Novelty Assessment Report

Paper: DiffuCoder: Understanding and Improving Masked Diffusion Models for Code Generation
PDF URL: https://openreview.net/pdf?id=58NA3unZj5
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-07

Abstract

Diffusion large language models (dLLMs) are compelling alternatives to autoregressive (AR) models because their denoising models operate over the entire sequence. The global planning and iterative refinement features of dLLMs are particularly useful for code generation. However, current training and inference mechanisms for dLLMs in coding are still under-explored. To demystify the decoding behavior of dLLMs and unlock their potential for coding, we systematically investigate their denoising processes and reinforcement learning (RL) methods. We train a 7B dLLM, DiffuCoder, on 130B tokens of code. Using this model as a testbed, we analyze its decoding behavior, revealing how it differs from that of AR models: (1) dLLMs can decide how causal their generation should be without relying on semi-AR decoding, and (2) increasing the sampling temperature diversifies not only token choices but also their generation order. This diversity creates a rich search space for RL rollouts. For RL training, to reduce the variance of token log-likelihood estimates and maintain training efficiency, we propose coupled-GRPO, a novel sampling scheme that constructs complementary mask noise for completions used in training. In our experiments, coupled-GRPO significantly improves DiffuCoder's performance on code generation benchmarks (+4.4\% on EvalPlus) and reduces reliance on AR bias during decoding. Our work provides deeper insight into the machinery of dLLM generation and offers an effective, diffusion-native RL training framework.

Core Task Landscape

This paper addresses: **Masked Diffusion Models for Code Generation**

A total of **16 papers** were analyzed and organized into a taxonomy with **10 categories**.

Taxonomy Overview

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

- **Core Diffusion Architectures and Training Frameworks**
- **Inference and Sampling Strategies**
- **Reinforcement Learning and Optimization for Diffusion Models**
- **Theoretical Analysis and Comparative Studies**

Complete Taxonomy Tree

- Masked Diffusion Models for Code Generation Survey Taxonomy
- Core Diffusion Architectures and Training Frameworks
  - Augmented State Space Diffusion Models (1 papers)
  - [1] Continuously augmented discrete diffusion model for categorical generative modeling (Zheng, 2025) View paper
  - Structure-Aware Code Diffusion Frameworks (2 papers)
  - [2] TreeDiff: AST-Guided Code Generation with Diffusion LLMs (Zeng, 2025) View paper
  - [4] CodeDiffuSe: A masked diffusion framework for structure-aware code completion and repair (Aytug Onan, 2025) View paper
  - Pre-trained Diffusion Language Models (2 papers)
  - [6] CodeFusion: A Pre-trained Diffusion Model for Code Generation (Mukul Singh, 2023) View paper
  - [13] LLaDA-MoE: A Sparse MoE Diffusion Language Model (Zhu Feng-qi, 2025) View paper
- Inference and Sampling Strategies
  - Lookahead and Path Planning Sampling (3 papers)
  - [3] Lookahead Unmasking Elicits Accurate Decoding in Diffusion Language Models (Sanghyun Lee, 2025) View paper
  - [7] Path Planning for Masked Diffusion Models with Applications to Biological Sequence Generation (FZ Peng, 2025) View paper
  - [10] Path Planning for Masked Diffusion Model Sampling (Bezemek, 2025) View paper
  - Soft-Masking and Probabilistic Unmasking (1 papers)
  - [8] Soft-Masked Diffusion Language Models (Hersche, 2025) View paper
  - Scheduled and Dilated Unmasking Strategies (2 papers)
  - [11] Plan for Speed--Dilated Scheduling for Masked Diffusion Language Models (Permuter, 2025) View paper
  - [16] Guided Star-Shaped Masked Diffusion (Meshchaninov, 2025) View paper
- Reinforcement Learning and Optimization for Diffusion Models
  - Latent Policy Adaptation and Reward-Guided Decoding (1 papers)
  - [5] Latent Adaptation with Masked Policy for Diffusion Language Models (G Sun, n.d.) View paper
  - Trajectory-Level Reinforcement Learning ★ (2 papers)
  - [0] DiffuCoder: Understanding and Improving Masked Diffusion Models for Code Generation (Anon et al., 2026) View paper
  - [9] Reinforcing the Diffusion Chain of Lateral Thought with Diffusion Language Models (Huang Ze-min, 2025) View paper
  - Distillation and Acceleration via Reinforcement Learning (1 papers)
  - [14] dUltra: Ultra-Fast Diffusion Language Models via Reinforcement Learning (Shirui Chen, 2025) View paper

- Theoretical Analysis and Comparative Studies (2 papers)
  - [12] On Powerful Ways to Generate: Autoregression, Diffusion, and Beyond (Yang Chenxiao, 2025) View paper
  - [15] Diffusion vs Autoregression: An Empirical Study on Code Comment Translation (A Dikov, 2025) View paper

Narrative

Core task: masked diffusion models for code generation. The field organizes around four main branches that reflect different aspects of applying diffusion techniques to discrete code synthesis. Core Diffusion Architectures and Training Frameworks establish foundational masking and denoising mechanisms, often exploring how to adapt continuous diffusion principles to token-level generation (e.g., CodeFusion[6], Soft-Masked Diffusion[8]). Inference and Sampling Strategies address how to efficiently decode from learned diffusion models, including scheduling variants like Dilated Scheduling[11] and lookahead techniques such as Lookahead Unmasking[3]. Reinforcement Learning and Optimization for Diffusion Models investigates trajectory-level or policy-based refinements to improve sample quality and task-specific performance. Finally, Theoretical Analysis and Comparative Studies examine trade-offs between diffusion and autoregressive paradigms, as seen in works like Diffusion vs Autoregression[15], providing empirical and conceptual grounding for design choices.

Within the reinforcement learning branch, a small cluster of works explores trajectory-level optimization to guide diffusion sampling toward higher-quality outputs. DiffuCoder[0] sits squarely in this area, emphasizing RL-driven refinement of masked diffusion trajectories for code generation tasks. It shares thematic overlap with Lateral Thought Diffusion[9], which similarly leverages trajectory-level reasoning, though the latter may focus on broader sequential decision-making contexts. Meanwhile, neighboring efforts like Latent Adaptation Masked Policy[5] investigate policy adaptation in latent spaces, highlighting an ongoing tension between end-to-end RL tuning and modular latent interventions. These contrasting approaches reflect open questions about where and how to inject optimization signals—whether at the token unmasking level, across entire generation rollouts, or within learned latent representations—underscoring the evolving interplay between diffusion mechanics and reinforcement learning in discrete generation domains.

Related Works in Same Category

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

1. Reinforcing the Diffusion Chain of Lateral Thought with Diffusion Language Models

Authors: Huang Ze-min, Chen Zhi-yang, Zemin Huang, Wang, Zijun, et al. (11 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

We introduce the Diffusion Chain of Lateral Thought (DCoLT), a reasoning framework for diffusion language models. DCoLT treats each intermediate step in the reverse diffusion process as a latent"thinking"action and optimizes the entire reasoning trajectory to maximize the reward on the correctness of the final answer with outcome-based Reinforcement Learning (RL). Unlike traditional Chain-of-Thought (CoT) methods that follow a causal, linear thinking process, DCoLT allows bidirectional, non-line...

Relationship Analysis

Both papers belong to the Trajectory-Level Reinforcement Learning category, applying RL methods to optimize entire diffusion trajectories using outcome-based rewards. They overlap in using GRPO-style algorithms to train diffusion language models for code generation tasks, with both focusing on reinforcing complete generation sequences rather than individual steps. The key difference is that the original paper (DiffuCoder) focuses on understanding and improving masked diffusion models specifically for code generation with a novel coupled-GRPO sampling scheme, while the candidate paper introduces DCoLT (Diffusion Chain of Lateral Thought) as a general reasoning framework that treats intermediate diffusion steps as lateral thinking actions and applies it to both math and code tasks with an unmasking policy module for discrete-time models.

Contributions Analysis

Overall novelty summary. The paper introduces DiffuCoder, a 7B diffusion language model trained on 130B code tokens, and proposes coupled-GRPO, a reinforcement learning algorithm tailored for diffusion-based code generation. According to the taxonomy, this work resides in the 'Trajectory-Level Reinforcement Learning' leaf under the broader 'Reinforcement Learning and Optimization for Diffusion Models' branch. This leaf contains only two papers total, including the original work, indicating a relatively sparse research direction within the masked diffusion for code generation landscape.

The taxonomy reveals that neighboring leaves explore alternative optimization strategies: 'Latent Policy Adaptation and Reward-Guided Decoding' focuses on external reward models guiding decoding, while 'Distillation and Acceleration via Reinforcement Learning' emphasizes efficiency through distillation. The sibling paper in the same leaf, Lateral Thought Diffusion, shares the trajectory-level optimization theme but may target broader sequential reasoning contexts. Meanwhile, the 'Core Diffusion Architectures' and 'Inference and Sampling Strategies' branches address orthogonal concerns—foundational training mechanisms and decoding algorithms—suggesting DiffuCoder's RL contributions occupy a distinct methodological niche.

Among 30 candidates examined, the DiffuCoder model contribution shows one refutable candidate out of ten examined, suggesting some prior work on large-scale diffusion models for code exists. The local/global AR-ness metrics contribution found no refutable candidates among ten examined, indicating potential novelty in analyzing diffusion decoding behavior. The coupled-GRPO algorithm shows two refutable candidates out of ten, implying moderate overlap with existing RL methods for diffusion models. These statistics reflect a limited semantic search scope, not exhaustive coverage of all relevant literature.

Based on the top-30 semantic matches examined, the work appears to occupy a moderately explored intersection of diffusion models and reinforcement learning for code generation. The trajectory-level RL focus sits in a sparse taxonomy leaf, though the broader RL-for-diffusion branch contains related efforts. The analysis does not cover potential work outside the semantic search radius or recent preprints, leaving open questions about comprehensiveness in rapidly evolving diffusion model research.

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

Contribution 1: DiffuCoder: 7B diffusion model for code generation

Description: The authors train DiffuCoder, a 7-billion parameter masked diffusion language model specialized for code generation, trained on 130B tokens. This model serves as a testbed for analyzing diffusion model behavior and developing new training methods.

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. Directional Diffusion-Style Code Editing Pre-training

URL: View paper

Brief Assessment

Directional Diffusion Code Editing[35] focuses on code editing tasks using a directional diffusion-style approach at the data level, not on training a 7B masked diffusion language model for general code generation like DiffuCoder.

---

2. Seed diffusion: A large-scale diffusion language model with high-speed inference

URL: View paper

Brief Assessment

Seed Diffusion[30] focuses on high-speed inference optimization for diffusion language models in code generation, not on training a 7B diffusion model as a testbed for analyzing diffusion behavior and developing new training methods like DiffuCoder.

---

3. dKV-Cache: The Cache for Diffusion Language Models

URL: View paper

Brief Assessment

dKV-Cache[32] focuses on inference acceleration mechanisms (KV-cache) for diffusion language models, not on training a specialized code generation model. The candidate does not demonstrate prior work on training diffusion models specifically for code generation tasks.

---

4. Dream-coder 7b: An open diffusion language model for code

URL: View paper

Brief Assessment

Dream-coder 7b[28] is also a 7B diffusion model for code generation trained on similar scale data, but focuses on different aspects (adaptive generation patterns, post-training techniques) rather than challenging the novelty of training a 7B diffusion model for code.

---

5. Beyond autoregression: An empirical study of diffusion large language models for code generation

URL: View paper

Brief Assessment

Beyond Autoregression Code[27] is an empirical study evaluating existing diffusion models including DiffuCoder, not claiming to develop DiffuCoder itself. The paper states 'we present the first empirical study of diffusion llms for code generation' and evaluates 9 models including diffucoder-7b-cpgrpo as one subject among others.

---

6. Diffusion-based Large Language Models Survey

URL: View paper

Brief Assessment

Diffusion LLMs Survey[34] only briefly mentions diffusion models for code generation in passing (e.g., 'language model focused on code generation') without presenting a specific 7B-scale model trained on 130B tokens or the detailed training methodology and analysis that DiffuCoder provides.

---

7. Dream 7b: Diffusion large language models

URL: View paper

Brief Assessment

Dream 7b[29] is a general-purpose diffusion language model evaluated across multiple domains (text, math, code), not specifically a code generation model. While it includes code benchmarks, it does not claim to be the first 7B diffusion model for code generation.

---

8. CodeFusion: A Pre-trained Diffusion Model for Code Generation

URL: View paper

Prior Art Analysis

CodeFusion[6] demonstrates that a pre-trained diffusion model for code generation was already developed and published prior to DiffuCoder. While CodeFusion[6] is smaller (75M parameters vs. 7B), it establishes the core concept of applying diffusion models to code generation tasks, including training on code data and evaluating on code generation benchmarks. Both models use masked diffusion approaches for iterative refinement of code sequences, challenging the novelty claim that DiffuCoder is the first diffusion model specialized for code generation.

Evidence

Evidence 1 - **Rationale**: Both papers describe training diffusion models specifically for code generation. CodeFusion[6] explicitly states it is a 'pre-trained diffusion code generation model,' establishing prior work in this exact domain before DiffuCoder. - **Original**: we train a 7b dllm, diffucoder, on 130b tokens of code - **Candidate**: We introduce codefusion, a pre-trained diffusion code generation model that addresses this limitation by iteratively denoising a complete program conditioned on the encoded natural language.

Evidence 2 - **Rationale**: Both papers evaluate diffusion models on code generation tasks. CodeFusion[6] was evaluated on multiple programming languages (Bash, Python, Excel), demonstrating that diffusion models for code generation existed and were empirically validated before DiffuCoder. - **Original**: To address these limitations, we first gain insight into decoding behaviors of dllms and then establish a diffusion-native reinforcement learning (rl) methodology. our investigation is grounded in the analysis of diffucoder, a 7b-scale mdm specialized for code generation (§3), trained on 130b effect... - **Candidate**: We evaluate codefusion on the task of natural language to code generation for bash, python, and microsoft excel conditional formatting (cf) rules.

Evidence 3 - **Rationale**: Both papers identify the same core motivation: diffusion models' iterative refinement capability is advantageous for code generation compared to autoregressive models. CodeFusion[6] explicitly articulates this advantage, showing prior recognition of diffusion models' benefits for code tasks. - **Original**: diffusion large language models (dllms) are compelling alternatives to autoregressive (ar) models because their denoising models operate over the entire sequence. the global planning and iterative refinement features of dllms are particularly useful for code generation. - **Candidate**: auto-regressive models for code generation from natural language have a similar limitation: they do not easily allow reconsidering earlier tokens generated. We introduce codefusion, a pre-trained diffusion code generation model that addresses this limitation by iteratively denoising a complete progr...

---

9. Mercury: Ultra-fast language models based on diffusion

URL: View paper

Brief Assessment

Mercury[31] focuses on commercial-scale diffusion models optimized for speed (1100+ tokens/sec) rather than analyzing diffusion model behavior or developing new training methods like DiffuCoder's coupled-GRPO.

---

10. DDPT: Diffusion-Driven Prompt Tuning for Large Language Model Code Generation

URL: View paper

Brief Assessment

DDPT[33] focuses on prompt optimization using diffusion models to generate optimal prompt embeddings for code generation, not on training a large-scale diffusion language model for code generation itself. The candidate addresses a different problem (prompt engineering automation) rather than model architecture development.

Contribution 2: Local and global AR-ness metrics for analyzing diffusion decoding

Description: The authors propose two metrics to quantify how closely diffusion models follow autoregressive (left-to-right) generation patterns. These metrics reveal that diffusion models can adaptively decide their generation order and that higher sampling temperatures increase non-autoregressive behavior.

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. An efficient diffusion-based non-autoregressive solver for traveling salesman problem

URL: View paper

Brief Assessment

Diffusion TSP Solver[44] focuses on solving the traveling salesman problem using diffusion models for combinatorial optimization, not on analyzing autoregressive versus non-autoregressive generation patterns in language models or measuring AR-ness metrics.

---

2. Ditar: Diffusion transformer autoregressive modeling for speech generation

URL: View paper

Brief Assessment

Ditar[41] focuses on speech generation using diffusion transformers combined with autoregressive modeling for patch-based generation. It does not propose metrics to measure autoregressive versus non-autoregressive generation patterns in diffusion models for code or text generation.

---

3. Ar-diffusion: Auto-regressive diffusion model for text generation

URL: View paper

Brief Assessment

Ar-diffusion Text[43] focuses on designing an auto-regressive diffusion model architecture with dynamic movement speeds, not on measuring AR-ness patterns in existing diffusion models. The paper does not propose metrics to quantify autoregressive versus non-autoregressive generation patterns.

---

4. Thermalizer: Stable autoregressive neural emulation of spatiotemporal chaos

URL: View paper

Brief Assessment

Thermalizer[45] focuses on stabilizing autoregressive surrogate models of spatiotemporal physical systems using diffusion models for denoising, not on measuring or analyzing autoregressive versus non-autoregressive generation patterns in language diffusion models.

---

5. ViD-GPT: Introducing GPT-style Autoregressive Generation in Video Diffusion Models

URL: View paper

Brief Assessment

ViD-GPT[38] focuses on video diffusion models with causal temporal attention for long video generation, not on measuring autoregressive versus non-autoregressive generation patterns in general diffusion models or analyzing how temperature affects generation order.

---

6. Block Diffusion: Interpolating Between Autoregressive and Diffusion Language Models

URL: View paper

Brief Assessment

Block Diffusion[39] focuses on a hybrid architecture that interpolates between autoregressive and diffusion models with block-based generation, rather than proposing metrics to measure autoregressive-ness in standard diffusion models. The candidate does not present measurement methodologies for quantifying AR patterns in diffusion decoding.

---

7. From Slow Bidirectional to Fast Autoregressive Video Diffusion Models

URL: View paper

Brief Assessment

Fast Autoregressive Video[37] focuses on converting bidirectional video diffusion models to autoregressive transformers for streaming generation. It does not propose metrics to measure autoregressive versus non-autoregressive generation patterns in diffusion models, which is the core novelty of the original paper's AR-ness metrics.

---

8. Ar-diffusion: Asynchronous video generation with auto-regressive diffusion

URL: View paper

Brief Assessment

Ar-diffusion Video[40] focuses on video generation with frame-level timestep constraints, not on measuring autoregressive patterns in text/code diffusion models. The candidate's timestep compositions relate to video frame noise levels, not sequential text generation analysis.

---

9. Progressive Autoregressive Video Diffusion Models

URL: View paper

Brief Assessment

Progressive Autoregressive Video[36] focuses on video generation with progressive noise schedules for autoregressive long video synthesis, not on measuring or analyzing autoregressive versus non-autoregressive generation patterns in diffusion models for text or code.

---

10. Amd: Autoregressive motion diffusion

URL: View paper

Brief Assessment

Autoregressive Motion Diffusion[42] focuses on human motion generation from text/audio, not on analyzing autoregressive patterns in diffusion language models for code generation.

Contribution 3: Coupled-GRPO: diffusion-native reinforcement learning algorithm

Description: The authors develop coupled-GRPO, a reinforcement learning method tailored for diffusion models that uses complementary mask noise pairs to reduce variance in token likelihood estimation while maintaining training efficiency. This method respects the non-autoregressive nature of diffusion models and significantly improves 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. Principled and Tractable RL for Reasoning with Diffusion Language Models

URL: View paper

Prior Art Analysis

Tractable RL Reasoning[18] demonstrates that prior work exists on principled RL algorithms for diffusion language models. The candidate paper presents AGRPO (Amortized Group Relative Policy Optimization), which addresses the same core challenge as Coupled-GRPO: computing unbiased policy gradients for diffusion models while maintaining computational tractability. Both methods aim to solve the fundamental problem that standard GRPO requires O(response length) forward passes for diffusion models, making it intractable. While the technical approaches differ (Coupled-GRPO uses complementary mask pairs; AGRPO uses Monte Carlo sampling over timesteps), both claim to be the first principled, tractable RL algorithm for diffusion LLMs, directly challenging the novelty claim.

Evidence

Evidence 1 - **Rationale**: Both papers claim to present novel RL algorithms specifically designed for diffusion LLMs. The candidate explicitly claims to be 'the first tractable, faithful adaptation of policy gradient methods for dllms', which directly contradicts the original paper's novelty claim for Coupled-GRPO. - **Original**: we propose coupledgrpo, a novel sampling scheme that constructs complementary mask noise for completions used in training. in our experiments, coupled-grpo significantly improves diffucoder's performance on code generation benchmarks (+4.4% on evalplus) and reduces reliance on ar bias during decodin... - **Candidate**: in this work, we present amortized group relative policy optimization (agrpo), a principled on-policy rl algorithm designed specifically for dllms. agrpo uses monte carlo sampling to compute an unbiased policy gradient estimate, making it the first tractable, faithful adaptation of policy gradient m...

Evidence 2 - **Rationale**: Both papers address the same fundamental challenge of computing accurate policy gradients for diffusion models. The candidate's claim to be the 'first principled adaptation' challenges the original's novelty, as both propose methods to achieve unbiased estimates while maintaining efficiency. - **Original**: for rl training, to reduce the variance of token log-likelihood estimates and maintain training efficiency, we propose coupledgrpo, a novel sampling scheme that constructs complementary mask noise for completions used in training. - **Candidate**: agrpo is the firstprincipledadaptation of policy gradient methods to dllms that computes anunbiasedpolicy gradient estimate by dropping unnecessary approximations in favor of exact token probabilities.

Evidence 3 - **Rationale**: Both papers propose novel sampling schemes to make GRPO tractable for diffusion models. While the technical approaches differ, both claim to provide theoretically grounded solutions to the same fundamental problem, challenging the original's claim to be first. - **Original**: we design coupled-grpo, an rl algorithm for dllms that avoids semi-ar decoding by using a novel coupled-sampling scheme for efficient and accurate policy gradient estimation (§5). we theoretically prove the variance reduction of coupled-grpo using antithetic variates. - **Candidate**: we show how to reinterpret the grpo objective as an expectation across timesteps with respect to the uniform measure... The rhs can now be estimated via monte carlo (mc) sampling by drawingk≪mtimesteps from the uniform distribution on{1, . . . , m}, computing theexactinner terms, and averaging

---

2. Improving Reasoning for Diffusion Language Models via Group Diffusion Policy Optimization

URL: View paper

Prior Art Analysis

Group Diffusion Policy Optimization[25] demonstrates that prior work exists on reinforcement learning algorithms specifically designed for diffusion language models. Both papers address the same core challenge: adapting GRPO to diffusion models by improving token likelihood estimation. While the original paper proposes coupled-GRPO with complementary mask noise pairs, the candidate paper introduces GDPO with semi-deterministic Monte Carlo schemes for variance reduction. Both methods aim to reduce variance in probability estimation while maintaining efficiency, and both are presented as diffusion-native RL algorithms that respect the non-autoregressive nature of diffusion models.

Evidence

Evidence 1 - **Rationale**: Both papers present novel RL algorithms specifically designed for diffusion language models that address variance reduction in likelihood estimation, demonstrating that the candidate's work predates or parallels the original's novelty claim. - **Original**: we propose coupledgrpo, a novel sampling scheme that constructs complementary mask noise for completions used in training. in our experiments, coupled-grpo significantly improves diffucoder's performance on code generation benchmarks (+4.4% on evalplus) and reduces reliance on ar bias during decodin... - **Candidate**: we introducegroup diffusion policy optimization (gdpo), a new rl algorithm tailored for dlms. gdpo leverages simple yet effective semi-deterministic monte carloschemes to mitigate the variance explosion of elbo estimators under vanilla double monte carlo sampling, yielding a provably lower-variance ...

Evidence 2 - **Rationale**: Both papers identify variance reduction in likelihood estimation as the key technical challenge and propose novel sampling schemes as solutions, indicating prior work on this specific contribution. - **Original**: for rl training, to reduce the variance of token log-likelihood estimates and maintain training efficiency, we propose coupledgrpo, a novel sampling scheme that constructs complementary mask noise for completions used in training. - **Candidate**: in this work, we revisit elbo estimation and disentangle its sources of variance. this decomposition motivates reducing variance through fast, deterministic integral approximations along a few pivotal dimensions. building on this insight, we introducegroup diffusion policy optimization (gdpo), a new...

Evidence 3 - **Rationale**: Both papers provide theoretical justification for variance reduction in their respective sampling schemes, with the candidate using deterministic integration and the original using antithetic variates, showing parallel development of diffusion-native RL methods. - **Original**: we design coupled-grpo, an rl algorithm for dllms that avoids semi-ar decoding by using a novel coupled-sampling scheme for efficient and accurate policy gradient estimation (§5). we theoretically prove the variance reduction of coupled-grpo using antithetic variates. - **Candidate**: to achieve low-variance estimates under tight evaluation budgets, we limit naive monte carlo sampling and adoptdeterministicintegration methods to avoid the slow mc convergence of o(n-1/2). deterministic time:motivated by the observation in figure 2(a), instead of considering the problem as a double...

---

3. Ctrldiff: Boosting large diffusion language models with dynamic block prediction and controllable generation

URL: View paper

Brief Assessment

Ctrldiff[21] focuses on dynamic block prediction and classifier-guided controllable generation for diffusion language models, not on reinforcement learning algorithms for policy optimization in diffusion models.

---

4. Reward-weighted sampling: Enhancing non-autoregressive characteristics in masked diffusion llms

URL: View paper

Brief Assessment

Reward-weighted Sampling[20] focuses on decoding-time reward guidance for token selection during inference, not on training-time policy optimization like coupled-GRPO. The candidate addresses a different problem (inference decoding strategy) rather than RL training methodology.

---

5. MMaDA-Parallel: Multimodal Large Diffusion Language Models for Thinking-Aware Editing and Generation

URL: View paper

Brief Assessment

MMaDA-Parallel[24] focuses on parallel multimodal diffusion for text-image generation with trajectory-level RL (ParaRL), not on variance reduction techniques for non-autoregressive diffusion language models in code generation. The technical contexts differ substantially.

---

6. Sequence-augmented Conversational Recommendation System Based on Diffusion Models for Personalized Cultural Exploration

URL: View paper

Brief Assessment

Conversational Recommendation Diffusion[26] focuses on conversational recommendation systems for cultural exploration using diffusion models, not on reinforcement learning algorithms for non-autoregressive diffusion language models in code generation tasks.

---

7. Taming Masked Diffusion Language Models via Consistency Trajectory Reinforcement Learning with Fewer Decoding Step

URL: View paper

Brief Assessment

Consistency Trajectory Reinforcement[22] focuses on consistency between rollout and optimization trajectories in masked diffusion models, while the original paper's coupled-GRPO addresses variance reduction through complementary mask noise pairs. These are distinct technical approaches to RL for diffusion models.

---

8. Consolidating Reinforcement Learning for Multimodal Discrete Diffusion Models

URL: View paper

Brief Assessment

Multimodal Discrete Diffusion[17] focuses on multimodal (text and image) discrete diffusion with different technical approaches (fading-out masking for text, probabilistic decoding for images), while the original paper specifically addresses code generation with complementary mask noise pairs for variance reduction in token likelihood estimation.

---

9. Text diffusion with reinforced conditioning

URL: View paper

Brief Assessment

Text Diffusion Reinforced Conditioning[23] focuses on reinforced self-conditioning for text diffusion models using reward signals to prevent degradation, not on variance reduction in token likelihood estimation for non-autoregressive diffusion models. The technical approaches and problem formulations differ fundamentally.

---

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

URL: View paper

Brief Assessment

Step-Aware Policy Optimization[19] focuses on process-based rewards to address unstructured refinement in reasoning tasks, while coupled-GRPO addresses variance reduction in token likelihood estimation through complementary mask noise pairs for general code generation.

Appendix: Text Similarity Detection

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

References

- [0] DiffuCoder: Understanding and Improving Masked Diffusion Models for Code Generation View paper
- [1] Continuously augmented discrete diffusion model for categorical generative modeling View paper
- [2] TreeDiff: AST-Guided Code Generation with Diffusion LLMs View paper
- [3] Lookahead Unmasking Elicits Accurate Decoding in Diffusion Language Models View paper
- [4] CodeDiffuSe: A masked diffusion framework for structure-aware code completion and repair View paper
- [5] Latent Adaptation with Masked Policy for Diffusion Language Models View paper
- [6] CodeFusion: A Pre-trained Diffusion Model for Code Generation View paper
- [7] Path Planning for Masked Diffusion Models with Applications to Biological Sequence Generation View paper
- [8] Soft-Masked Diffusion Language Models View paper
- [9] Reinforcing the Diffusion Chain of Lateral Thought with Diffusion Language Models View paper
- [10] Path Planning for Masked Diffusion Model Sampling View paper
- [11] Plan for Speed--Dilated Scheduling for Masked Diffusion Language Models View paper
- [12] On Powerful Ways to Generate: Autoregression, Diffusion, and Beyond View paper
- [13] LLaDA-MoE: A Sparse MoE Diffusion Language Model View paper
- [14] dUltra: Ultra-Fast Diffusion Language Models via Reinforcement Learning View paper
- [15] Diffusion vs Autoregression: An Empirical Study on Code Comment Translation View paper
- [16] Guided Star-Shaped Masked Diffusion View paper
- [17] Consolidating Reinforcement Learning for Multimodal Discrete Diffusion Models View paper
- [18] Principled and Tractable RL for Reasoning with Diffusion Language Models View paper
- [19] Step-Aware Policy Optimization for Reasoning in Diffusion Large Language Models View paper
- [20] Reward-weighted sampling: Enhancing non-autoregressive characteristics in masked diffusion llms View paper
- [21] Ctrldiff: Boosting large diffusion language models with dynamic block prediction and controllable generation View paper
- [22] Taming Masked Diffusion Language Models via Consistency Trajectory Reinforcement Learning with Fewer Decoding Step View paper
- [23] Text diffusion with reinforced conditioning View paper

- [24] MMaDA-Parallel: Multimodal Large Diffusion Language Models for Thinking-Aware Editing and Generation View paper
- [25] Improving Reasoning for Diffusion Language Models via Group Diffusion Policy Optimization View paper
- [26] Sequence-augmented Conversational Recommendation System Based on Diffusion Models for Personalized Cultural Exploration View paper
- [27] Beyond autoregression: An empirical study of diffusion large language models for code generation View paper
- [28] Dream-coder 7b: An open diffusion language model for code View paper
- [29] Dream 7b: Diffusion large language models View paper
- [30] Seed diffusion: A large-scale diffusion language model with high-speed inference View paper
- [31] Mercury: Ultra-fast language models based on diffusion View paper
- [32] dKV-Cache: The Cache for Diffusion Language Models View paper
- [33] DDPT: Diffusion-Driven Prompt Tuning for Large Language Model Code Generation View paper
- [34] Diffusion-based Large Language Models Survey View paper
- [35] Directional Diffusion-Style Code Editing Pre-training View paper
- [36] Progressive Autoregressive Video Diffusion Models View paper
- [37] From Slow Bidirectional to Fast Autoregressive Video Diffusion Models View paper
- [38] ViD-GPT: Introducing GPT-style Autoregressive Generation in Video Diffusion Models View paper
- [39] Block Diffusion: Interpolating Between Autoregressive and Diffusion Language Models View paper
- [40] Ar-diffusion: Asynchronous video generation with auto-regressive diffusion View paper
- [41] Ditar: Diffusion transformer autoregressive modeling for speech generation View paper
- [42] Amd: Autoregressive motion diffusion View paper
- [43] Ar-diffusion: Auto-regressive diffusion model for text generation View paper
- [44] An efficient diffusion-based non-autoregressive solver for traveling salesman problem View paper
- [45] Thermalizer: Stable autoregressive neural emulation of spatiotemporal chaos View paper