data Generation to Test-Time Inference View paper
- [37] Inference Scaled GraphRAG: Improving Multi Hop Question Answering on Knowledge Graphs View paper
- [38] Forest-of-Thought: Scaling Test-Time Compute for Enhancing LLM Reasoning View paper
- [39] Improving open-ended text generation via adaptive decoding View paper
- [40] Wider or Deeper? Scaling LLM Inference-Time Compute with Adaptive Branching Tree Search View paper
- [41] M1: Towards scalable test-time compute with mamba reasoning models View paper
- [42] Evaluating Test-Time Scaling LLMs for Legal Reasoning: OpenAI o1, DeepSeek-R1, and Beyond View paper
- [43] Accelerated Test-Time Scaling with Model-Free Speculative Sampling View paper
- [44] MindStar: Enhancing Math Reasoning in Pre-trained LLMs at Inference Time View paper
- [45] MindJourney: Test-Time Scaling with World Models for Spatial Reasoning View paper
- [46] Don't Overthink it. Preferring Shorter Thinking Chains for Improved LLM Reasoning View paper
- [47] s1: Simple test-time scaling View paper

- [48] SecInfer: Preventing Prompt Injection via Inference-time Scaling View paper
- [49] Probabilistic Optimality for Inference-time Scaling View paper
- [50] Aurelia: Test-time Reasoning Distillation in Audio-Visual LLMs View paper
- [51] Absolute zero: Reinforced self-play reasoning with zero data View paper
- [52] Seg-zero: Reasoning-chain guided segmentation via cognitive reinforcement View paper
- [53] Reflexion: language agents with verbal reinforcement learning View paper
- [54] ThinkLess: A Training-Free Inference-Efficient Method for Reducing Reasoning Redundancy View paper
- [55] RFG: Test-Time Scaling for Diffusion Large Language Model Reasoning with Reward-Free Guidance View paper
- [56] VEM: Environment-Free Exploration for Training GUI Agent with Value Environment Model View paper
- [57] Better zero-shot reasoning with self-adaptive prompting View paper
- [58] Seg-R1: Segmentation Can Be Surprisingly Simple with Reinforcement Learning View paper
- [59] Training-free Generation of Temporally Consistent Rewards from VLMs View paper
- [60] Eliciting language model behaviors using reverse language models View paper
- [61] A Bayesian mixture model for Poisson network autoregression View paper
- [62] Amortizing intractable inference in diffusion models for vision, language, and control View paper
- [63] Mix and match: Learning-free controllable text generation using energy language models View paper
- [64] Designing proteins with language models View paper
- [65] Principled gradient-based MCMC for conditional sampling of text View paper
- [66] Toward automated story generation with markov chain monte carlo methods and deep neural networks View paper
- [67] Sequential Monte Carlo Methods in the nimble and nimbleSMC R Packages View paper
- [68] Posterior sampling via autoregressive generation View paper
- [69] Bayesian estimation of an autoregressive model using Markov chain Monte Carlo View paper
- [70] Enhancing Large Language Models with Graph-Based Node Sampling for Fault Attribution in Power Distribution Networks View paper
- [71] Training a Reasoning Large Language Model for Improving Power Flow Convergence View paper
- [72] Uncertainty-Driven Adaptive Sampling for Resource-Efficient Language Model Inference View paper
- [73] Build a Multimodal Interaction and Multi-Agent Collaborative Decision-Making Mechanism Enhanced by Large Models in the Intelligent Decision-Making System for â¦ View paper
- [74] Automating High Energy Physics Data Analysis with LLM-Powered Agents View paper
- [75] Power-Flow: Unlocking LLMs with -Power Distribution Fine-Tuning View paper