

# Novelty Assessment Report

**Paper:** Continuously Augmented Discrete Diffusion model for Categorical Generative Modeling

**PDF URL:** <https://openreview.net/pdf?id=JNAZ3e7Bwt>

**Venue:** ICLR 2026 Conference Submission

**Year:** 2026

**Report Generated:** 2025-12-30

## Abstract

Standard discrete diffusion models treat all unobserved states the same way, typically mapping them to an absorbing [MASK] token. This creates an "information void" where global semantic information that may be inferred for the masked tokens from the unmasked tokens is not directly passed from one denoising step to another. We introduce **Continuously Augmented Discrete Diffusion (CADD)**, a framework that augments the discrete state space with a paired diffusion in a continuous latent space. This yields graded, gradually corrupted states in which masked tokens are represented by noisy yet informative latent vectors rather than information voids. At each reverse step, CADD uses the continuous latent as a semantic hint to guide discrete denoising. The design is clean and compatible with existing discrete diffusion training. At sampling time, the strength and estimator of the continuous latent vector enables a controlled trade-off between mode-coverage (diversity-oriented) and mode-seeking (context-localization-oriented). Empirically, we demonstrate CADD improves generative quality over mask-based diffusion across text generation, image synthesis, and code modeling, with consistent gains on both qualitative and quantitative metrics against strong discrete baselines.

### 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 for categorical data generation**

A total of **50 papers** were analyzed and organized into a taxonomy with **21 categories**.

### Taxonomy Overview

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

- **Theoretical Foundations and Training Objectives**
- **Architectural Innovations and Model Design**
- **Sampling and Inference Methods**
- **Controllable and Conditional Generation**
- **Domain-Specific Applications**
- **Continuous Embedding and Latent Space Methods**

### Complete Taxonomy Tree

- discrete diffusion for categorical data generation Survey Taxonomy
- Theoretical Foundations and Training Objectives
  - Score-Based and Ratio-Matching Formulations (3 papers)
  - [3] Discrete diffusion modeling by estimating the ratios of the data distribution (Lou, 2023) [View paper](#)
  - [25] Efficient Perplexity Bound and Ratio Matching in Discrete Diffusion Language Models (Haxholli, 2025) [View paper](#)
  - [48] Information-Theoretic Discrete Diffusion (Shin, 2025) [View paper](#)
  - Continuous-Time Discrete Markov Chain Frameworks (3 papers)
  - [5] Score-based continuous-time discrete diffusion models (Sun Hao-Ran, 2022) [View paper](#)
  - [18] Unified Discrete Diffusion for Categorical Data (Zhao, 2024) [View paper](#)
  - [24] Discrete-state continuous-time diffusion for graph generation (XU Zhe, 2024) [View paper](#)
  - Convergence Analysis and Theoretical Guarantees (2 papers)
  - [20] Convergence analysis of discrete diffusion model: Exact implementation through uniformization (Chen Hongrui, 2024) [View paper](#)
  - [22] Coevolutionary continuous discrete diffusion: Make your diffusion language model a latent reasoner (Zhou Cai, 2025) [View paper](#)
- Architectural Innovations and Model Design
  - Hybrid Continuous-Discrete Representations ★ (4 papers)
  - [0] Continuously Augmented Discrete Diffusion model for Categorical Generative Modeling (Anon et al., 2026) [View paper](#)
  - [19] CANDI: Hybrid Discrete-Continuous Diffusion Models (Shi, 2025) [View paper](#)
  - [23] Latent Discrete Diffusion Models (Durmus, 2025) [View paper](#)
  - [33] Disco-diff: Enhancing continuous diffusion models with discrete latents (Xu, 2024) [View paper](#)
  - Masked and Absorbing Diffusion Variants (4 papers)
  - [6] Masked diffusion models are secretly time-agnostic masked models and exploit inaccurate categorical sampling (Zheng, 2024) [View paper](#)
  - [15] Simplified and generalized masked diffusion for discrete data (Shi, 2024) [View paper](#)
  - [32] Glauber generative model: Discrete diffusion models via binary classification (Nagaraj, 2024) [View paper](#)
  - [41] Generalized Interpolating Discrete Diffusion (von Ráth, 2025) [View paper](#)
  - Structured Noise and Transition Matrices (2 papers)
  - [4] First hitting diffusion models for generating manifold, graph and categorical data (Ye Mao, 2022) [View paper](#)

- [12] Structured denoising diffusion models in discrete state-spaces (Jacob Austin, 2021) [View paper](#)
- Embedding and Representation Learning (3 papers)
- [9] Analog bits: Generating discrete data using diffusion models with self-conditioning (Chen Ting, 2022) [View paper](#)
- [44] Mitigating Embedding Collapse in Diffusion Models for Categorical Data (Nguyen, 2024) [View paper](#)
- [47] Authentic Discrete Diffusion Model (Li Xiao, 2025) [View paper](#)
- Sampling and Inference Methods
  - Distillation and Acceleration (2 papers)
  - [35] Minibatch Optimal Transport and Perplexity Bound Estimation in Discrete Flow Matching (Haxholli, 2024) [View paper](#)
  - [36] Distillation of Discrete Diffusion by Exact Conditional Distribution Matching (Yansong Gao, 2025) [View paper](#)
  - Deterministic and Derandomized Sampling (1 papers)
  - [50] Deterministic Discrete Denoising (Suzuki Hideyuki, 2025) [View paper](#)
  - Dependency Modeling in Denoising (1 papers)
  - [49] Discrete Copula Diffusion (Liu, 2024) [View paper](#)
- Controllable and Conditional Generation
  - Classifier-Based and Classifier-Free Guidance (1 papers)
  - [17] Simple guidance mechanisms for discrete diffusion models (Schiff, 2024) [View paper](#)
  - Constraint Enforcement and Optimization (1 papers)
  - [37] Constrained Discrete Diffusion (Michael Cardei, 2025) [View paper](#)
  - Conditional Tabular and Structured Data Synthesis (2 papers)
  - [14] Card: Classification and regression diffusion models (Han, 2022) [View paper](#)
  - [43] Controllable Tabular Data Synthesis Using Diffusion Models (Tongyu Liu, 2024) [View paper](#)
- Domain-Specific Applications
  - Natural Language and Text Generation (2 papers)
  - [11] Discrete Diffusion Models for Language Generation (Weligalle, 2025) [View paper](#)
  - [42] A Reparameterized Discrete Diffusion Model for Text Generation (Zheng Lin, 2023) [View paper](#)
  - Graph and Relational Structure Generation (1 papers)
  - [2] Digress: Discrete denoising diffusion for graph generation (Vignac, 2022) [View paper](#)
  - Tabular Data Synthesis (4 papers)
  - [13] Tabddpm: Modelling tabular data with diffusion models (Kotelnikov, 2023) [View paper](#)
  - [21] A comparative exploration of two diffusion generative models on tabular data synthesis (Neetu Kumari, 2025) [View paper](#)
  - [28] MissHDD: Hybrid Deterministic Diffusion for Hetrogeneous Incomplete Data Imputation (Youran Zhou, 2025) [View paper](#)
  - [30] Tabdiff: a mixed-type diffusion model for tabular data generation (Shi, 2024) [View paper](#)
  - Healthcare and Biomedical Data (3 papers)
  - [1] Guided discrete diffusion for electronic health record generation (Han, 2024) [View paper](#)
  - [45] A Hyperbolic Discrete Diffusion 3D RNA Inverse Folding Model for Functional RNA Design. (Dongyue Hou, 2025) [View paper](#)
  - [46] Domain-Generalized Discrete Diffusion Model for Cross-Domain Medical Image Segmentation. (Heran Yang, 2025) [View paper](#)
  - 3D Shape and Spatial Data Generation (4 papers)
  - [10] Diffusion probabilistic models for scene-scale 3d categorical data (Lee Jumin, 2023) [View paper](#)
  - [27] Generalized deep 3d shape prior via part-discretized diffusion process (Yuhan Li, 2023) [View paper](#)
  - [29] ShapeMorph: 3D Shape Completion via Blockwise Discrete Diffusion (Jiahui Li, 2025) [View paper](#)
  - [38] 3dqd: Generalized deep 3d shape prior via part-discretized diffusion process (Li YuHan, 2023) [View paper](#)
  - Design and Creative Applications (3 papers)
  - [7] SketchDNN: Joint Continuous-Discrete Diffusion for CAD Sketch Generation (Femiani, 2025) [View paper](#)
  - [26] LayoutDM: Discrete Diffusion Model for Controllable Layout Generation (Naoto Inoue, 2023) [View paper](#)
  - [31] Discrete Diffusion Probabilistic Models for Symbolic Music Generation (Plasser, 2023) [View paper](#)
  - Trajectory and Sequential Behavior Modeling (3 papers)
  - [16] Breaking determinism: Fuzzy modeling of sequential recommendation using discrete state space diffusion model (Xie Wenjia, 2024) [View paper](#)
  - [39] Synthetic location trajectory generation using categorical diffusion models (Hong Ye, 2024) [View paper](#)
  - [40] ScanDL: A Diffusion Model for Generating Synthetic Scanpaths on Texts (Lena S. Bolliger, 2023) [View paper](#)
- Continuous Embedding and Latent Space Methods (2 papers)
  - [8] Diffusing gaussian mixtures for generating categorical data (Regol, 2023) [View paper](#)
  - [34] Global context with discrete diffusion in vector quantised modelling for image generation (Minghui Hu, 2022) [View paper](#)

## Narrative

Core task: discrete diffusion for categorical data generation. The field has evolved into several major branches that reflect different strategic emphases. Theoretical Foundations and Training Objectives address the mathematical underpinnings and loss formulations needed to handle discrete state spaces, while Architectural Innovations and Model Design explore how network structures can be tailored to categorical variables—ranging from purely discrete transition matrices to hybrid continuous-discrete representations that embed categories into latent spaces. Sampling and Inference Methods focus on efficient denoising schedules and accelerated generation, Controllable and Conditional Generation investigates guidance mechanisms for steering outputs toward desired attributes, and Domain-Specific Applications demonstrate successes in areas such as tabular synthesis, molecular design, and layout generation. Continuous Embedding and Latent Space Methods form a complementary strand that leverages smooth representations to sidestep some of the challenges inherent in purely discrete transitions.

A particularly active line of work centers on hybrid continuous-discrete representations, where methods like CANDI[19], Latent Discrete Diffusion[23], and Disco-diff[33] embed categorical tokens into continuous spaces to enable smoother gradient flow and more expressive modeling. Continuously Augmented Discrete Diffusion[0] sits squarely within this cluster, proposing to augment discrete states with continuous auxiliary variables to improve training dynamics and sample quality. This contrasts with purely discrete approaches such as Discrete Diffusion Ratios[3] or Score-based Continuous-time Discrete[5], which operate directly on categorical distributions and face challenges related to gradient estimation and mode coverage. The trade-off between representational flexibility and computational overhead remains a central open question: hybrid methods often achieve stronger empirical performance on complex data but introduce additional hyperparameters and architectural complexity. By bridging discrete and continuous paradigms, the original paper[0] aligns closely with recent efforts to harness the best of both worlds, offering a pathway to more stable and scalable categorical generation.

## Related Works in Same Category

---

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

### 1. CANDI: Hybrid Discrete-Continuous Diffusion Models

**Authors:** Shi, Jiaxin, Patrick Pynadath, Zhang Ru-qi, Jiaxin Shi, et al. (6 authors total) | **Year/Venue:** 2025 | **URL:** [View paper](#)

#### Abstract

While continuous diffusion has shown remarkable success in continuous domains such as image generation, its direct application to discrete data has underperformed compared to purely discrete formulations. This gap is counterintuitive, given that continuous diffusion learns score functions that enable joint evolution across multiple positions. To understand this gap, we introduce token identifiability as an analytical framework for understanding how Gaussian noise corrupts discrete data through t...

#### Relationship Analysis

Both papers belong to the Hybrid Continuous-Discrete Representations category, combining continuous latent spaces with discrete state spaces for categorical data generation. They share the core approach of augmenting discrete diffusion with continuous embeddings to provide semantic hints during denoising, addressing the information loss problem in masked diffusion models. However, CADD focuses on augmenting masked diffusion by pairing discrete masking with conditional Gaussian diffusion on embeddings to preserve graded semantic proximity, while CANDI addresses the temporal dissonance between discrete identity corruption and continuous rank degradation by decoupling discrete masking from Gaussian noise schedules to enable simultaneous learning of conditional structure and continuous geometry.

---

### 2. Latent Discrete Diffusion Models

**Authors:** Durmus, Alain, Dario Shariatian, Peluchetti, Stefano, et al. (7 authors total) | **Year/Venue:** 2025 | **URL:** [View paper](#)

#### Abstract

We study discrete diffusion for language and other categorical data and focus on a common limitation of masked denoisers: reverse transitions typically factorize across positions, which can weaken joint structure and degrade quality in few-step generation. We propose **Latent Discrete Diffusion Models** (LDDMs), which couple a masked discrete diffusion over tokens with a continuous diffusion over latent embeddings. The latent channel provides a softer signal and carries cross-token dependenci...

#### Relationship Analysis

Both papers belong to the Hybrid Continuous-Discrete Representations category, combining continuous latent spaces with discrete state spaces for categorical data generation. They overlap in augmenting masked discrete diffusion with continuous embeddings to provide semantic hints during denoising, addressing the information void problem of pure masking approaches. The key difference is that CADD uses a simpler element-wise fusion of discrete and continuous embeddings with a single shared backbone, while Latent Discrete Diffusion Models (LDDMs) employ a more structured variational framework with separate encoder-decoder architectures and offer two distinct variants (FUJI for joint denoising and SEQ for sequential denoising).

---

### 3. Disco-diff: Enhancing continuous diffusion models with discrete latents

**Authors:** Xu, Yilun, Corso, Gabriele, Yilun Xu, et al. (15 authors total) | **Year/Venue:** 2024 | **URL:** [View paper](#)

#### Abstract

Diffusion models (DMs) have revolutionized generative learning. They utilize a diffusion process to encode data into a simple Gaussian distribution. However, encoding a complex, potentially multimodal data distribution into a single continuous Gaussian distribution arguably represents an unnecessarily challenging learning problem. We propose **Discrete-Continuous Latent Variable Diffusion Models** (DisCo-Diff) to simplify this task by introducing complementary discrete latent variables. We augment D...

#### Relationship Analysis

Both papers belong to the Hybrid Continuous-Discrete Representations category, combining continuous latent spaces with discrete state spaces for enhanced expressivity in diffusion models. While the original paper (CADD) augments masked discrete diffusion with a paired continuous Gaussian diffusion to provide semantic hints for token denoising, DisCo-Diff takes the opposite approach by augmenting continuous diffusion models with discrete latent variables to simplify the noise-to-data mapping and reduce ODE curvature. The key difference is that CADD starts from discrete diffusion and adds continuous guidance, whereas DisCo-Diff starts from continuous diffusion and adds discrete structure.

---

## Contributions Analysis

**Overall novelty summary.** The paper introduces Continuously Augmented Discrete Diffusion (CADD), which augments discrete diffusion with a paired continuous latent space to provide semantic hints during denoising. It resides in the 'Hybrid Continuous-Discrete Representations' leaf of the taxonomy, which contains four papers total (including this one). This leaf sits within the broader 'Architectural Innovations and Model Design' branch, indicating a moderately active research direction focused on combining continuous and discrete representations. The taxonomy shows this is a recognized but not overcrowded area, with sibling papers exploring similar embedding strategies.

The taxonomy reveals that CADD's leaf is adjacent to 'Masked and Absorbing Diffusion Variants' (four papers) and 'Structured Noise and Transition Matrices' (two papers), both of which operate primarily in discrete space without continuous augmentation. The 'Embedding and Representation Learning' leaf (three papers) addresses related concerns about representation quality but focuses on learning embeddings rather than joint diffusion processes. The 'Continuous Embedding and Latent Space Methods' branch (two papers) maps discrete data to continuous spaces but does not maintain the hybrid structure CADD proposes. This positioning suggests CADD bridges multiple research threads while occupying a distinct methodological niche.

Among thirty candidates examined, the contribution-level analysis shows mixed novelty signals. The core CADD framework (Contribution 1) examined ten candidates and found one refutable match, suggesting some prior work explores continuous-discrete augmentation. The 'graded semantic hints' mechanism (Contribution 2) and the 'mode-coverage versus mode-seeking trade-off' (Contribution 3) each examined ten candidates with zero refutations, indicating these specific design choices appear less directly anticipated in the limited search scope. The statistics reflect a focused but not exhaustive literature review, leaving open the possibility of additional relevant work beyond the top-thirty semantic matches.

Given the limited search scope of thirty candidates, the analysis suggests CADD occupies a recognizable but not densely populated research direction. The hybrid continuous-discrete approach has precedent in the taxonomy's sibling papers, yet the specific mechanism of using continuous latents as semantic hints during discrete denoising appears less directly covered. The controlled trade-off between diversity and context-localization represents a design contribution that, within the examined candidates, lacks clear prior instantiation. A broader literature search might reveal additional overlaps, particularly in adjacent fields like variational autoencoders or semi-discrete generative models.

---

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

## Contribution 1: Continuously Augmented Discrete Diffusion (CADD) framework

**Description:** CADD augments masked discrete diffusion models with a continuous latent space that preserves semantic information for masked tokens. Instead of collapsing masked positions into information voids, the framework maintains noisy yet informative latent vectors that guide discrete denoising at each reverse step.

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. Disco-diff: Enhancing continuous diffusion models with discrete latents

URL: [View paper](#)

#### Brief Assessment

Disco-diff[33] focuses on augmenting continuous diffusion models with discrete latents to simplify the noise-to-data mapping, whereas CADD augments discrete masked diffusion models with continuous latents to preserve semantic information for masked tokens. These are fundamentally different approaches addressing opposite modeling directions.

---

### 2. Coevolutionary continuous discrete diffusion: Make your diffusion language model a latent reasoner

URL: [View paper](#)

#### Prior Art Analysis

Coevolutionary Continuous Discrete[22] demonstrates that prior work exists on augmenting discrete diffusion with continuous latent spaces. The candidate paper explicitly proposes 'coevolutionary continuous discrete diffusion (ccdd), which defines a joint diffusion process on both the discrete state space through *ctmc* and the continuous probability or embedding space through *sde*'. This directly addresses the same core concept as CADD: combining discrete and continuous diffusion processes where masked tokens are represented by continuous latent vectors rather than pure absorbing states. Both papers solve the identical problem of information loss in masked diffusion by maintaining semantic information through continuous representations.

#### Evidence

Evidence 1 - **Rationale:** Both papers propose frameworks that augment discrete diffusion with continuous latent spaces. CADD uses 'paired diffusion in a continuous latent space' while CCDD uses 'joint multimodal diffusion process on the union of a continuous representation space and a discrete token space'. The core innovation of representing masked tokens with continuous vectors is identical. - **Original:** we introduce continuously augmented discrete diffusion (cadd), a framework that augments the discrete state space with a paired diffusion in a continuous latent space. this yields graded, gradually corrupted states in which masked tokens are represented by noisy yet informative latent vectors rather... - **Candidate:** we therefore propose coevolutionary continuous discrete diffusion (ccdd), which defines a joint multimodal diffusion process on the union of a continuous representation space and a discrete token space, leveraging a single model to simultaneously denoise in the joint space. by combining two modalities, ...

Evidence 2 - **Rationale:** Both papers describe joint forward processes that evolve discrete tokens and continuous latents together. CADD's 'paired conditional gaussian diffusion' and CCDD's 'joint corruption process' serve the same purpose of maintaining semantic information for masked positions through continuous representations. - **Original:** cadd retains the discrete masking trajectory and augments it with a paired conditional gaussian diffusion in a continuous semantic embedding space. our forward process jointly evolves the token sequence and its latent, so positions that are masked in the discrete path are accompanied by noisy yet in... - **Candidate:** now we introduce coevolutionary continuous discrete diffusion (ccdd), a diffusion model on the joint of discrete and continuous space  $x \times z$  (figure 3). in particular, we consider a joint corruption process  $(x_t, z_t) \in x \times z$  that applies noise individually to each component, and a denoising process that con...

Evidence 3 - **Rationale:** Both papers identify the same fundamental problem: discrete masked diffusion loses information by collapsing masked states. This shared problem motivation demonstrates that CCDD addresses the identical challenge that CADD claims as novel. - **Original:** standard discrete diffusion models treat all unobserved states identically by mapping them to an absorbing [mask] token. this creates an "information void" where semantic information that could be inferred from unmasked tokens is lost between denoising steps. - **Candidate:** intriguingly, discrete diffusions with masked noises tend to outperform those with uniform noises (amin et al., 2025), at the cost of losing self-correction capabilities. analogously to llms, discrete dlms also reason in the explicit token space and may partially lose information of previous decodin...

---

### 3. Continuous latent variables

URL: [View paper](#)

#### Brief Assessment

Continuous Latent Variables[67] discusses latent-space rotations and categorical variables in a general context, but does not present a framework for augmenting masked discrete diffusion models with continuous latent spaces that preserve semantic information during denoising.

---

### 4. Continuous Diffusion Model for Language Modeling

URL: [View paper](#)

#### Brief Assessment

Continuous Diffusion Language[64] focuses on modeling discrete data through continuous flow on the statistical manifold using Riemannian geometry, not on augmenting masked discrete diffusion with continuous latent spaces that preserve semantic information for masked tokens.

---

### 5. Argmax flows and multinomial diffusion: Learning categorical distributions

URL: [View paper](#)

#### Brief Assessment

Argmax Flows[63] focuses on learning categorical distributions through argmax transformations and multinomial diffusion processes, not on augmenting masked discrete diffusion with continuous latent spaces that preserve semantic information for masked tokens.

---

### 6. Latent diffusion models for controllable rna sequence generation

URL: [View paper](#)

#### Brief Assessment

Latent Diffusion RNA[65] focuses on RNA sequence generation using a latent diffusion model with pretrained encoders and Q-formers, not on augmenting masked discrete diffusion with continuous latent spaces for general categorical data.

---

### 7. Length-aware motion synthesis via latent diffusion

URL: [View paper](#)

#### Brief Assessment

Length-aware Motion Synthesis[66] focuses on motion generation with length-dependent latent codes for controlling sequence duration and dynamics, not on augmenting discrete diffusion models with continuous latent spaces for categorical data generation.

---

### **8. Ddmi: Domain-agnostic latent diffusion models for synthesizing high-quality implicit neural representations**

URL: [View paper](#)

#### **Brief Assessment**

DDMI[62] focuses on generating adaptive positional embeddings for implicit neural representations across multiple domains (images, 3D shapes, videos), not on augmenting discrete diffusion models with continuous latent spaces for categorical data generation as in CADD.

---

### **9. Tabdiff: a mixed-type diffusion model for tabular data generation**

URL: [View paper](#)

#### **Brief Assessment**

TabDiff[30] focuses on tabular data generation with mixed numerical and categorical features, not on augmenting masked discrete diffusion with continuous latent spaces for general categorical data modeling.

---

### **10. Ladir: Latent diffusion enhances llms for text reasoning**

URL: [View paper](#)

#### **Brief Assessment**

LaDiR[61] focuses on latent diffusion for text reasoning using a VAE-based latent space with blockwise diffusion, not on augmenting masked discrete diffusion models with continuous latent spaces for categorical generative modeling as in CADD.

---

## **Contribution 2: Graded semantic hints for token prediction**

**Description:** The continuous latent provides graded proximity information to ground-truth embeddings for masked positions, reducing ambiguity in token prediction. This addresses the information loss problem in standard masked diffusion where all unobserved states are treated identically.

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. ExLM: Rethinking the Impact of Tokens in Masked Language Models**

URL: [View paper](#)

#### **Brief Assessment**

ExLM[77] addresses semantic corruption in MLMs through state expansion and dependency modeling, while the original paper focuses on continuous latent diffusion augmentation for discrete diffusion models. These are fundamentally different technical approaches to different model architectures.

---

### **2. Latent resonance pathways for large language models through gradient-synchronized semantic fluxion**

URL: [View paper](#)

#### **Brief Assessment**

Latent Resonance Pathways[74] discusses semantic propagation via 'gated transport' and 'lexical ambiguity' in masked models, but does not present a concrete mechanism for providing graded proximity information to ground-truth embeddings for masked token prediction as described in the original contribution.

---

### **3. Tackling Ambiguity from Perspectives of Uncertainty Inference and Affinity Diversification for Weakly Supervised Semantic Segmentation**

URL: [View paper](#)

#### **Brief Assessment**

Uncertainty Affinity Diversification[70] addresses ambiguity in weakly supervised semantic segmentation using uncertainty estimation and affinity diversification for image-level labels, not masked token prediction in generative modeling. The technical domains and problem settings are fundamentally different.

---

### **4. Semantic depth redistribution in large language models to contextual embedding preservation**

URL: [View paper](#)

#### **Brief Assessment**

Semantic Depth Redistribution[76] focuses on redistributing semantic information across hierarchical layers in language models and uses dense token masking as a methodology. This is architecturally distinct from CADD's approach of augmenting discrete diffusion with continuous latent variables to provide graded proximity hints for masked token prediction.

---

### **5. Stochastic lexical dissonance injection for self-consistent reasoning in large language models: A quantitative investigation**

URL: [View paper](#)

#### **Brief Assessment**

Stochastic Lexical Dissonance[68] focuses on input noise and semantic cues in pre-training strategies, not on continuous latent representations providing graded proximity information for masked token prediction in diffusion models.

---

### **6. Context-aware alignment and mutual masking for 3d-language pre-training**

URL: [View paper](#)

#### **Brief Assessment**

Context-aware 3D-Language[75] focuses on aligning 3D point cloud features with language features for vision-language tasks, not on masked token prediction in discrete diffusion models for text generation.

---

### **7. EviGraph-LLMRec: Evidential Graph-Language Model Fusion for Uncertainty-Aware Recommendation**

URL: [View paper](#)

#### **Brief Assessment**

EviGraph-LLMRec[71] focuses on recommendation systems with uncertainty-aware fusion of graph and language models, not on masked token prediction or discrete diffusion models for generative modeling.

---

## 8. Towards a novel architecture for semantic pattern resolution in large language models

URL: [View paper](#)

### Brief Assessment

Semantic Pattern Resolution[73] focuses on semantic pattern resolution and linguistic relationships in language models, not on graded proximity information in masked diffusion for reducing token prediction ambiguity.

---

## 9. Context-aware masking and learnable diffusion-guided patch refinement in transformers via sparse supervision for hyperspectral image classification

URL: [View paper](#)

### Brief Assessment

Context-aware Masking Hyperspectral[72] focuses on hyperspectral image classification using diffusion-guided feature refinement and contextualized masking for vision transformers, not on discrete token prediction in language/categorical modeling. The technical domains and problem settings are fundamentally different.

---

## 10. OCR-Assisted Masked BERT for Homoglyph Restoration towards Multiple Phishing Text Downstream Tasks

URL: [View paper](#)

### Brief Assessment

OCR-Assisted Masked BERT[69] focuses on homoglyph restoration in phishing text using OCR preprocessing and character-level BERT, not on graded semantic hints in masked diffusion models for general token prediction.

---

### Contribution 3: Controlled mode-coverage versus mode-seeking trade-off

**Description:** The framework enables flexible control between diversity and precision during inference through the choice of continuous latent estimator (hard versus soft) and resampling strategies. This allows users to balance between generating diverse outputs and contextually precise outputs.

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. End-to-end learning of gaussian mixture priors for diffusion sampler

URL: [View paper](#)

#### Brief Assessment

Gaussian Mixture Priors[52] addresses mode-seeking vs. mode-coverage in the context of diffusion-based sampling from unnormalized densities, not in discrete categorical generation. The original paper's contribution focuses on controlling this trade-off during inference in discrete diffusion models through continuous latent estimators and resampling strategies, which is a different technical approach and application domain.

---

### 2. One-step Diffusion Models with F-Divergence Distribution Matching

URL: [View paper](#)

#### Brief Assessment

One-step f-Divergence[60] focuses on controlling mode-seeking versus mode-coverage through choice of f-divergence in one-step diffusion distillation, not through continuous latent estimators and resampling strategies during inference as in the original paper's framework.

---

### 3. Importance Weighted Score Matching for Diffusion Samplers with Enhanced Mode Coverage

URL: [View paper](#)

#### Brief Assessment

Importance Weighted Score[54] focuses on training diffusion samplers to achieve mode coverage through importance sampling techniques, not on providing flexible inference-time control between diversity and precision via different latent estimators and resampling strategies as in the original paper.

---

### 4. Unrolled generative adversarial networks

URL: [View paper](#)

#### Brief Assessment

Unrolled GANs[55] addresses mode collapse in GANs through unrolled discriminator optimization, not through continuous latent estimators or resampling strategies for controlling diversity-precision trade-offs during inference.

---

### 5. Mcl-gan: Generative adversarial networks with multiple specialized discriminators

URL: [View paper](#)

#### Brief Assessment

MCL-GAN[58] focuses on using multiple discriminators to address mode collapse in GANs, not on controlling mode-coverage versus mode-seeking trade-offs during inference through latent estimator choices and resampling strategies as in the original paper.

---

### 6. One-step Diffusion Models with -Divergence Distribution Matching

URL: [View paper](#)

#### Brief Assessment

One-step Divergence Matching[51] focuses on controlling mode-seeking versus mode-coverage through choice of f-divergence (e.g., forward-KL vs reverse-KL) in one-step diffusion distillation, not through continuous latent estimators and resampling strategies in discrete diffusion as proposed in the original paper.

---

### 7. Improving generative adversarial networks via adversarial learning in latent space

URL: [View paper](#)

#### Brief Assessment

Adversarial Latent Learning[57] focuses on improving GANs through adversarial learning in latent space for image generation, not on discrete diffusion models or the specific mode-coverage versus mode-seeking trade-off mechanism described in the original paper's continuous latent estimator design.

---

### 8. Flow to the Mode: Mode-Seeking Diffusion Autoencoders for State-of-the-Art Image Tokenization

URL: [View paper](#)

## Brief Assessment

Flow to Mode[59] focuses on image tokenization using diffusion autoencoders with mode-seeking post-training for perceptual reconstruction quality. The ORIGINAL paper addresses mode-coverage versus mode-seeking trade-offs in discrete text/code generation through continuous latent estimators and resampling strategies. These are fundamentally different application domains and technical approaches.

---

## 9. Dual discriminator generative adversarial nets

URL: [View paper](#)

### Brief Assessment

Dual Discriminator GANs[56] addresses mode collapse in GANs through dual discriminators that balance KL and reverse KL divergences, but does not provide inference-time control mechanisms for trading off diversity versus precision like CADD's continuous latent estimator choices and resampling strategies.

---

## 10. Training neural samplers with reverse diffusive kl divergence

URL: [View paper](#)

### Brief Assessment

Reverse Diffusive KL[53] addresses mode-coverage vs mode-seeking in the context of training neural samplers for Boltzmann distributions using diffusive KL divergence, not in discrete diffusion models for categorical generation with continuous latent estimators and resampling strategies.

---

## Appendix: Text Similarity Detection

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

---

## References

- [0] Continuously Augmented Discrete Diffusion model for Categorical Generative Modeling [View paper](#)
- [1] Guided discrete diffusion for electronic health record generation [View paper](#)
- [2] Digress: Discrete denoising diffusion for graph generation [View paper](#)
- [3] Discrete diffusion modeling by estimating the ratios of the data distribution [View paper](#)
- [4] First hitting diffusion models for generating manifold, graph and categorical data [View paper](#)
- [5] Score-based continuous-time discrete diffusion models [View paper](#)
- [6] Masked diffusion models are secretly time-agnostic masked models and exploit inaccurate categorical sampling [View paper](#)
- [7] SketchDNN: Joint Continuous-Discrete Diffusion for CAD Sketch Generation [View paper](#)
- [8] Diffusing gaussian mixtures for generating categorical data [View paper](#)
- [9] Analog bits: Generating discrete data using diffusion models with self-conditioning [View paper](#)
- [10] Diffusion probabilistic models for scene-scale 3d categorical data [View paper](#)
- [11] Discrete Diffusion Models for Language Generation [View paper](#)
- [12] Structured denoising diffusion models in discrete state-spaces [View paper](#)
- [13] Tabddpm: Modelling tabular data with diffusion models [View paper](#)
- [14] Card: Classification and regression diffusion models [View paper](#)
- [15] Simplified and generalized masked diffusion for discrete data [View paper](#)
- [16] Breaking determinism: Fuzzy modeling of sequential recommendation using discrete state space diffusion model [View paper](#)
- [17] Simple guidance mechanisms for discrete diffusion models [View paper](#)
- [18] Unified Discrete Diffusion for Categorical Data [View paper](#)
- [19] CANDI: Hybrid Discrete-Continuous Diffusion Models [View paper](#)
- [20] Convergence analysis of discrete diffusion model: Exact implementation through uniformization [View paper](#)
- [21] A comparative exploration of two diffusion generative models on tabular data synthesis [View paper](#)
- [22] Coevolutionary continuous discrete diffusion: Make your diffusion language model a latent reasoner [View paper](#)
- [23] Latent Discrete Diffusion Models [View paper](#)
- [24] Discrete-state continuous-time diffusion for graph generation [View paper](#)
- [25] Efficient Perplexity Bound and Ratio Matching in Discrete Diffusion Language Models [View paper](#)
- [26] LayoutDM: Discrete Diffusion Model for Controllable Layout Generation [View paper](#)
- [27] Generalized deep 3d shape prior via part-discretized diffusion process [View paper](#)
- [28] MissHDD: Hybrid Deterministic Diffusion for Heterogeneous Incomplete Data Imputation [View paper](#)
- [29] ShapeMorph: 3D Shape Completion via Blockwise Discrete Diffusion [View paper](#)
- [30] Tabdiff: a mixed-type diffusion model for tabular data generation [View paper](#)
- [31] Discrete Diffusion Probabilistic Models for Symbolic Music Generation [View paper](#)
- [32] Glauber generative model: Discrete diffusion models via binary classification [View paper](#)
- [33] Disco-diff: Enhancing continuous diffusion models with discrete latents [View paper](#)
- [34] Global context with discrete diffusion in vector quantised modelling for image generation [View paper](#)
- [35] Minibatch Optimal Transport and Perplexity Bound Estimation in Discrete Flow Matching [View paper](#)
- [36] Distillation of Discrete Diffusion by Exact Conditional Distribution Matching [View paper](#)
- [37] Constrained Discrete Diffusion [View paper](#)
- [38] 3dqd: Generalized deep 3d shape prior via part-discretized diffusion process [View paper](#)
- [39] Synthetic location trajectory generation using categorical diffusion models [View paper](#)
- [40] ScanDL: A Diffusion Model for Generating Synthetic Scanpaths on Texts [View paper](#)
- [41] Generalized Interpolating Discrete Diffusion [View paper](#)
- [42] A Reparameterized Discrete Diffusion Model for Text Generation [View paper](#)
- [43] Controllable Tabular Data Synthesis Using Diffusion Models [View paper](#)
- [44] Mitigating Embedding Collapse in Diffusion Models for Categorical Data [View paper](#)
- [45] A Hyperbolic Discrete Diffusion 3D RNA Inverse Folding Model for Functional RNA Design. [View paper](#)
- [46] Domain-Generalized Discrete Diffusion Model for Cross-Domain Medical Image Segmentation. [View paper](#)
- [47] Authentic Discrete Diffusion Model [View paper](#)
- [48] Information-Theoretic Discrete Diffusion [View paper](#)

- [49] Discrete Copula Diffusion [View paper](#)
- [50] Deterministic Discrete Denoising [View paper](#)
- [51] One-step Diffusion Models with  $\beta$ -Divergence Distribution Matching [View paper](#)
- [52] End-to-end learning of gaussian mixture priors for diffusion sampler [View paper](#)
- [53] Training neural samplers with reverse diffusive kl divergence [View paper](#)
- [54] Importance Weighted Score Matching for Diffusion Samplers with Enhanced Mode Coverage [View paper](#)
- [55] Unrolled generative adversarial networks [View paper](#)
- [56] Dual discriminator generative adversarial nets [View paper](#)
- [57] Improving generative adversarial networks via adversarial learning in latent space [View paper](#)
- [58] Mcl-gan: Generative adversarial networks with multiple specialized discriminators [View paper](#)
- [59] Flow to the Mode: Mode-Seeking Diffusion Autoencoders for State-of-the-Art Image Tokenization [View paper](#)
- [60] One-step Diffusion Models with F-Divergence Distribution Matching [View paper](#)
- [61] Ladir: Latent diffusion enhances llms for text reasoning [View paper](#)
- [62] Ddmi: Domain-agnostic latent diffusion models for synthesizing high-quality implicit neural representations [View paper](#)
- [63] Argmax flows and multinomial diffusion: Learning categorical distributions [View paper](#)
- [64] Continuous Diffusion Model for Language Modeling [View paper](#)
- [65] Latent diffusion models for controllable rna sequence generation [View paper](#)
- [66] Length-aware motion synthesis via latent diffusion [View paper](#)
- [67] Continuous latent variables [View paper](#)
- [68] Stochastic lexical dissonance injection for self-consistent reasoning in large language models: A quantitative investigation [View paper](#)
- [69] OCR-Assisted Masked BERT for Homoglyph Restoration towards Multiple Phishing Text Downstream Tasks [View paper](#)
- [70] Tackling Ambiguity from Perspectives of Uncertainty Inference and Affinity Diversification for Weakly Supervised Semantic Segmentation [View paper](#)
- [71] EviGraph-LLMRec: Evidential Graph-Language Model Fusion for Uncertainty-Aware Recommendation [View paper](#)
- [72] Context-aware masking and learnable diffusion-guided patch refinement in transformers via sparse supervision for hyperspectral image classification [View paper](#)
- [73] Towards a novel architecture for semantic pattern resolution in large language models [View paper](#)
- [74] Latent resonance pathways for large language models through gradient-synchronized semantic fluxion [View paper](#)
- [75] Context-aware alignment and mutual masking for 3d-language pre-training [View paper](#)
- [76] Semantic depth redistribution in large language models to contextual embedding preservation [View paper](#)
- [77] ExLM: Rethinking the Impact of Tokens in Masked Language Models [View paper](#)