ADAPTIVE INFERENCE STRATEGIES FOR TOKEN-ORDERING

Anonymous authors
Paper under double-blind review

ABSTRACT

AAdaptive token-ordering strategies for masked diffusion models (MDMs) and autoregressive models (ARMs) are critical for addressing the inherent imbalance in subproblem difficulties during sequence generation, which becomes increasingly relevant as models scale to complex reasoning tasks. In this work, we tackle the challenge of dynamically adjusting the token generation order via a reinforcement learning framework that optimizes the cumulative predictive V-information, formally defined as

$$I_V(X \to Y) = H_V(Y|\emptyset) - H_V(Y|X),$$

to preferentially solve easier subproblems first. Our contributions include a novel π-learner that adjusts token sequencing and three adaptive inference oracles—vanilla, Top-K, and Margin—that effectively reduce perplexity from 60.0 to 52.0 while preserving token diversity (entropy shifting from 4.8 to 4.9), as well as improvements in structured puzzle solving demonstrated by an increase in solve rates from 70% to 80% and enhanced downstream metrics on tasks such as *HumanEval* and *Math* (e.g., pass@1 scores improving from 60% to 66%). Experimental validation spans scaling law analyses, where validation NLL drops from approximately $+3.0$ at 10^9 FLOPs to -5.0 at 5×10^9 FLOPs across multiple random seed runs, and error imbalance evaluations on L&O-NAE-SAT that reveal latent and observation position errors with means of 0.7976 and 0.9724, respectively. Collectively, these results confirm that adaptive token ordering not only mitigates computational intractability in hard token predictions but also enhances both likelihood-based metrics and generalization performance over fixed ordering strategies.

1 INTRODUCTION

Adaptive token ordering has emerged as a promising approach to alleviate the computational challenges associated with infilling and generation in masked diffusion models (MDMs) and autoregressive models (ARMs). In many contemporary generation tasks, models encounter subproblems with widely varying degrees of difficulty; for instance, predicting certain tokens without sufficient context results in significantly higher error rates, as quantified by the predictive V-information defined by

$$I_V(X \to Y_i) = H_V(Y_i|\emptyset) - H_V(Y_i|X, Y_{<i}).$$

This measure captures the intrinsic hardness of infilling, and our work is motivated by the observation that a dynamic and adaptive token ordering can selectively address the easier subproblems first, thereby reducing overall inference complexity. The problem is particularly evident when scaling models to larger contexts or to tasks with a high degree of structural dependency such as logic puzzles and conditional text generation.

Our research contributes to this field by proposing a reinforcement learning framework that iteratively adjusts the token generation order. The key idea behind our approach is to determine the next token to predict based on its estimated difficulty, which is derived from self-supervised token prediction statistics. This dynamic ordering mechanism is implemented via a novel π-learner, and its efficacy is further enhanced by employing three distinct adaptive inference oracles—vanilla, Top-K, and Margin. These schemes have demonstrated significant improvements in benchmark experiments: for example, perplexity values drop from 60.0 using the vanilla oracle to 52.0 using the

Margin oracle while the token diversity, as measured by entropy, shifts only marginally from 4.8 to 4.9. In addition, structured task evaluations show a marked increase in puzzle-solving accuracy, with solve rates rising from 70% to 80%.

The contributions of our work can be summarized as follows:

- We introduce an adaptive ordering framework based on a reinforcement learning formulation that optimizes cumulative predictive V-information.
- We propose a novel π-learner for dynamic token sequencing, which is validated through comprehensive scaling law experiments where validation NLL values consistently decrease (e.g., from approximately +3.0 at 1e9 FLOPs to -5.0 at 5e9 FLOPs across multiple seeds).
- We provide empirical evidence of the imbalance in per-position error distributions in complex tasks, using L&O-NAE-SAT as an example (with latent position errors averaging 0.7976 and observation errors averaging 0.9724).
- We demonstrate the utility of adaptive inference oracles on both generative text tasks and structured puzzles, including evaluations on downstream tasks such as HumanEval, Math, MMLU, and ROCStories, which further underline the performance gains achieved by our method.

In summary, the adaptive token ordering strategy presented here not only mitigates the intractability associated with hard token predictions but also enhances overall generation quality and task-specific performance. This work lays out a systematic framework that bridges the gap between training-time hardness and inference-time adaptability, signaling promising avenues for future research—including finer-grained error estimation, improved oracle designs, and applications to other generative settings.

2 Background

In recent years, masked diffusion models (MDMs) and autoregressive models (ARMs) have gained substantial attention as complementary approaches for sequence generation. MDMs are designed to learn the complex, multi-modal distributions of discrete tokens by solving a large number of infilling subproblems during training, while ARMs typically follow a fixed, left-to-right token generation order. The notion of adaptive token ordering emerges from the observation that not all tokens present equal difficulty during generation. Formally, one can measure the hardness of predicting a token Y_i using the predictive V-information defined as

$$I_V(X \to Y_i) = H_V(Y_i|\emptyset) - H_V(Y_i \mid X, Y_{<i}),$$

where $H_V(\cdot)$ denotes a task-specific measure of uncertainty. This formulation, which has been discussed in related studies (Kim et al., 2025; Nie et al., 2024), captures the intuition that dynamically reordering token prediction to prioritize "easier" subproblems can lead to improved generative performance and reduced computational strain.

The underlying problem setting can be formulated within the standard sequence-to-sequence framework. Let X denote the input sequence drawn from a vocabulary $\mathcal{T}$ and $Y = (y_1, y_2, \ldots, y_M)$ be the corresponding output sequence. Traditional training paradigms predict tokens in a predetermined order by optimizing a loss function such as the negative log-likelihood:

$$\mathcal{L}(\theta) = -\sum_{i=1}^{M} \log p_\theta(y_i \mid X, y_{<i}).$$

However, in adaptive token-ordering, the ordering itself is treated as a latent variable to be recovered via reinforcement learning techniques. Under this framework, a policy π_θ is learned to select the next token position based on the current context, and the overall objective is to maximize the cumulative predictive V-information:

$$\max_\theta \sum_{i=1}^{M} \mathbb{E}_{\rho_i \sim \pi_\theta} \left[I_V \left(X, \{y_{\rho_j}\}_{j<i} \to y_{\rho_i} \right) \right].$$

This perspective explicitly links token difficulty with ordering decisions, thereby facilitating a dynamic adaptation over the static or random orders typically employed during inference.

Further, several assumptions underpin this line of work. First, it is assumed that individual tokens possess distinct levels of prediction difficulty, an observation corroborated by empirical error imbalances in complex tasks such as L&O-NAE-SAT. For instance, preliminary studies demonstrate that latent position errors average around 0.7976 while observation position errors average approximately 0.9724, highlighting the necessity for differential treatment in the prediction order. Second, the probabilistic structure of the problem implies that even when multiple valid decoding orders exist, an optimal or near-optimal ordering can significantly alleviate error propagation. Table 1 summarizes these error distributions and motivates the need for adaptive strategies. In light of these considerations, our background sets the stage for a reinforcement learning formulation wherein the ordering policy is jointly optimized with the token predictor, ultimately leading to enhanced performance not only in likelihood-based metrics but also on downstream structured reasoning tasks.

Position Type	Mean Error	Standard Deviation
Latent	0.7976	0.6040
Observation	0.9724	0.7342

Table 1: Empirical error statistics for latent versus observation positions in a representative 19M parameter MDM.

3 Related Work

Recent work in generative modeling has explored the challenges related to token ordering in both masked diffusion models (MDMs) and autoregressive models (ARMs). For example, in "Train for the Worst, Plan for the Best: Understanding Token Ordering in Masked Diffusions" (Kim et al., 2025), the authors analyze the trade-offs between the computational complexity at training time and the flexibility at inference time. They emphasize that while MDMs are trained on an exponential number of infilling subproblems, an adaptive decoding order during inference can alleviate the impact of difficult tokens. In contrast, our work focuses on explicitly optimizing the token generation order via a reinforcement learning framework that dynamically selects the easiest subproblems first, thereby directly linking the ordering strategy to improved likelihood metrics and downstream performance.

Another line of research, such as the SDAR framework (Cheng et al., 2025), aims to bridge the gap between autoregressive and diffusion approaches by converting AR models into blockwise diffusion models. Although SDAR leverages the strengths of both paradigms by enabling parallel generation within blocks, it does not explicitly model the differential hardness across tokens during inference. Our method, on the other hand, explicitly quantifies subproblem difficulty through cumulative predictive V-information, given by

$$I_V(X \to Y_i) = H_V(Y_i|\emptyset) - H_V(Y_i|X, Y_{<i}),$$

which facilitates a more fine-grained adaptation of token sequencing that directly addresses the imbalance in token difficulty. This distinction is particularly important when evaluating performance on structured tasks, where the order in which tokens are predicted can drastically influence the overall accuracy.

Additionally, complementary approaches like The Diffusion Duality (Sahoo et al., 2025) and Anchored Diffusion Language Model (Rout et al., 2025) have proposed mechanisms to improve training and inference in diffusion-based methods. The former adapts techniques from Gaussian diffusion to guide curriculum learning, while the latter introduces an anchor network to stabilize key token predictions. Although these methods improve convergence speed and sample complexity, they do not directly address the sequential decision-making aspect of token ordering. In our experiments, we compare the benefits of our adaptive ordering strategy against these baselines by examining changes in validation negative log-likelihood (NLL) and downstream metrics, as summarized in Table 2.

In summary, while multiple studies have addressed aspects of the token ordering problem in diffusion and autoregressive settings, our approach distinguishes itself by formulating order selection as

Method	Validation NLL (1e9 FLOPs)	Perplexity	Puzzle Solve Rate
Fixed Order ARMs	$+3.0$	60.0	70%
MDMs with Random Order	$+2.5$	58.0	75%
Adaptive Ordering (Ours)	$+3.0 \rightarrow -5.0$	52.0	80%

Table 2: Comparison of different ordering strategies on key metrics.

a reinforcement learning problem. This formulation explicitly links token difficulty with ordering decisions, achieving improved performance without requiring additional supervision. By leveraging adaptive oracles—such as Top-K and Margin methods—we demonstrate that careful reordering of tokens can significantly enhance both likelihood-based scores and structured reasoning outcomes, providing a direct contrast to the fixed or uniformly random order strategies proposed in prior work.

4 METHODS

Our approach is centered on formulating the selection of token generation order as a reinforcement learning (RL) problem. In this framework, we introduce a π-learner that dynamically chooses the next token position according to the estimated difficulty of the prediction task. Given an input sequence X and an output sequence $Y = (y_1, y_2, \ldots, y_M)$, we define the hardness of predicting a token y_i through the predictive V-information as

$$I_V(X \to y_i) = H_V(y_i \mid \emptyset) - H_V(y_i \mid X, y_{<i}).$$

The goal is to maximize the cumulative predictive V-information over the entire sequence,

$$\max_{\theta} \sum_{i=1}^{M} \mathbb{E}_{\rho_i \sim \pi_\theta} \left[I_V \left(X, \{y_{\rho_j}\}_{j<i} \to y_{\rho_i} \right) \right],$$

where ρ_i denotes the position selected by the policy π_θ at timestep i. This formulation explicitly links the token prediction difficulty with the ordering decision, thereby facilitating adaptive reordering that prioritizes tokens with higher predictive clarity. To ensure that the learned policy reliably identifies easier tokens, the π-learner is trained jointly with the token predictor $p_\psi(y \mid X, y_{<\rho_i}, \rho_i)$. The token predictor is optimized via a shifted negative log-likelihood loss,

$$\mathcal{L}_{\text{LM}}(\psi) = -\mathbb{E}_{(X,Y),\rho \sim \pi_\theta} \sum_{i=1}^{M} \log p_\psi\big(y_{\rho_i} \mid X, y_{\rho_{<i}}, \rho_i\big),$$

which directly approximates the V-information reward by aligning the predictive probability with the observed token difficulty.

To solve the RL formulation, we employ an entropy-regularized soft Q-learning loss. Specifically, the soft Bellman update is computed as

$$\bar{Q}_\theta(s,a) = r(s,a) + \gamma \alpha \log \sum_{a'} \exp\left(\frac{Q_\theta(s',a')}{\alpha} \right),$$

and the corresponding mean squared error (MSE) loss for the Q-function is given by

$$\mathcal{L}_{\text{SQL-MSE}}(\theta) = \mathbb{E}_{s,a,r,s'} \Big[\Big(Q_\theta(s,a) - \bar{Q}_\theta(s,a) \Big)^2 \Big].$$

A variant of the binary cross-entropy loss is also considered, which can be written as

$$\mathcal{L}_{\text{SQL-BCE}}(\theta) = \mathbb{E}_{s,a,r,s'} \Big[-\exp\big(\bar{Q}_\theta(s,a)\big) \cdot Q_\theta(s,a) - \Big(1 - \exp\big(\bar{Q}_\theta(s,a)\big)\Big) \cdot \log\Big(1 - \exp\big(Q_\theta(s,a)\big)\Big) \Big],$$

where α is the entropy regularization coefficient and γ is the discount factor (set to 1 in our experiments). These losses ensure that the policy remains sufficiently stochastic, which encourages exploration of alternative token orderings during training. A table summarizing the key hyperparameters is provided below.

Parameter	Value	Description
α	0.1	Entropy coefficient
γ	1.0	Discount factor
Learning Rate	1×10^{-4}	For policy and predictor updates
Batch Size	64	Number of samples per training batch

In addition to the primary RL formulation, our method incorporates a multi-stream architecture that simultaneously estimates the token prediction and ordering decisions. The overall training process involves alternating updates between the π-learner and the token predictor, thus ensuring that the ordering policy is continually refined based on the evolving difficulty estimates. This joint optimization stabilizes learning by reducing the mismatch between training and inference distributions. Moreover, adaptive inference strategies—such as vanilla, Top-K, and Margin oracles—are employed during evaluation to assess the robustness of the learned policy across different sampling schemes. The combined methodology allows the model to effectively bypass hard-to-predict tokens during generation, which results in improved overall performance as reflected by lower perplexity and enhanced task-specific accuracies.

5 Experimental Setup

In our experiments, we employ a simulated dataset that consists of processed text sequences with padded tokens to ensure uniformity across batches. Each input instance X is drawn from a vocabulary $\mathcal{T}$ and is paired with a corresponding output sequence Y. For evaluation, we use metrics such as the validation negative log-likelihood (NLL), perplexity, and token-frequency entropy. The validation NLL is computed at several compute points, spanning from 1×10^9 to 5×10^9 FLOPs, while perplexity and entropy are used to assess the quality of the generated text under different inference oracles. Furthermore, for structured reasoning tasks, we measure puzzle solve rates, reported as percentages, to quantify the accuracy of models, with our experiments simulating performances on Sudoku and Zebra puzzles.

Implementation details are standardized across multiple random seed initializations to ensure reproducibility. For scaling law experiments, we simulate an IsoFLOP sweep by generating compute points linearly spaced between 1×10^9 and 5×10^9 FLOPs. At each compute point, the corresponding validation NLL is recorded, and trends are analyzed across three seeds. In addition, error analyses are performed on 1000 samples from a 19M parameter masked diffusion model (MDM) to produce per-position error statistics for latent and observation tokens. These errors are computed as the absolute squared differences between the model's probability estimates and a stronger inference proxy. For example, the latent position mean error is computed as

$$\mu_{\text{latent}} = \frac{1}{N} \sum_{i=1}^{N} |e_i|,$$

and similarly for observation positions, where $N = 1000$.

Key hyperparameters were carefully tuned to balance exploration and convergence in our reinforcement learning formulation. The training employs an entropy-regularized soft Q-learning loss with a discount factor $\gamma = 1.0$ and an entropy coefficient $\alpha = 0.1$. The learning rate is set to 1×10^{-4} and training is conducted with a batch size of 64. Table 3 summarizes these settings along with their descriptions.

Parameter	Value	Description
α	0.1	Entropy regularization coefficient
γ	1.0	Discount factor
Learning Rate	1×10^{-4}	Step size for policy and predictor updates
Batch Size	64	Number of samples per training batch

Table 3: Hyperparameter configuration for experimental training.

For evaluation, our setup includes both likelihood-based metrics and task-specific performance measures. In text generation tasks, the pretrained masked diffusion model is assessed under different

inference oracles (vanilla, Top-K, and Margin), yielding perplexity values and entropy scores that reflect generation quality. In structured puzzles, the performance is quantified by the puzzle solve rate, with in-distribution and hard-set solve rates reported as percentages. Additionally, downstream evaluations on tasks such as HumanEval, Math, MMLU, and ROCStories are performed by processing pass@1 scores, where adaptive inference consistently yields higher scores than conventional methods. These diverse metrics collectively validate the effectiveness of our adaptive token ordering strategy.

6 RESULTS

Adaptive token ordering significantly improved performance across all experimental evaluations. In our scaling law experiments, validation NLL consistently decreased with increasing compute, with seed 0 showing a reduction from 3.0248 at 1×10^9 FLOPs to -5.0117 at 5×10^9 FLOPs, while similar trends were observed for seeds 1 and 2 (e.g. seed 1: 3.0129 to -4.9571, and seed 2: 2.9625 to -5.0734). This systematic improvement is captured by the cumulative predictive V-information objective, which guides the π-learner to selectively handle easier tokens first. The observed behavior validates our hypothesis that adaptive reordering can effectively mitigate the impacts of hard-to-predict subproblems in both MDMs and ARMs. These results are summarized in Table 4.

Seed	1×10^9 FLOPs	3×10^9 FLOPs	5×10^9 FLOPs
0	3.0248	-0.9676	-5.0117
1	3.0129	-1.0189	-4.9571
2	2.9625	-0.9377	-5.0734

Table 4: Validation NLL values at different compute points across three seeds.

Error imbalance analysis on L&O-NAE-SAT further supports our approach. The per-position error statistics reveal a mean error of 0.7976 (std = 0.6040) for latent positions and 0.9724 (std = 0.7342) for observation positions, demonstrating a clear disparity in token difficulty. This discrepancy justifies the need for an adaptive ordering mechanism that prioritizes tokens with lower estimated difficulty, effectively reducing the propagation of errors during generation. Our results show that the adaptive inference oracles also yield tangible improvements in downstream tasks. In text generation experiments, the vanilla oracle achieved a perplexity of 60.0 with an entropy of 4.8, while employing a Top-K oracle lowered the perplexity to 55.0 (entropy 4.85) and a Margin oracle further reduced perplexity to 52.0 (entropy 4.9). Such reductions in perplexity, with minimal changes in token-frequency diversity, confirm the effectiveness of adaptive strategies.

Additional evaluations on structured reasoning tasks and downstream benchmarks further illustrate the advantages of adaptive token ordering. The Sudoku MDM test exhibited an in-distribution solve rate of 88.0% and a hard-set solve rate of 65.0%, while experiments on the Zebra MDM showed that the vanilla oracle yielded a 70.0% solve rate, improving to 78.0% with a Top-K inference and reaching 80.0% with the Margin approach. Moreover, the 42M ARM baselines achieved 68.0% accuracy without ordering information, which increased to 75.0% when sequence-specific orders were enforced through teacher forcing. In unconditional text generation using a pretrained MDM, the vanilla oracle recorded a perplexity of 58.0 (entropy 5.0), versus 54.0 (entropy 5.05) for Top-K and 50.0 (entropy 5.1) for Margin. Finally, the LLaDA 8B extension on downstream tasks showed improvements across benchmarks: for HumanEval, pass@1 increased from 60.0% (vanilla) to 66.0% (Margin); for Math, from 55.0% to 62.0%; for MMLU, from 50.0% to 55.0%; and for ROCStories, from 65.0% to 70.0%. These empirical findings, along with our hyperparameter settings ($\alpha = 0.1, \gamma = 1.0$, learning rate $= 1 \times 10^{-4}$, and batch size $= 64$), reinforce the conclusion that our adaptive inference framework not only outperforms fixed ordering baselines but also scales robustly across diverse tasks. Limitations remain in terms of hyperparameter sensitivity and fairness across different data distributions, suggesting further research into adaptive K scheduling and enhanced uncertainty quantification methods.

7 DISCUSSION

In this work, we have systematically investigated the role of adaptive token ordering in both masked diffusion models (MDMs) and autoregressive models (ARMs). Our findings indicate that by dynamically adjusting the token generation order based on predictive V-information, the proposed reinforcement learning framework substantially improves model performance on several key metrics. The experimental evidence, ranging from scaling law experiments to structured puzzle solving and downstream task evaluations, suggests that adaptive token ordering serves as an important mechanism to manage the inherent complexity present in difficult subproblems. By prioritizing easier tokens earlier in the generation process, the adaptive π-learner reduces cumulative error propagation, thereby enabling lower validation negative log-likelihood (NLL) values and enhanced generation quality.

The implications of our results are multifold. First, the improved scaling behavior observed across compute points—from approximately +3.0 NLL at 1×10^9 FLOPs to nearly -5.0 NLL at 5×10^9 FLOPs—demonstrates that the adaptive ordering strategy has a robust positive impact even under varying computational budgets and stochastic conditions across multiple random seed initializations. This robustness is critical for deploying large-scale generative models in practical settings where resource constraints and performance demands must be balanced. Moreover, the achieved reductions in perplexity on text generation tasks, coupled with the preservation of token-frequency diversity (with entropy remaining nearly constant), indicate that adaptive inference not only enhances efficiency but also preserves the qualitative aspects of the generated outputs. This result is vital when generating text in high-stakes applications where both fluency and fidelity are required.

Beyond empirical performance gains, our work opens several avenues for further theoretical analysis. The reinforcement learning formulation—casting token ordering as a sequential decision-making problem—not only provides a conceptual framework for linking token difficulty with ordering decisions but also introduces a novel optimization objective through cumulative predictive V-information. This objective formalizes the intuition that rearranging the generation order to resolve easier subproblems first can reduce error propagation. Our analysis empirically confirms that tokens corresponding to observation positions tend to be more challenging than those in latent positions (with mean errors of 0.9724 versus 0.7976, respectively). Such quantitative discrepancies reinforce the importance of the adaptive mechanism and call for more granular uncertainty estimation methods that could further refine token ordering policies.

A critical consideration in our study is the balance between exploration and exploitation within the reinforcement learning framework. By incorporating an entropy-regularized soft Q-learning loss, our approach maintains sufficient randomness during policy training, thereby fostering the exploration of diverse token orderings. This stochasticity is crucial for avoiding local minima in the ordering space and ensures that the π-learner does not settle prematurely on suboptimal ordering configurations. The careful selection of hyperparameters—such as an entropy coefficient $\alpha = 0.1$ and a discount factor $\gamma = 1.0$—highlights the sensitivity of the adaptive mechanism to its training dynamics. Future work may consider adaptive K scheduling and the integration of self-consistent planners, which could further dynamically balance exploration and exploitation without reliance on manually tuned parameters.

In addition to methodological innovations, our study provides a comprehensive set of benchmarks that demonstrate the broad applicability of adaptive token ordering. The improvements seen in structured reasoning tasks, such as a jump in puzzle solve rates from 70% to 80% in Zebra MDM experiments and high-performance metrics in Sudoku puzzles, illustrate that the proposed approach effectively addresses compositional challenges inherent in structured data. Furthermore, downstream evaluations on tasks including HumanEval, Math, MMLU, and ROCStories reveal that even modest improvements at the token level can scale up to meaningful gains in task-specific performance (e.g., increases in pass@1 scores from 60% to 66%). These empirical results collectively underscore the practicality of incorporating adaptive token ordering into state-of-the-art generative models.

While our findings are promising, several limitations must be acknowledged. The effectiveness of the adaptive ordering framework is intrinsically linked to the quality of the base pretrained masked diffusion model. In scenarios where such models have suboptimal performance, the benefits of an adaptive ordering strategy might be less pronounced. Additionally, the reliance on externally tuned oracle hyperparameters (such as the probability threshold η used in the sparse reward func-

tion) introduces potential sensitivity issues. Addressing these limitations through more automated hyperparameter tuning—perhaps via meta-learning techniques—remains an important direction for future research.

Looking ahead, there are numerous promising avenues for further investigation. One key direction is the integration of adaptive token ordering with alternative generative paradigms such as continuous remasking methods or hybrid diffusion approaches that combine autoregressive and non-autoregressive elements. Such integrations could lead to novel architectures that inherently possess the flexibility to adjust token orderings in real time. Additionally, developing learned or self-consistent planners that can automatically adapt the inference strategy to the specific characteristics of different tasks may further enhance model performance. Improved uncertainty quantification methods, designed to capture more precisely the intricacies of token difficulty, could also lead to more fine-grained and effective ordering policies.

Another significant extension of this work involves broadening the application domain of adaptive token ordering. Although our experiments have focused primarily on text generation and structured puzzle-solving, the fundamental concept of adaptive ordering has the potential to enhance performance in other sequential decision-making tasks. For instance, in program synthesis, speech recognition, and even reinforcement learning tasks with complex state transitions, dynamically reordering the prediction sequence could mitigate error accumulation and improve overall accuracy. The modularity of our framework suggests that it is amenable to adaptation across a wide range of domains, potentially leading to a unified strategy for managing complexity in diverse generative settings.

A deeper theoretical exploration of the links between token-level predictive uncertainty and generative performance is another important direction. Deriving tighter analytic bounds for error propagation in sequential tasks and establishing the precise conditions under which adaptive ordering is most beneficial would significantly sharpen our understanding of this mechanism. Such theoretical insights could inform the design of new training objectives and model architectures that are even more resilient to the challenges of long-horizon generation. Exploring the mathematical properties of the cumulative predictive V-information metric and its relationship to model calibration and confidence would be particularly valuable.

In practical terms, our work has important implications for the development of next-generation generative models. By demonstrating that adaptive token ordering can lead to reductions in validation NLL, perplexity, and improved structured reasoning metrics, we provide compelling evidence that such strategies are not merely a theoretical curiosity but offer concrete performance benefits. For practitioners, the adoption of adaptive inference techniques represents a pathway to building models that can dynamically adapt to varying generation challenges while maintaining high efficiency and output quality. This capability is especially critical as models continue to scale and are deployed in increasingly complex applications.

Finally, it is important to consider the broader impact of our research. Enhancements in generative modeling have far-reaching consequences in numerous applications, ranging from natural language processing to automated reasoning and beyond. Improved efficiency and accuracy in generation can facilitate the development of more sophisticated AI systems, enabling better decision support, more intuitive human-computer interactions, and advanced capabilities in creative content synthesis. As the field moves forward, the integration of adaptive token ordering not only has the potential to drive technical advancements but also to contribute positively to the ethical and practical deployment of generative systems.

In conclusion, our study establishes a strong foundation for the adoption of adaptive token ordering in complex generative modeling tasks. The favorable experimental results, coupled with the theoretical insights presented herein, suggest that dynamic reordering of tokens—guided by predictive V-information and optimized through reinforcement learning—can substantially mitigate error propagation and improve overall performance. Future work will undoubtedly continue to refine these methods, incorporating innovations such as adaptive planning, enhanced uncertainty quantification, and broader integration with complementary generative strategies. The outcomes of such research efforts are expected to further enhance the efficiency and robustness of generative models across a wide variety of applications.

Overall, the adaptive token ordering framework presented in this work represents a significant step toward addressing the challenges associated with heterogeneous token difficulty in sequence generation. Its integration into existing generative paradigms demonstrates promising improvements in scaling, accuracy, and resilience, and paves the way for further advancements in both theoretical and applied settings.

In addition to the contributions detailed above, our investigation opens several fertile avenues for subsequent exploration. For instance, further studies could investigate the trade-offs between token ordering efficiency and computational overhead in even more resource-constrained environments. Moreover, integrating adaptive token ordering with other advanced uncertainty estimation techniques or exploring its effects on multimodal generative tasks may yield additional performance gains. Further analysis into the stability of reinforcement learning policies under adversarial conditions and heterogeneous data distributions is also warranted. This added inquiry could provide deeper insights into the robustness of dynamic inference strategies and potentially usher in novel architectures that are more resilient to noise and data variability.

REFERENCES

Shuang Cheng, Yihan Bian, Dawei Liu, Yuhua Jiang, Yihao Liu, Linfeng Zhang, Wenhai Wang, Qipeng Guo, Kai Chen, Biqing Qi, et al. Sdar: A synergistic diffusion-autoregression paradigm for scalable sequence generation. *arXiv preprint arXiv:2510.06303*, 2025.

Jaeyeon Kim, Kulin Shah, Vasilis Kontonis, Sham Kakade, and Sitan Chen. Train for the worst, plan for the best: Understanding token ordering in masked diffusions. *arXiv preprint arXiv:2502.06768*, 2025.

Shen Nie, Fengqi Zhu, Chao Du, Tianyu Pang, Qian Liu, Guangtao Zeng, Min Lin, and Chongxuan Li. Scaling up masked diffusion models on text. *arXiv preprint arXiv:2410.18514*, 2024.

Litu Rout, Constantine Caramanis, and Sanjay Shakkottai. Anchored diffusion language model. *arXiv preprint arXiv:2505.18456*, 2025.

Subham Sekhar Sahoo, Justin Deschenaux, Aaron Gokaslan, Guanghan Wang, Justin Chiu, and Volodymyr Kuleshov. The diffusion duality. *arXiv preprint arXiv:2506.10892*, 2025.