

Novelty Assessment Report

Paper: Beyond Masks: Efficient, Flexible Diffusion Language Models via Deletion-Insertion Processes

PDF URL: <https://openreview.net/pdf?id=VbvXjs5f72>

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-29

Abstract

While Masked Diffusion Language Models (MDLMs) relying on token masking and unmasking have shown promise in language modeling, their computational efficiency and generation flexibility remain constrained by the masking paradigm. In this paper, we propose Deletion-Insertion Diffusion language models (DID) that rigorously formulate token deletion and insertion as discrete diffusion processes, replacing the masking and unmasking processes in current MDLMs. DID improves training and inference efficiency by eliminating two major sources of computational overhead in MDLMs: the computations on non-informative 1) $\langle \text{MASK} \rangle$ tokens inherent to its paradigm, and 2) $\langle \text{PAD} \rangle$ tokens introduced in variable-length settings. Furthermore, DID offers greater flexibility by: 1) natively supporting variable-length sequences without requiring fixed-length padding, and 2) an intrinsic self-correction mechanism during generation due to insertion that dynamically adjusts token positions. To train DID, we design a score-based approach that assigns scores to token insertion operations and derive appropriate training objectives. The objectives involve subsequence counting problems, which we efficiently solve via a parallelized dynamic programming algorithm. Our experiments across fixed and variable-length settings demonstrate the advantage of DID over baselines of MDLMs and existing insertion-based LMs, in terms of modeling performance, sampling quality, and training/inference speed.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **Discrete Diffusion Language Modeling via Deletion and Insertion Processes**

A total of **13 papers** were analyzed and organized into a taxonomy with **11 categories**.

Taxonomy Overview

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

- **Core Deletion-Insertion Diffusion Frameworks**
- **Masking-Based Discrete Diffusion**
- **Multimodal and Vision-Language Discrete Diffusion**
- **Applications and Specialized Domains**
- **Survey and Review Literature**

Complete Taxonomy Tree

- Discrete Diffusion Language Modeling via Deletion and Insertion Processes Survey Taxonomy
- Core Deletion-Insertion Diffusion Frameworks
 - Deletion-Insertion Process Formulations ★ (1 papers)
 - [0] Beyond Masks: Efficient, Flexible Diffusion Language Models via Deletion-Insertion Processes (Anon et al., 2026) [View paper](#)
 - Edit-Based Reconstruction Approaches (1 papers)
 - [1] Diffuser: Discrete diffusion via edit-based reconstruction (Reid, 2022) [View paper](#)
 - General Insertion and Deletion Corruption (1 papers)
 - [5] Beyond in-place corruption: Insertion and deletion in denoising probabilistic models (Johnson, 2021) [View paper](#)
 - Continuous Time Markov Chain Insertion Models (1 papers)
 - [13] A Continuous Time Markov Chain framework for Insertion Language Models (D Patel, n.d.) [View paper](#)
- Masking-Based Discrete Diffusion
 - Generalized Masked Diffusion (1 papers)
 - [7] Generalized Interpolating Discrete Diffusion (von Rã¼tte, 2025) [View paper](#)
 - Conditional and Controllable Masked Diffusion (2 papers)
 - [6] Conditional Discrete Diffusion Language Model (H Koh, 2025) [View paper](#)
 - [9] Classifiers Guided Controllable Text Generation for Discrete Diffusion Language Models (Hang Jiang, 2024) [View paper](#)
 - Sparse and Accelerated Masked Diffusion (1 papers)
 - [8] Sparse-LaViDa: Sparse Multimodal Discrete Diffusion Language Models (Shufan Li, 2025) [View paper](#)
- Multimodal and Vision-Language Discrete Diffusion (1 papers)
 - [12] LaViDa: A Large Diffusion Model for Vision-Language Understanding (S Li, n.d.) [View paper](#)
- Applications and Specialized Domains
 - Security and Watermarking (2 papers)
 - [2] STEAD: Robust Provably Secure Linguistic Steganography with Diffusion Language Model (Y Qi, 2025) [View paper](#)
 - [4] Watermarking Discrete Diffusion Language Models (Bhimaraju, 2025) [View paper](#)
 - Style Transfer and Machine Translation (1 papers)
 - [3] Curriculum pre-training for stylized neural machine translation (Aixiao Zou, 2024) [View paper](#)
- Survey and Review Literature (2 papers)
 - [10] A Survey of Discrete Diffusion for Text and Genomic Sequence Generation (L de Groot, n.d.) [View paper](#)

Narrative

Core task: discrete diffusion language modeling via deletion and insertion processes. The field has organized itself around several complementary directions. At the foundation lie core deletion-insertion diffusion frameworks that directly model text generation through iterative removal and addition of tokens, exemplified by works such as Deletion Insertion Diffusion[0] and Insertion Deletion Denoising[5]. A parallel branch focuses on masking-based discrete diffusion, which treats masked tokens as a special corruption mechanism rather than explicit deletions. Multimodal and vision-language discrete diffusion extends these ideas beyond pure text, incorporating image or cross-modal conditioning. Applications and specialized domains explore how these generative processes can be adapted to tasks like stylized translation or watermarking, while survey and review literature provides broader perspectives on the landscape.

Within the core deletion-insertion frameworks, a handful of works have explored different formulations of the corruption and denoising processes. Some emphasize curriculum-based strategies for controlling generation complexity, as seen in Curriculum Stylized Translation[3], while others investigate generalized interpolation schemes that unify deletion and insertion with other discrete transitions, such as Generalized Interpolating Diffusion[7]. Deletion Insertion Diffusion[0] sits squarely in this foundational branch, proposing a principled formulation of how tokens are iteratively deleted and reinserted during the diffusion process. Compared to earlier insertion-based models like Insertion Deletion Denoising[5], which laid groundwork for non-autoregressive generation, the original paper refines the theoretical underpinnings and explores richer dynamics between deletion and insertion steps. This positioning highlights ongoing efforts to balance tractability, expressiveness, and sample quality in discrete diffusion for language.

Related Works in Same Category

No sibling papers were found in the same taxonomy leaf. A taxonomy-subtopic-level comparison will be produced instead.

Taxonomy-Level Summary

The original leaf focuses on rigorous mathematical formulations of deletion-insertion processes as discrete diffusion with score-based training objectives. The sibling subtopics explore related but distinct approaches: continuous-time Markov chain formulations, edit-based reconstruction without formal diffusion frameworks, and general insertion-deletion corruption processes. Together, these represent different mathematical and algorithmic perspectives on non-masking discrete diffusion for language modeling.

Similarities: - All subtopics involve deletion and/or insertion operations as core mechanisms for discrete diffusion in language modeling - All move beyond simple masking-based approaches to enable variable-length sequence generation - All address the challenge of modeling discrete text through diffusion-like processes rather than traditional autoregressive generation

Differences: - The original leaf emphasizes score-based training objectives and rigorous diffusion formulations, while Edit-Based Reconstruction explicitly excludes formal diffusion frameworks - Continuous Time Markov Chain models use continuous-time formulations, contrasting with the discrete-step diffusion in the original leaf - General Insertion and Deletion Corruption focuses on broader corruption processes as a category distinction, while the original leaf emphasizes the mathematical rigor of specific formulations - The original leaf's scope is narrower and more mathematically precise, while siblings cover complementary algorithmic approaches (CTMC) and broader conceptual categories (general corruption)

Suggested Search Directions: - Investigate connections between score-based objectives in discrete diffusion and continuous-time Markov chain likelihood formulations - Explore whether edit-based reconstruction methods can be recast with formal diffusion objectives to bridge these categories - Examine how different mathematical formulations (discrete-step vs continuous-time, score-based vs likelihood-based) affect generation quality and computational efficiency

Sibling Subtopics

- **Continuous Time Markov Chain Insertion Models** (leaves: 1, papers: 1)
 - Scope: Formulations using continuous time Markov chains for modeling insertion and deletion processes in language generation.
 - Exclude: Excludes discrete-step diffusion formulations; see Deletion-Insertion Process Formulations.
- **Edit-Based Reconstruction Approaches** (leaves: 1, papers: 1)
 - Scope: Denoising diffusion models using edit operations for text reconstruction without fixed masking paradigms.
 - Exclude: Excludes formal deletion-insertion diffusion formulations; see Deletion-Insertion Process Formulations.
- **General Insertion and Deletion Corruption** (leaves: 1, papers: 1)
 - Scope: Broader corruption processes beyond in-place noise that insert and delete sequence elements during diffusion.
 - Exclude: Excludes models limited to in-place token corruption; see Masking-Based Discrete Diffusion.

Contributions Analysis

Overall novelty summary. The paper proposes Deletion-Insertion Diffusion (DID) language models that formulate token deletion and insertion as discrete diffusion processes, replacing masking paradigms in existing masked diffusion language models. According to the taxonomy, this work resides in the 'Deletion-Insertion Process Formulations' leaf under 'Core Deletion-Insertion Diffusion Frameworks'. Notably, this leaf contains only the original paper itself with zero sibling papers, indicating a relatively sparse research direction within the broader discrete diffusion landscape of thirteen total papers across eleven leaf nodes.

The taxonomy reveals that the broader field organizes around masking-based approaches (with three distinct subtopics including generalized, conditional, and sparse variants) versus explicit deletion-insertion frameworks. The original paper's leaf sits alongside three other leaves in the core frameworks branch: edit-based reconstruction, general insertion-deletion corruption, and continuous-time Markov chain formulations. The scope note explicitly excludes masking-based approaches and edit-based methods without formal diffusion formulation, positioning DID as pursuing rigorous mathematical foundations for deletion-insertion dynamics distinct from neighboring paradigms.

Among ten candidates examined for the simplified DICE objective contribution, zero were identified as refutable, though all ten were classified as non-refutable-or-unclear. The other two contributions—the core DID framework and the DISE training objective—had zero candidates examined, suggesting the literature search focused primarily on training methodology rather than the fundamental deletion-insertion formulation. Given the limited search scope of ten candidates total and the sparse taxonomy leaf (no siblings), the analysis provides initial signals but cannot comprehensively assess novelty across the full discrete diffusion literature.

Based on top-ten semantic matches, the work appears to occupy a distinct position within discrete diffusion language modeling, though the small candidate pool and absence of sibling papers in the taxonomy limit definitive conclusions. The analysis captures immediate neighborhood relationships but does not exhaustively cover all potential overlaps with masking-based methods or continuous-time formulations in adjacent taxonomy branches.

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

Contribution 1: Deletion-Insertion Diffusion language models (DID)

Description: The authors introduce DID, a novel discrete diffusion paradigm that replaces masking-unmasking in MDLMs with deletion-insertion processes. This eliminates and tokens, improving computational efficiency and enabling native variable-length sequence support with intrinsic self-correction during generation.

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

Contribution 2: Denoising Insertion Score Entropy (DISE) training objective

Description: The authors develop DISE, a score-based training objective for learning DID's insertion process. They define an insertion score modeling the probability of inserting any token at any position, derive the DISE objective involving subsequence count ratios, and provide an efficient parallelized dynamic programming algorithm to compute these ratios.

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

Contribution 3: Simplified DICE objective for fixed-length settings

Description: For fixed-length data, the authors show that the insertion score becomes time-independent and satisfies a sequence-level normalization property. This enables a simplified Denoising Insertion Cross Entropy (DICE) objective that improves parameterization and learning efficiency in fixed-length language modeling benchmarks.

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. Planner Aware Path Learning in Diffusion Language Models Training

URL: [View paper](#)

Brief Assessment

Planner Aware Path[20] focuses on aligning training with planner-based inference dynamics in diffusion language models, not on simplifying objectives for fixed-length settings through time-independent insertion scores and sequence-level normalization properties.

2. Optimizing medical image report generation through a discrete diffusion framework

URL: [View paper](#)

Brief Assessment

Medical Report Diffusion[16] focuses on medical image report generation using diffusion models with cross-entropy loss. It does not address fixed-length discrete diffusion language modeling or the specific insertion score normalization properties claimed in the original contribution.

3. Efficient Perplexity Bound and Ratio Matching in Discrete Diffusion Language Models

URL: [View paper](#)

Brief Assessment

Perplexity Ratio Matching[15] focuses on ratio-matching through score-entropy and denoising cross-entropy for discrete diffusion models, but does not address fixed-length sequence-level normalization properties or the simplified DICE objective derived from insertion score time-independence in fixed-length settings.

4. Conditional Discrete Diffusion Language Model

URL: [View paper](#)

Brief Assessment

Conditional Discrete Diffusion[6] focuses on conditional masked language models integrated with discrete diffusion for general text generation tasks, not specifically on fixed-length optimization or cross-entropy objectives for insertion-based diffusion models.

5. Classifier-free Diffusion Models for Machine Translation

URL: [View paper](#)

Brief Assessment

Classifier Free Translation[22] focuses on machine translation using diffusion models with classifier-free guidance, not on developing simplified cross-entropy objectives for fixed-length discrete diffusion language modeling.

6. Simplified and Generalized Masked Diffusion for Discrete Data

URL: [View paper](#)

Brief Assessment

Simplified Masked Diffusion[14] focuses on simplifying masked diffusion models with cross-entropy objectives for general discrete data, not specifically on insertion-based diffusion with fixed-length optimizations like DICE.

7. Authentic Discrete Diffusion Model

URL: [View paper](#)

Brief Assessment

Authentic Discrete Diffusion[19] focuses on discrete diffusion in one-hot space using cross-entropy objectives, but does not address fixed-length sequence normalization properties or the specific DICE simplification described in the original paper.

8. Structured Denoising Diffusion Models in Discrete State-Spaces

URL: [View paper](#)

Brief Assessment

Structured Discrete Diffusion[17] focuses on discrete state-space diffusion with transition matrices and auxiliary cross entropy losses, not on insertion-based diffusion or fixed-length sequence normalization properties that enable the DICE objective.

9. CGDD: Contrastive Gaussian-Dirac Diffusion Model

URL: [View paper](#)

Brief Assessment

CGDD[21] focuses on hybrid Gaussian-Dirac noise diffusion for text generation with contrastive learning, not on simplifying cross-entropy objectives for fixed-length discrete diffusion models.

10. A Cheaper and Better Diffusion Language Model with Soft-Masked Noise

URL: [View paper](#)

Brief Assessment

Soft Masked Noise[18] focuses on a different diffusion approach for language modeling using soft-masking strategies and cross-entropy objectives, rather than the deletion-insertion framework and subsequence counting methods central to the DICE objective.

Appendix: Text Similarity Detection

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

References

- [0] Beyond Masks: Efficient, Flexible Diffusion Language Models via Deletion-Insertion Processes [View paper](#)
- [1] Diffuser: Discrete diffusion via edit-based reconstruction [View paper](#)
- [2] STEAD: Robust Provably Secure Linguistic Steganography with Diffusion Language Model [View paper](#)
- [3] Curriculum pre-training for stylized neural machine translation [View paper](#)
- [4] Watermarking Discrete Diffusion Language Models [View paper](#)
- [5] Beyond in-place corruption: Insertion and deletion in denoising probabilistic models [View paper](#)
- [6] Conditional Discrete Diffusion Language Model [View paper](#)
- [7] Generalized Interpolating Discrete Diffusion [View paper](#)
- [8] Sparse-LaViDa: Sparse Multimodal Discrete Diffusion Language Models [View paper](#)
- [9] Classifiers Guided Controllable Text Generation for Discrete Diffusion Language Models [View paper](#)
- [10] A Survey of Discrete Diffusion for Text and Genomic Sequence Generation [View paper](#)
- [11] Diffusion Models in Text Generation: A [View paper](#)
- [12] LaViDa: A Large Diffusion Model for Vision-Language Understanding [View paper](#)
- [13] A Continuous Time Markov Chain framework for Insertion Language Models [View paper](#)
- [14] Simplified and Generalized Masked Diffusion for Discrete Data [View paper](#)
- [15] Efficient Perplexity Bound and Ratio Matching in Discrete Diffusion Language Models [View paper](#)
- [16] Optimizing medical image report generation through a discrete diffusion framework [View paper](#)
- [17] Structured Denoising Diffusion Models in Discrete State-Spaces [View paper](#)
- [18] A Cheaper and Better Diffusion Language Model with Soft-Masked Noise [View paper](#)
- [19] Authentic Discrete Diffusion Model [View paper](#)
- [20] Planner Aware Path Learning in Diffusion Language Models Training [View paper](#)
- [21] CGDD: Contrastive Gaussian-Dirac Diffusion Model [View paper](#)
- [22] Classifier-free Diffusion Models for Machine Translation [View paper](#)