FAST LANGUAGE GENERATION THROUGH DISCRETE DIFFUSION DIVERGENCE INSTRUCT

Anonymous authors
Paper under double-blind review

ABSTRACT

The fast generation of language texts is the holy grail that people pursue in the AI era. In this work, we introduced **Di**screte **Di**ffusion Divergence **Instruct** (**DiDi-Instruct**), a training-based method that leads to fast language generation models by initializing from a pre-trained (masked) discrete diffusion language model (dLLM). The resulting DiDi-Instruct model outperforms the dLLM counterparts and the GPT-2 baseline with 64× acceleration. In the theoretical part of the paper, we build the foundation of DiDi-Instruct in a framework of integral KL-divergence minimization, with practical training algorithms. We also introduce techniques like *grouped reward normalization*, *intermediate-state matching*, and the *reward-guided ancestral sampler* (RGAS) that significantly improve the training stability, the model coverage, and the inference performances. On OpenWebText, **DiDi-Instruct** outperforms **all** accelerated language generation models as well as the GPT-2 baseline and the standard dLLMs, achieving sample perplexities ranging from 62.2 (8 NFEs) to 18.4 (128 NFEs). These performance gains are accomplished with a negligible entropy loss of about 1% and $20\times$ less additional training wall-clock time. We further validate the robustness and effectiveness of DiDi-Instruct through extensive ablation studies, model scaling, and the generation of discrete protein sequences. In conclusion, DiDi-Instruct is an efficient yet effective distillation method, enabling language generation in the blink of an eye. We will release our code and models along with the paper.

1 INTRODUCTION

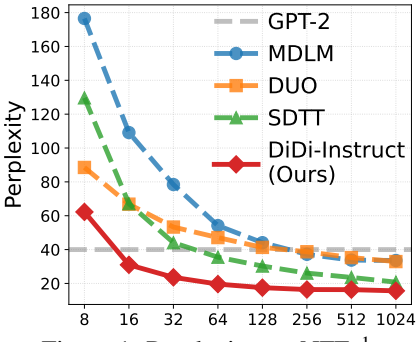

Figure 1: Perplexity vs. NFEs[1].

Fast language sequence generation has been a long-standing goal for large artificial intelligence systems for decades. Being one of the most typical large language models (LLMs), the Auto-regressive (AR) models have achieved remarkable success across a wide spectrum of natural language tasks (Brown et al., 2020; Meta AI, 2023; Achiam et al., 2023; Team et al., 2024). By predicting the next token from left to right using causal attention of transformer neural networks (Vaswani et al., 2017), such models can scale to billions of parameters and deliver state-of-the-art performance (Touvron et al., 2023). However, the interior mechanism that underpins their success imposes an inherent bottleneck: tokens are generated sequentially one at a time, limiting parallelism and throughput at scales (Leviathan et al., 2023; Kim et al., 2023; Ning et al., 2024). *Even with advanced computational techniques, such as KV caches, AR models still have a significant throughput ceiling.*

Emerging as a competitive language generation paradigm, the Discrete diffusion large language models (dLLMs) (Li et al., 2022a) reinterpret text generation as an iterative text tokens denoising process inspired by continuous diffusion models for images (Song et al., 2021b; Ho et al., 2020). Starting from a fully corrupted token sequence, the dLLM first trains and then applies a learned

[1]We compare DiDi-Instruct (ours, solid red line) with MDLM, DUO, SDTT (dashed lines), and GPT-2 small baseline (≈ 40.0 PPL). Our approach consistently achieves lower perplexity across all sampling budgets.

sequence denoiser several times to recover the clean text token sequences. As a result, this inference paradigm enables dLLMs to use the bidirectional attention inherited from the transformer network, and therefore, can generate sequences with fewer times of model evaluations than AR models.

Despite their good performances and improved efficiency, dLLMs still face the inference efficiency bottleneck in generation steps. For instance, on the OpenWebText benchmark, *a dLLM still needs up to 256 generation steps to match the performance of a GPT2 model when generating a sequence of a length of 1024.* To mitigate these issues, several distillation methods for dLLMs have been proposed to facilitate few-step generators to produce high-quality samples without performance degradation. For example, Deschenaux & Gulcehre (2025) proposed the SDTT that trains a student to match a teacher's predictions across time-steps, allowing the student to generate at least 32 tokens at once and achieving speedups up to eight times over AR models. Sahoo et al. (2025) introduced DUO that connects USDMs to Gaussian diffusion and introduces curriculum learning and discrete consistency distillation to enable few-step sampling. Zhu et al. (2025b) proposed the DiMO that matches the token-level distribution of MDMs and uses a token initialization strategy to produce a one-step generator. We defer a detailed, side-by-side introduction and discussions of their distillation mechanisms in Appendix B.3. *Though these early studies have pioneered the frontier of fast language generation, their generation throughput still has much room to improve: on the OpenWebText benchmark, even the most recent SDTT can not match the performance of GPT2 baseline within 32 generation steps.* Moreover, existing distillation methods are more straightforward in principle rather than well-established theoretical frameworks.

In this work, we introduce **Di**screte **Di**ffusion Divergence **Instruct** (**DiDi-Instruct**), a novel distillation framework for fast language generation. The principle of **DiDi-Instruct** is to minimize an Integral KL divergence (IKL) introduced by the seminal work of Luo et al. (2023b) in the continuous diffusion model. By minimizing IKL between the distributions of a few-step student generation model and a pre-trained dLLM, the few-step student can match the ability of dLLM in a distribution-matching manner with significantly improved efficiency. However, directly adapting IKL divergence to MDMs presents unique challenges due to discrete, non-differentiable operations that build the foundation of language sequence generation. To overcome the challenges, one of our key contributions is a comprehensive and effective solution through innovations in objective design, training stability, and inference efficiency. Specifically, we make the following contributions:

- ***Principled Training Method for Fast (Language) Sequence Generation***: We reformulate the distillation objective from a general policy gradient perspective, deriving a simple yet tractable update rule for the few-step student according to some reward function. With an adversarial language discriminator to estimate the log-density ratio (reward) between the teacher dLLM and the student, we introduce a practical **DiDi-Instruct** algorithm that trains the few-step student, accompanied by an assistant discriminator;
- ***Simple yet Effective Techniques in Training***: We introduce *grouped reward normalization*, *intermediate-state matching*, and the *reward-guided ancestral sampler* (RGAS) that significantly improve the training stability, the model coverage, and the inference performances, reducing the generation perplexity by 30%.
- ***State-of-the-art Fast Sequence Generation:*** **DiDi-Instruct** models achieve new state-of-the-art results on OpenWebText (OWT) benchmark: consistently lower PPL across 8-128 number of function evaluations (NFEs) (Figure 1), negligible entropy loss, and over 20$\times$ faster distillation; detailed ablations, model scaling, and protein generation further confirm robustness.

To the best of our knowledge, DiDi-Instruct is the first framework to successfully apply distribution-matching distillation to MDM-based text generation, along with record-breaking performances in few-step language sequence generation.

2 Related Work

Continuous-space diffusion distillation. Diffusion models (Song et al., 2021b; Ho et al., 2020; Sohl-Dickstein et al., 2015) added Gaussian noises to data with a forward diffusion process and then learn the backward stochastic differential equation (SDE) with so-called score neural networks. Early acceleration reduced the steps of iterative samplers using training-free high-order ODE solvers (Song et al., 2021a; Lu et al., 2022; Karras et al., 2022; Xue et al., 2023). However, solver-based acceleration suffered from discretization error in less than 10 generation steps. Luhman & Luhman

(2021) and Salimans & Ho (2022) first introduce training the few-step student models by learning the consecutive trajectory mapping of diffusion solvers. The seminal work of Song et al. (2023) and Luo et al. (2023b) opens the one-step generation of diffusion models through different distillation principles. Specifically, consistency models (Song et al., 2023) distill the few-step models with the trajectory consistency principle, resulting in strong performances (Song & Dhariwal, 2024; Lu & Song, 2025; Geng et al., 2025b; Luo et al., 2023a; Kim et al., 2024; Geng et al., 2025a). Diff-Instruct(Luo et al., 2023b) introduces the distribution matching principle that distills one-step generative models, resulting in leading efficient image generative models (Wang et al., 2025; Luo et al., 2025a; 2024; Luo, 2025; Xu et al., 2025; Yin et al., 2024b;a; Xie et al., 2024; Fan et al., 2023; Zhou et al., 2024; Huang et al., 2024; Luo et al., 2025b;c). It is worth noting that, in this paper, we are inspired by the distribution-matching principle introduced by Diff-Instruct, but in the context of challenging discrete language sequence generation.

Discrete diffusion for language. Discrete modeling began with categorical transitions via argmax flows (Hoogeboom et al., 2021). Early dLLMs (He et al., 2023; Sahoo et al., 2024) essentially train BERT (Devlin et al., 2019) under a hierarchy of noise levels, motivated by a scaling-based rationale for generation. Later, Austin et al. (2021) introduced structured transition matrices and auxiliary cross-entropy training. For language modeling, two principal families of dLLMs have emerged. Masked diffusion models (MDMs) (Lou et al., 2024; Sahoo et al., 2025; Shi et al., 2024; Ou et al., 2025) treat the forward process as progressive masking of tokens. At the highest noise level, every token is replaced by a special `[MASK]` symbol; the reverse process gradually unmask tokens. Once a token is unmasked, it remains fixed, but many tokens can be denoised simultaneously. Uniform-state diffusion models (USDMs) (Austin et al., 2021; Gulrajani & Hashimoto, 2023; Schiff et al., 2025) instead corrupt tokens by sampling from a uniform distribution over the vocabulary. Consequently, each token can change multiple times during sampling, enabling self-correction and strong theoretical connections to continuous Gaussian diffusion. Subsequent work both elucidates training and decoding challenges (Zheng et al., 2025; Kim et al., 2025) and scales models to the billion-parameter range (Nie et al., 2025; Zhu et al., 2025a; Song et al., 2025). Our distillation framework, which adopts ideas from continuous-time IKL, is specifically designed for dLLMs to ensure tractable training and inference.

3 Preliminary

This section introduces the notations of the masked diffusion model (Sec. 3.1) and the integral KL divergence (Sec. 3.2), for preparation of the DiDi-Instruct algorithm in Section 4. We denote a vocabulary of size K. A one-hot vector represents each element in this vocabulary. The set of all such vectors is defined as $\mathcal{V} = \{\mathbf{x} \in \{0,1\}^K \mid \sum_{j=1}^K x_j = 1\}$. Within this vocabulary, the K-th token is reserved for the special `[MASK]` token. Its one-hot encoding is the standard basis vector $\mathbf{m}$, where the K-th element is 1 and all others are 0. To generate a sequence $\mathbf{x}^{1:L} = (\mathbf{x}^1, \dots, \mathbf{x}^L) \in \mathcal{V}^L$ of length L, we assume that the process factorizes across positions, i.e., each index $\ell \in \{1, \dots, L\}$ evolves independently. We adopt a continuous time index $t \in [0,1]$ with a monotonically decreasing noise schedule $\alpha_t \in [0,1]$ satisfying $\alpha_0 \approx 1$ and $\alpha_1 \approx 0$. Moreover, we denote K-simplex as Δ^K.

3.1 Masked Diffusion Models

Forward process. In the forward process, once a token transitions to the masked state $\mathbf{m}$, it never reverts. With independent per-token transitions, $\mathbf{x}^{1:L}$ converges to fully masked sequence $(\mathbf{m}, \dots, \mathbf{m})$ with probability 1 as $t \to 1$ (Sahoo et al., 2024). In the forward corruption process, the clean token $\mathbf{x} \in \mathcal{V}$ transitions to a latent state $\mathbf{z}_t \in \mathcal{V}$ at time $t \in (0,1]$. This absorbing state process ensures that $\mathbf{z}_t$ either remains the original token $\mathbf{x}$ with probability α_t or transitions into $\mathbf{m}$ with probability $1 - \alpha_t$. The forward corruption kernel is formalized as:

$$\mathcal{Q}(\mathbf{z}_t \mid \mathbf{x}) = \mathrm{Cat}\big(\mathbf{z}_t; \alpha_t \mathbf{x} + (1-\alpha_t)\mathbf{m}\big), \tag{1}$$

where $\mathrm{Cat}\big(\mathbf{z}; \boldsymbol{\pi}\big)$ denotes a Categorical distribution over the one-hot vector $\mathbf{z} \in \mathcal{V}$ with $\boldsymbol{\pi} \in \Delta^K$. For $s < t$, the exact posterior takes two cases:

$$\mathcal{Q}(\mathbf{z}_s \mid \mathbf{z}_t, \mathbf{x}) = \begin{cases} \mathrm{Cat}(\mathbf{z}_s; \mathbf{z}_t), & \mathbf{z}_t \neq \mathbf{m}, \\ \mathrm{Cat}\left(\mathbf{z}_s; \dfrac{(1-\alpha_s)\mathbf{m} + (\alpha_s - \alpha_t)\mathbf{x}}{1-\alpha_t}\right), & \mathbf{z}_t = \mathbf{m}. \end{cases} \tag{2}$$

This result is direct: an unmasked token implies a deterministic past, while a masked token implies a past probabilistic mixture governed by the noise schedule.

Backward process. The reverse process of MDMs iteratively reconstructs the original uncorrupted sequence from the noised input by replacing masked tokens. This is achieved through a learned neural network $\mathbf{p}_\theta : \mathcal{V} \times [0,1] \to \Delta^K$ that predicts the clean token $\mathbf{x}$ from its corrupted version $\mathbf{z}_t$ at time t. The output of $\mathbf{p}_\theta(\mathbf{z}_t, t)$ satisfies that (i) the predicted distribution forms valid class probabilities that sum to 1, (ii) the prediction $\mathbf{p}_\theta(\mathbf{z}_t, t)$ assigns zero probability mass to the `[MASK]` token, and (iii) once a token is unmasked, its state remains unchanged in subsequent steps.

When applied to sequences $\mathbf{x}^{1:L}$, the denoising process is assumed to be token-wise independent. A network $\mathbf{p}_\theta^{1:L}(\mathbf{z}_t^{1:L}, t)$ predicts the distribution for the entire sequence, where $\mathbf{p}_\theta^\ell(\mathbf{z}_t^{1:L}, t)$ specifies the distribution for the ℓ-th token. The reverse transition from $\mathbf{z}_t$ to $\mathbf{z}_s$ is defined by $\mathcal{P}_\theta(\mathbf{z}_s \mid \mathbf{z}_t) = \mathcal{Q}(\mathbf{z}_s \mid \mathbf{z}_t, \mathbf{x} = \mathbf{p}_\theta(\mathbf{z}_t, t))$. We extend this to sequences by factorizing the process over the sequence length L: $\mathcal{P}_\theta(\mathbf{z}_s^{1:L} \mid \mathbf{z}_t^{1:L}) = \prod_{\ell=1}^{L} \mathcal{Q}(\mathbf{z}_s^\ell \mid \mathbf{z}_t^\ell, \mathbf{p}_\theta^\ell(\mathbf{z}_t^{1:L}, t))$ In continuous time, minimizing the negative evidence lower bound (NELBO) provides a tractable training objective (Shi et al., 2024):

$$\mathcal{L}_{\text{NELBO}}^{\infty^L}(\theta) = \mathbb{E}_q \left[\int_0^1 \frac{\mathrm{d}\alpha_t}{\mathrm{d}t} \frac{1}{1-\alpha_t} \Big[\sum_{\ell : \mathbf{z}_t^\ell = \mathbf{m}} \mathbf{x}^{\ell\top} \log \mathbf{p}_\theta^\ell(\mathbf{z}_t^{1:L}, t) \Big] \mathrm{d}t \right]. \tag{3}$$

3.2 Integral KL divergence

Let $\mathbf{q}_\theta(\mathbf{z_t}, t)$ denote the forward marginal and $\mathbf{q}_\nu(\mathbf{z_t}, t)$ the student marginal parameterized by ν[1]. To distill knowledge from $\mathbf{q}_\theta$ to $\mathbf{q}_\nu$ across different noise levels $t \in [0,1]$, we minimize the integral Kullback-Leibler (IKL) divergence (introduced in Luo et al. (2023b)), which aggregates the discrepancy between $\mathbf{q}_\nu$ and $\mathbf{q}_\theta$:

$$\mathcal{D}_{\text{IKL}}\big(\mathbf{q}_\nu \| \mathbf{q}_\theta\big) := \int_0^1 \omega(t) \mathrm{KL}\big(\mathbf{q}_\nu \| \mathbf{q}_\theta\big) \mathrm{d}t = \int_0^1 \omega(t) \mathbb{E}_{\mathbf{z}_t \sim \mathbf{q}_\nu} \left[\log \frac{\mathbf{q}_\nu(\mathbf{z}_t, t)}{\mathbf{q}_\theta(\mathbf{z}_t, t)} \right] \mathrm{d}t, \tag{4}$$

where $\omega(t)$ is a positive weighting function[2]. Intuitively, both $\mathbf{q}_\theta$ and $\mathbf{q}_\nu$ are full-support for any $t \in [0,1]$ due to diffusion, making each term $\mathrm{KL}(\mathbf{q}_\nu \| \mathbf{q}_\theta)$ finite and smooth. IKL integrates this continuum of reliable comparisons, ensuring the student learns the teacher's complete denoising behavior, which leads to more stable and effective training than only matching the final output.

It is clear that if we can minimize the IKL divergence (4) between a few-step student language model and a pre-trained dLLM teacher model, the student is able to match the teacher's generation ability with much improved efficiency.

4 Methodology

The extension of IKL to distill MDMs presents several inherent challenges. This section details the comprehensive solution via advances in objective design, training stability, and inference efficiency.

4.1 Discrete Diffusion Divergence Instruction

A primary challenge in distilling MDMs is the discrete nature of the state space. The gradient formulation in Luo et al. (2023b) relies on differentiating through $\mathbf{z}_t$. In MDMs, this forward process involves non-differentiable operations (e.g., $\arg\max$), which renders this gradient inapplicable (Zekri & Boullé, 2025). Zhu et al. (2025b) proposed a proxy model to approximate $\mathbf{z}_t$ and achieved promising results in text-to-image tasks. However, its effectiveness in text generation remains underexplored. We instead take inspiration from the policy gradient (Schulman et al., 2017; Fan et al., 2023) to adapt a mathematically rigorous solution. By decomposing the gradient of the objective as a score function weighted by a log-density ratio, we derive the following gradient estimation.

[1]Unless otherwise specified, $\mathbf{q}_\theta$ and $\mathbf{q}_\nu$ denote the full collections $\mathbf{q}_\theta^{1:L}(\mathbf{z_t}, t)$ and $\mathbf{q}_\nu^{1:L}(\mathbf{z_t}, t)$.

[2]To simplify notation, we abbreviate $\mathbf{z}_t^{1:L}$ as $\mathbf{z}_t$, and $\mathbf{x}^{1:L}$ as $\mathbf{x}$ unless stated otherwise.

Theorem 4.1 (Score-Function Identity). *Let the objective $\mathcal{L}(\nu)$ be the weighted integral of KL divergences between student and teacher marginals, $\mathbf{q}_\nu$ and $\mathbf{q}_\theta$. The gradient of the objective admits:*

$$\nabla_\nu \mathcal{L}(\nu) = \mathbb{E}_{t\sim\pi(t),\mathbf{x}\sim\mathbf{p}_\nu,\mathbf{z}_t\sim\mathcal{Q}}\left[\frac{\omega(t)}{\pi(t)} \cdot R(\mathbf{z}_t, t) \cdot \nabla_\nu \log \mathbf{p}_\nu(\mathbf{z}_t = \mathbf{m}, t = 1)\right], \tag{5}$$

where the expectation over time t is sampled from distribution $\pi(t)$, and $R(\mathbf{z}_t, t) := \log \mathbf{p}_\nu(\mathbf{z}_t, t) - \log \mathbf{p}_\theta(\mathbf{z}_t, t)$ denotes the reward (log-density ratio between student and teacher) at $\mathbf{z}_t$.

The proof is deferred to Appendix C. Theorem 4.1 shows that the IKL gradient admits a score-function form that does not differentiate through the discrete sampling path. We sample $\mathbf{x} \sim \mathbf{p}_\nu(\mathbf{z}_t = \mathbf{m}, t = 1)$, $\mathbf{z}_t \sim \mathcal{Q}$ and differentiate only $\log \mathbf{p}_\nu(\mathbf{z}_t = \mathbf{m}, t = 1)$. The reward $R(\mathbf{z}_t, t)$ then weights this score to guide distillation toward matching the teacher's marginal at time t.

4.2 Density-Ratio Estimation

The reward is a mathematically natural choice for distribution matching. However, both $\log \mathbf{q}_\nu$ and $\log \mathbf{q}_\theta$ are intractable to compute directly. To overcome this, we avoid modeling the individual densities and instead approximate their ratio. Inspired by Wang et al. (2025) in continuous-time diffusion, we train an auxiliary discriminator specifically for this it. See Appendix D for details.

Lemma 4.2 (Density Ratio Representation). *Let $D_\lambda : \mathcal{V}^L \times [0, 1] \to (0, 1)^L$ be a parameterized discriminator to distinguish samples generated by $\mathbf{q}_\nu$ and $\mathbf{q}_\theta$. For the optimal discriminator $D_{\lambda^\star}(\mathbf{z}_t, t)$, the density ratio can be expressed as*

$$\frac{\mathbf{q}_\nu(\mathbf{z}_t, t)}{\mathbf{q}_\theta(\mathbf{z}_t, t)} = \frac{D_{\lambda^\star}(\mathbf{z}_t, t)}{1 - D_{\lambda^\star}(\mathbf{z}_t, t)}.$$

Following Lemma 4.2, we subsequently construct a tractable reward signal $R(\mathbf{z}_t, t)$ by aggregating the log-density ratios over all masked positions. With M denoting the number of masked tokens in the sequence, the reward can be estimated from the discriminator network D_λ as follows:

$$R(\mathbf{z}_t, t) = \frac{1}{M} \sum_{\ell, \mathbf{z}_t^\ell = \mathbf{m}} \log \frac{D_\lambda^\ell(\mathbf{z}_t, t)}{1 - D_\lambda^\ell(\mathbf{z}_t, t)}, \tag{6}$$

where $D_\lambda^\ell(\cdot, \cdot)$ denotes the ℓ-th element of $D_\lambda(\cdot, \cdot)$. This procedure provides a tractable and computable reward signal for our objective according to D_λ.

4.3 Grouped Reward Normalization

While the reward estimator in (6) is tractable, its direct use in score-function gradients can exhibit high variance (Williams, 1992). We therefore adopt Group Relative Policy Optimization (Shao et al., 2024; Team et al., 2025) to standardize rewards within each mini-batch. Given a mini-batch of size G, we draw $\{t_i\}_{i=1}^G$ from $\pi(t)$, estimate $\{\mathbf{z}_{t_i}\}_{i=1}^G$ and $\{R(\mathbf{z}_{t_i}, t_i)\}_{i=1}^G$ correspondingly[3]. The final stabilized reward $\widetilde{R}_i$ is obtained by normalizing R_i with the mean μ_g and variance σ_g^2:

$$\widetilde{R}_i = \frac{R_i - \mu_g}{\sigma_g + \epsilon}, \qquad \text{where } \mu_g = \tfrac{1}{G}\sum_{i=1}^G R_i, \text{ and } \sigma_g^2 = \tfrac{1}{G}\sum_{i=1}^G (R_i - \mu_g)^2. \tag{7}$$

Here, a small constant $\epsilon > 0$ is used for numerical stability. By substituting this stabilized reward $\tilde{R}_i$ for the raw reward R_i in our gradient estimator, we arrive at a more robust update rule.

4.4 The Proposed Algorithms

Adversarial reward estimation. Following Wang et al. (2025), we train a parameterized discriminator D_λ to separate corrupted samples from $\mathbf{q}_\nu$ and $\mathbf{q}_\theta$ at time t. For each mini-batch index i, we start from the all masked sequence to sample $\mathbf{x} \sim \mathbf{p}_\nu$ and $\mathbf{x}' \sim \mathbf{p}_\theta$. We then corrupt both to the

[3]For clarity and concision, we denote $R(\mathbf{z}_{t_i}^{1:L}, t_i)$ as R_i, and $D_\lambda(\mathbf{z}_{t_i}^{1:L}, t_i)$ as D_i throughout.

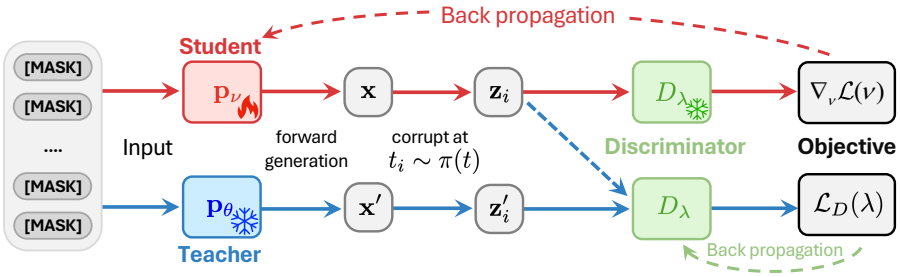

Figure 2: The pipeline of DiDi-Instruct (Algorithm 1). Given a fully masked input $\mathbf{z}_t$ ($t = 1$), both the student $\mathbf{p}_\nu$ and the teacher (assistant model) $\mathbf{p}_\theta$ produce clean samples $\mathbf{x}$ and $\mathbf{x}'$, which are corrupted at $t_i \sim \pi(t)$ to form $\mathbf{z}_i$ and $\mathbf{z}_i'$. The discriminator D_λ is trained to classify these outputs, while its reward signal (6) enables the gradient update (5) for the student. The red line represents the gradient flow for the student's update step, and the blue line represents the gradient flow for the auxiliary model's update step.

same noise level t_i via the forward process (2) to obtain $\mathbf{z}_i \sim \mathcal{Q}$ and $\mathbf{z}_i' \sim \mathcal{Q}$. The discriminator is optimized with balanced binary cross-entropy:

$$\mathcal{L}_D(\lambda) = -\frac{1}{G}\sum_{i=1}^{G}\Big[\log D_\lambda(\mathbf{z}_i, t_i) + \log(1 - D_\lambda(\mathbf{z}_i', t_i))\Big]. \tag{8}$$

At optimality $\lambda^\star$, $D_{\lambda^\star}(\mathbf{z}_t, t) \approx \frac{\mathbf{q}_\nu(\mathbf{z}_t,t)}{\mathbf{q}_\nu(\mathbf{z}_t,t)+\mathbf{q}_\theta(\mathbf{z}_t,t)}$, which indicates that $\operatorname{logit} D_{\lambda^\star}$ can provide a tractable estimate of the log–density ratio used in (6).

Score function decomposition. While sampling directly from $\mathbf{z}_t{=}\mathbf{m}$ ($t = 1$) to $\mathbf{x}$ is suitable for one-step generators, it induces collapse in multi-step regimes: the student never conditions on intermediate states and tends toward low-entropy, mode-seeking behavior (Zhu et al., 2025b). To expose the student to intermediate corruption levels, we approximate the score from (5) by decomposing it at a randomly sampled time $t_i \sim \pi(t)$ and its corresponding state $\mathbf{z}_i$, which gives:

$$\nabla_\nu \log \mathbf{p}_\nu(\mathbf{z}_t{=}\mathbf{m}, t = 1) \approx \nabla_\nu \log \mathcal{P}_\nu(\mathbf{z}_i|\mathbf{z}_t{=}\mathbf{m}) + \nabla_\nu \log \mathbf{p}_\nu(\mathbf{z}_i, t_i). \tag{9}$$

Training with the split score (9) exposes the student to a distribution of intermediate states and mitigates entropy collapse, while remaining compatible with the IKL estimator.

End-to-end training algorithm. The DiDi-Instruct training procedure alternates between updating the discriminator D_λ and the student $\mathbf{p}_\nu$, which is summarized in Algorithm 1 and visualized in Figure 2. Each iteration consists of two phases: first, the discriminator is updated to better distinguish between corrupted samples from $\mathbf{p}_\theta$ and $\mathbf{p}_\nu$; we apply (5) with the normalized reward and the decomposed score to stabilize gradient updates. This alternating scheme yields a robust few-step generator that closely matches the teacher's marginals over different corrupt levels.

Algorithm 1 DiDi-Instruct

for each training step **do**
  Sample $t_i \sim \pi(t)$, and
  $\mathbf{x} \sim \mathbf{p}_\nu(\mathbf{z}_t, t = 1)$, $\mathbf{x}' \sim \mathbf{p}_\theta(\mathbf{z}_t, t = 1)$.
  Corrupt $\mathbf{x}$, $\mathbf{x}'$ to partially masked $\mathbf{z}_i$, $\mathbf{z}_i'$.
  Update discriminator D_λ with (8).
  Update student $\mathbf{p}_\nu$ with (5)-(7), and (9).
return student $\mathbf{p}_\nu$ and discriminator D_λ

Algorithm 2 RGAS

Initialize $\mathbf{z}_N \leftarrow (\mathbf{m}, \ldots, \mathbf{m})$.
for $n = N, \ldots, 1$ **do**
  if n in early stage **then**
    Sample $\mathbf{z}_{n-1}$ with (10) ($h > 0$, $M = 1$).
  else
    Sample $\mathbf{z}_{n-1}$ with (10)-(11) ($h = 0$, $M \geq 1$).
Set $\mathbf{x} = \mathbf{z}_0$ and **return** sequence $\mathbf{x}$

Reward-guided ancestral sampler. We further propose a decoding strategy that leverages the trained discriminator to guide ancestral sampling (AS) (Shi et al., 2024; Zheng et al., 2025). Starting from a fully masked sequence $\mathbf{z}_N = (\mathbf{m}, \ldots, \mathbf{m})$ at $t_N = 1$, the procedure generates samples by iteratively denoising from t_n to t_{n-1} for $n = N, \ldots, 1$, following the student's backward distribution $\mathbf{p}_\nu(\mathbf{z}_{n-1}|\mathbf{z}_n)$. Specifically, for $\mathbf{z}_n^\ell \in \mathcal{V}$ ($\ell = 1, \cdots, L$) we consider a tilted transition:

$$\mathbf{z}_{n-1}^\ell = \begin{cases} \mathbf{z}_n^\ell, & \text{if } \mathbf{z}_n^\ell \neq \mathbf{m}, \\ \sim \mathrm{Cat}\left[\dfrac{(1-\alpha_{n-1})\mathbf{m} + (\alpha_{n-1} - \alpha_n)\,\mathbf{p}_\nu^\ell(\mathbf{z}_n, t_n)}{1-\alpha_n}\right], & \text{if } \mathbf{z}_n^\ell = \mathbf{m}, \end{cases} \tag{10}$$

where the logits for unmasked tokens are augmented with the reward gradient:

$$\mathbf{p}_\nu^\ell(\mathbf{z}_n, t_n) = \begin{cases} \mathbf{e}_{z_n^\ell}, & \text{if } \mathbf{z}_n^\ell \neq \mathbf{m}, \\ \big[\mathrm{softmax}\big(\boldsymbol{\mu}_\nu^\ell(\mathbf{z}_n, t_n) + h\nabla R(\mathbf{z}_n, t_n)\big), 0\big], & \text{if } \mathbf{z}_n^\ell = \mathbf{m}. \end{cases}$$

Here, $\mathbf{e}_{z_n^\ell} \in \mathcal{V}$ denotes the one-hot vector with a 1 at index z_n^ℓ and 0 elsewhere, and h denotes the tilting scale. RGAS adjusts h and the number of candidates M across the denoising process: For early steps ($t_n \approx 1$), we use *gradient tilting* ($h > 0, M = 1$) to steer global structure toward high-reward regions. For late steps ($t_n \approx 0$), we switch to *multi-candidate re-ranking* ($h = 0, M > 1$): we draw M candidates $\{\mathbf{z}_{n-1}^{(m)}\}_{m=1}^M$[4] from (10) with $h = 0$, then select $\mathbf{z}_{n-1}$ following:

$$\mathbf{z}_{n-1} \sim \mathrm{Cat}\left[\frac{\exp\left[R(\mathbf{z}_{n-1}^{(m)}, t_{n-1})\right]}{\sum_{m=1}^M \exp\left[R(\mathbf{z}_{n-1}^{(m)}, t_{n-1})\right]}\right]. \tag{11}$$

To decompose score estimation in (9) and sample the intermediate state $\mathbf{z}_i$ at t_i, we set $h = 0$ and $M = 1$ to mitigate the risk of reward hacking (Skalse et al., 2022; Gao et al., 2023). The complete procedure is detailed in Algorithm 2.

5 Experiments

Our experiments are designed to distill a pre-trained teacher model into an efficient few-step student generator with DiDi-Instruct. All models are trained on OWT (Gokaslan et al., 2019). Following standard practices (Sahoo et al., 2025), we tokenize the corpus using the GPT-2 tokenizer, pack sequences to a context length of 1024, and hold out the last 100,000 documents for validation.

Our teacher is a 169M parameter MDLM with a Diffusion Transformer (Peebles & Xie, 2023) (12 layers, 12 attention heads, 768 hidden dimension). We pre-trained this model from scratch for 1024 NFEs, achieving a perplexity of 38.53 and an entropy of 5.22. The student model shares an identical architecture to ensure a fair comparison. The reward model is a 131M parameter network based on the same backbone, but with a new randomly initialized classification head. This head consists of two linear layers with spectral normalization and a SiLU activation function. During distillation, the reward model and the student generator are trained adversarially in an alternating fashion.

The distillation process runs 10,000 iterations using the AdamW optimizer with a learning rate of 10^{-6} and no warm-up. The teacher model was pre-trained on 8 NVIDIA H100 GPUs. In contrast, our DiDi-Instruct distillation is highly efficient, requiring only a single H100 GPU. All training procedures leverage `bfloat16` for acceleration. For more details, please refer to Appendix E.

5.1 Experimental Result and Analysis

We present the empirical results of our proposed DiDi-Instruct. The results show that DiDi-Instruct not only achieves state-of-the-art performance in sample quality, particularly in the few-step generation regime, but also offers substantial improvements in training and inference efficiency.

Generation results. We evaluate generative quality by sampling text from the distilled student from 8 to 128 NFEs and measuring GPT-2 Large generative PPL and average sequence entropy. Figure 1 shows that DiDi-Instruct consistently outperforms all baselines in PPL across all sampling steps. Notably, with only **16** NFEs, our model's PPL already surpasses that of the 1024-step teacher model. At 1024 NFEs, DiDi-Instruct achieves a final PPL of 15.62, a reduction of over 24% compared to the strongest baseline. These performance gains are achieved with a negligible loss in diversity; the generative entropy from 8 to 128 NFEs is 5.17, 5.21, 5.18, 5.15, and 5.15, respectively (e.g., 1024-step teacher model is 5.22), indicating sample diversity is well-preserved.

Efficiency-performance tradeoff. DiDi-Instruct offers substantial computational advantages in both training and inference. Our single-round distillation framework completes training in around one H100 GPU hour, in contrast to the 20+ GPU hours required by multi-round methods (Sahoo et al., 2024; Deschenaux & Gulcehre, 2025). During inference, RGAS provides a superior perplexity-latency trade-off. As shown in Figure 3, it achieves a PPL of 23.54 at a latency of 1.31

[4]For simplicity, we denote the m-th candidate of $\mathbf{z}_{t_n}^{1:L}$ as $\mathbf{z}_n^{(m)}$.

s/sequence (32 NFEs), whereas the ancestral sampler only reaches a PPL of 44.54 at a similar latency (0.71 s/sequence, 32 NFEs).

Zero-shot likelihood. Distilling a model for few-step generations can risk degrading its core language understanding and causing mode collapse. We assess this trade-off by evaluating zero-shot PPL on seven out-of-domain corpora (PTB, WikiText, LM1B, LAMBADA, AG News, PubMed, and ArXiv), with results presented in Table 7. The evaluation shows that DiDi-Instruct strikes an effective balance. It consistently improves upon the DUO distilled baseline, validating our superior distillation process. Crucially, while it trails some of the full, undistilled models as expected, it maintains a highly competitive level of performance. These results confirm that DiDi-Instruct achieves its goal of sampling efficiency while preserving robust zero-shot generalization.

Sample quality. Appendix F.7 presents full text excerpts with per-sample metrics (PPL and entropy) for the 1024-step teacher and DiDi-Instruct students from 8 to 128 NFEs, which confirms a clear improvement in narrative quality. The 8-step student exhibits expected repetition, a common artifact of rapid sampling. However, this issue is resolved by 16 NFEs, at which point the student model already surpasses the 1024-step teacher in paragraph-level coherence and topic adherence. This ability to construct focused and specific narratives strengthens as NFEs increase to 128, indicating that DiDi-Instruct not only preserves fluency but actively enhances the model's ability to generate structured, coherent text.

5.2 Ablation Studies

We conduct comprehensive ablation studies to validate the contribution of each component in DiDi-Instruct. We perform two types of analyses: a cumulative study (Table 1) that progressively adds techniques to a baseline, showing their synergistic benefits, and a leave-one-out study (Table 2) that removes individual components to confirm their necessity. Implementation details can refer to F.3. Our ablation studies reveal distinct roles for each component in our framework. We find that a two-step score decomposition is a non-negotiable cornerstone, providing essential stability without which the model fails to train. The most significant performance gains are driven by coupling t in (5) and (9). Loss shaping with $\omega(t)$ and $\pi(t)$ further smooths optimization (especially around 16 NFEs, while effects at very small/large budgets are modest). Finally, we identify two budget-dependent components: Regularization is crucial for stability at very few steps ($\leq$ 8 NFEs) but detrimental at higher budgets, while Guided Inference boosts quality at low NFEs and enhances diversity at high NFEs. These findings highlight a hierarchy of importance and nuanced interactions between the techniques. More details are presented in Appendix F.4.

Table 1: Cumulative ablation study. We start from a baseline model and *progressively add* the listed tricks on top of the previous row (top→bottom). Metrics are reported as PPL↓ and Entropy↑ for different sampling steps. The teacher with 1024 steps yields entropy 5.22.

Configurations	8 Steps		16 Steps		32 Steps		64 Steps		128 Steps	
	PPL↓	Entropy↑	PPL↓	Entropy↑	PPL↓	Entropy↑	PPL↓	Entropy↑	PPL↓	Entropy ↑
Baseline (no tricks)	803.922	5.85	311.450	5.76	174.789	5.70	113.112	5.61	96.649	5.59
+ Score Decompose	667.830	5.83	289.720	5.76	165.809	5.70	105.880	5.61	89.350	5.59
+ Coupled Time t	101.019	5.16	75.188	5.46	48.441	5.35	35.833	5.37	30.574	5.33
+ $\omega(t)$ Correction	94.955	5.21	75.607	5.22	31.651	5.20	25.271	5.16	20.980	5.12
+ $\pi(t)$ Weighting	92.100	5.15	43.997	5.17	32.276	5.21	26.079	5.21	21.377	5.13
+ Regularization	88.274	5.11	43.980	5.16	28.444	5.12	21.946	5.06	18.325	5.00
+ Guided Inference	62.236	5.17	38.188	5.21	24.971	5.18	21.905	5.15	18.446	5.15

Table 2: Leave-one-out ablation study. Each row shows performance *without (w/o)* one trick while keeping the others unchanged. Metrics are PPL↓ and Entropy↑ over different sampling steps. The lowest PPL in each step column is underlined.

Configurations	8 Steps		16 Steps		32 Steps		64 Steps		128 Steps	
	PPL↓	Entropy↑	PPL↓	Entropy↑	PPL↓	Entropy↑	PPL↓	Entropy↑	PPL↓	Entropy ↑
w/o Score Decompose	33584	6.77	28962	6.77	23134	6.75	14634	6.64	7983	6.51
w/o Coupled Time t	360.75	5.42	159.43	5.43	94.859	5.45	64.639	5.35	51.121	5.39
w/o $\omega(t)$ Correction	82.489	5.12	41.034	5.13	30.313	5.09	25.125	5.04	18.806	5.02
w/o $\pi(t)$ Weighting	69.656	5.22	40.499	5.17	25.799	5.15	21.503	5.16	19.616	5.14
w/o Regularization	84.594	5.20	30.994	5.22	23.603	5.20	19.609	5.18	17.499	5.17
w/o Guided Inference	88.274	5.11	43.980	5.16	28.444	5.12	21.946	5.06	18.325	5.00
Baseline (with all tricks)	62.236	5.17	38.188	5.21	24.971	5.18	21.905	5.15	18.446	5.15

5.3 Model Scaling Up

We perform an incremental scaling experiment to 424M parameters while holding the training and inference pipeline fixed to the 169M configuration. The teacher model is a pre-trained 424M MDLM, which is a DiT-style transformer with a hidden size of 1024, 24 blocks, and 16 attention heads (check Table 3 for detailed network structure). The student is distilled on a single H100 using the same data, masking policy, optimizer, and schedule. To balance the training, we also scale the reward discriminator to 373M and employ a deeper network with more trainable parameters.

The generative PPL and entropy are shown in Figure 4, and we report improvements relative to the teacher at matched NFEs here: from 8 to 1024 NFEs, DiDi-Instruct yields large and consistent PPL reductions of 88.5% (8 NFEs), 87.7% (16 NFEs), 85.3% (32 NFEs), 82.7% (64 NFEs), and 79.0% (128 NFEs). Moreover, with only 16 steps, the distilled model achieves a perplexity of 32.79, marking an 11.4% improvement over the 1024-step teacher baseline. Entropy remains comparable throughout, indicating that diversity is preserved while accuracy improves. These results confirm that the quality–efficiency advantages of DiDi-Instruct persist at increasing scale with minimal procedural changes. For more details, please refer to Appendix F.5.

5.4 Another Application: Protein Sequences Generation

To demonstrate the applicability of our distillation framework beyond natural language generation, we apply DiDi-Instruct to unconditional protein sequence generation. We adopt the Diffusion Protein Language Model (DPLM) (Wang et al., 2024) , pretrained on UniRef50 (Suzek et al., 2014) with 150M parameters, as the teacher model and distill it into a few-step student generator. Following (Wang et al., 2024), we evaluate sequence quality using the predicted local distance difference test (pLDDT) score, which reflects structural plausibility and foldability. The distilled model retains the teacher's ability to generate variable-length protein sequences while substantially reducing inference cost, as shown in Figure 5.

Our results demonstrate that the distilled student consistently achieves superior pLDDT scores across generation settings ranging from 8 to 512 steps. Compared to the teacher model, the student not only preserves the ability to generate variable-length protein sequences but also enhances structural quality in most cases. Moreover, our model surpasses the high-confidence threshold (pLDDT > 70) with as few as 8–32 steps, while the teacher requires substantially more NFEs to reach a comparable level. These results highlight that our distillation framework not only improves structural confidence but also delivers stable performance across sequence lengths and generation budgets. For experimental details on protein sequence distillation, please refer to Appendix F.6, where we also provide visualizations of low-confidence protein sequences generated by DPLM under fewer NFEs in Figure 6, along with high-quality examples generated by DiDi-Instruct across different NFEs in Figure 7.

6 Conclusion and Future Work

In this work, we introduced DiDi-Instruct, a training-based acceleration framework for fast language generation that distills a high-quality teacher into a few-step student. Our design targets three axes simultaneously: (i) *objective design* via a tractable policy–gradient update driven by a discriminator–estimated reward; (ii) *training stability* through score decomposition and grouped reward normalization; and (iii) *inference efficiency* through reward-guided ancestral sampling with gradient tilting and re-ranking. Experiments demonstrate strong gains in generation quality, large reductions in training/inference time, and competitive zero-shot generalization, corroborated by comprehensive cumulative and leave-one-out ablations.

We plan to scale DiDi-Instruct to billion-parameter models, which presents a practical challenge due to the memory requirements of concurrently maintaining the teacher, student, and discriminator. Nevertheless, our findings already establish a new state-of-the-art trade-off among comparable methods, with the student model excelling in quality at low NFEs and maintaining diversity at higher computational budgets. We posit that DiDi-Instruct offers a foundational recipe (principled objectives, training stability, and efficient guidance) for developing high-performance generative models.

Ethics Statement

All authors have read and commit to adhering to the ICLR Code of Ethics. This work uses only publicly available datasets and open-source models; no human subjects or sensitive personal data are involved, and Institutional Review Board approval is not required. We briefly discuss potential biases and fairness considerations of the datasets and algorithms in the main paper, and confirm that the work does not target discriminatory or harmful applications. Any potential conflicts of interest have been disclosed. We do not foresee misuse beyond the general risks associated with large language models, and we report methods and results transparently.

Reproducibility Statement

To facilitate reproduction, the main text and Appendix describe the model architecture, training procedures, and evaluation protocols, and the Appendix includes complete proofs of theoretical results. All datasets are publicly available, and data preprocessing steps are documented in the submission. We include training and evaluation code and brief instructions in the supplementary material and refer to the relevant sections rather than repeating details here.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774*, 2023.

Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg. Structured Denoising Diffusion Models in Discrete State-Spaces. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:17981–17993, 2021.

Nicholas M Boffi, Michael S Albergo, and Eric Vanden-Eijnden. Flow Map Matching. *arXiv preprint arXiv:2406.07507*, 2024.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:1877–1901, 2020.

Andrew Campbell, Joe Benton, Valentin De Bortoli, Thomas Rainforth, George Deligiannidis, and Arnaud Doucet. A Continuous Time Framework for Discrete Denoising Models. *Advances in Neural Information Processing Systems (NeurIPS)*, 35:28266–28279, 2022.

Tianyu Chen, Zhendong Wang, and Mingyuan Zhou. Diffusion Policies Creating a Trust Region for Offline Reinforcement Learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:50098–50125, 2024.

Ting Chen, Ruixiang Zhang, and Geoffrey Hinton. Analog Bits: Generating Discrete Data using Diffusion Models with Self-Conditioning. *Proc. of the International Conference on Learning Representation (ICLR)*, 2023.

Justin Deschenaux and Caglar Gulcehre. Beyond Autoregression: Fast LLMs via Self-Distillation Through Time. In *Proc. of the International Conference on Learning Representation (ICLR)*. OpenReview.net, 2025.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proc. of the Annual Meeting of the Association Computational Linguistics (ACL)*, pp. 4171–4186, 2019.

Sander Dieleman, Laurent Sartran, Arman Roshannai, Nikolay Savinov, Yaroslav Ganin, Pierre H Richemond, Arnaud Doucet, Robin Strudel, Chris Dyer, Conor Durkan, et al. Continuous Diffusion for Categorical Data. *arXiv preprint arXiv:2211.15089*, 2022.

Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. DPOK: Reinforcement Learning for Fine-Tuning Text-to-Image Diffusion Models. *Advances in Neural Information Processing Systems (NeurIPS)*, 36:79858–79885, 2023.

Leo Gao, John Schulman, and Jacob Hilton. Scaling Laws for Reward Model Overoptimization. In *Proc. of the International Conference on Machine Learning (ICML)*, pp. 10835–10866. PMLR, 2023.

Zhengyang Geng, Ashwini Pokle, and J Zico Kolter. One-Step Diffusion Distillation via Deep Equilibrium Models. *Advances in Neural Information Processing Systems (NeurIPS)*, 36:41914–41931, 2023.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J. Zico Kolter, and Kaiming He. Mean Flows for One-Step Generative Modeling. *arXiv preprint arXiv:2505.13447*, 2025a.

Zhengyang Geng, Ashwini Pokle, Weijian Luo, Justin Lin, and J. Zico Kolter. Consistency Models Made Easy. In *Proc. of the International Conference on Learning Representation (ICLR)*. OpenReview.net, 2025b.

Aaron Gokaslan, Vanya Cohen, Ellie Pavlick, and Stefanie Tellex. OpenWebText Corpus. http://Skylion007.github.io/OpenWebTextCorpus, 2019.

Shansan Gong, Mukai Li, Jiangtao Feng, Zhiyong Wu, and Lingpeng Kong. DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models. In *Proc. of the International Conference on Learning Representation (ICLR)*, 2023.

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 27, 2014.

Jiatao Gu, Shuangfei Zhai, Yizhe Zhang, Lingjie Liu, and Josh Susskind. BOOT: Data-free Distillation of Denoising Diffusion Models with Bootstrapping. *arXiv preprint arXiv:2306.05544*, 2023.

Ishaan Gulrajani and Tatsunori B Hashimoto. Likelihood-Based Diffusion Language Models. *Advances in Neural Information Processing Systems (NeurIPS)*, 36:16693–16715, 2023.

Zhengfu He, Tianxiang Sun, Qiong Tang, Kuanning Wang, Xuanjing Huang, and Xipeng Qiu. DiffusionBERT: Improving Generative Masked Language Models with Diffusion Models. *Proc. of the Annual Meeting of the Association Computational Linguistics (ACL)*, 2023.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:6840–6851, 2020.

Emiel Hoogeboom, Didrik Nielsen, Priyank Jaini, Patrick Forré, and Max Welling. Argmax Flows and Multinomial Diffusion: Learning Categorical Distributions. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:12454–12465, 2021.

Zemin Huang, Zhengyang Geng, Weijian Luo, and Guo-jun Qi. Flow Generator Matching. *arXiv preprint arXiv:2410.19310*, 2024.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the Design Space of Diffusion-Based Generative Models. *Advances in Neural Information Processing Systems (NeurIPS)*, 35, 2022.

Dongjun Kim, Chieh-Hsin Lai, Wei-Hsiang Liao, Naoki Murata, Yuhta Takida, Toshimitsu Uesaka, Yutong He, Yuki Mitsufuji, and Stefano Ermon. Consistency Trajectory Models: Learning Probability Flow ODE Trajectory of Diffusion. In *Proc. of the International Conference on Learning Representation (ICLR)*, 2024.

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. *Proc. of the International Conference on Machine Learning (ICML)*, 2025.

Sehoon Kim, Karttikeya Mangalam, Suhong Moon, Jitendra Malik, Michael W. Mahoney, Amir Gholami, and Kurt Keutzer. Speculative Decoding with Big Little Decoder. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Yaniv Leviathan, Matan Kalman, and Yossi Matias. Fast Inference from Transformers via Speculative Decoding. In *Proc. of the International Conference on Machine Learning (ICML)*, 2023.

Xiang Li, John Thickstun, Ishaan Gulrajani, Percy S Liang, and Tatsunori B Hashimoto. Diffusion-LM Improves Controllable Text Generation. *Advances in Neural Information Processing Systems (NeurIPS)*, 35:4328–4343, 2022a.

Xiang Lisa Li, John Thickstun, Ishaan Gulrajani, Percy Liang, and Tatsunori B. Hashimoto. Diffusion-LM Improves Controllable Text Generation. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, 2022b.

Shanchuan Lin, Anran Wang, and Xiao Yang. SDXL-Lightning: Progressive Adversarial Diffusion Distillation. *arXiv preprint arXiv:2402.13929*, 2024.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. *Science*, 379(6637):1123–1130, 2023.

Sulin Liu, Juno Nam, Andrew Campbell, Hannes Stärk, Yilun Xu, Tommi Jaakkola, and Rafael Gómez-Bombarelli. Think While You Generate: Discrete Diffusion with Planned Denoising. *Proc. of the International Conference on Learning Representation (ICLR)*, 2025.

Xingchao Liu, Mei Zhen, et al. Learning to Generate and Transfer Data with Rectified Flow. *arXiv preprint arXiv:2209.03003*, 2022.

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete Diffusion Modeling by Estimating the Ratios of the Data Distribution. In *Proc. of the International Conference on Machine Learning (ICML)*, 2024.

Cheng Lu and Yang Song. Simplifying, Stabilizing and Scaling Continuous-time Consistency Models. In *Proc. of the International Conference on Learning Representation (ICLR)*. OpenReview.net, 2025.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps. *Advances in Neural Information Processing Systems (NeurIPS)*, 35, 2022.

Eric Luhman and Troy Luhman. Knowledge Distillation in Iterative Generative Models for Improved Sampling Speed. *arXiv preprint arXiv:2101.02388*, 2021.

Shanchuan Luo, Xiaohui Wang, Fan Zhang, et al. Latent Consistency Models: Synthesizing High-Resolution Images with Few-Step Inference. *arXiv preprint arXiv:2310.04378*, 2023a.

Weijian Luo. A Comprehensive Survey on Knowledge Distillation of Diffusion Models. *arXiv preprint arXiv:2304.04262*, 2023.

Weijian Luo. Diff-Instruct++: Training One-step Text-to-image Generator Model to Align with Human Preferences. *Trans. Mach. Learn. Res.*, 2025, 2025.

Weijian Luo, Tianyang Hu, Shifeng Zhang, Jiacheng Sun, Zhenguo Li, and Zhihua Zhang. Diff-Instruct: A Universal Approach for Transferring Knowledge From Pre-Trained Diffusion Models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023b.

Weijian Luo, Ting Zhu, Shifeng Zhang, Zhenguo Li, and Zhihua Zhang. One-Step Diffusion Distillation through Score Implicit Matching. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Weijian Luo, colin zhang, Debing Zhang, and Zhengyang Geng. David and Goliath: Small One-step Model Beats Large Diffusion with Score Post-training. In *Proc. of the International Conference on Machine Learning (ICML)*, 2025a.

Yihong Luo, Xiaolong Chen, Xinghua Qu, Tianyang Hu, and Jing Tang. You Only Sample Once: Taming One-Step Text-to-Image Synthesis by Self-Cooperative Diffusion GANs. In *Proc. of the International Conference on Learning Representation (ICLR)*. OpenReview.net, 2025b.

Yihong Luo, Tianyang Hu, Weijian Luo, Kenji Kawaguchi, and Jing Tang. Reward-Instruct: A Reward-Centric Approach to Fast Photo-Realistic Image Generation. *arXiv preprint arxiv.org/abs/2503.13070*, 2025c.

Chenlin Meng, Ruiqi Gao, Diederik P Kingma, Stefano Ermon, Jonathan Ho, and Tim Salimans. On Distillation of Guided Diffusion Models. *arXiv preprint arXiv:2210.03142*, 2022.

Meta AI. Introducing LLaMA: A foundational, 65-billion-parameter large language model. https://ai.meta.com/blog/large-language-model-llama-meta-ai/, 2023. Accessed: 2025-09-14.

Thuan Hoang Nguyen and Anh Tran. SwiftBrush: One-Step Text-to-Image Diffusion Model with Variational Score Distillation. *arXiv preprint arXiv:2312.05239*, 2023.

Shen Nie, Fengqi Zhu, Zebin You, Xiaolu Zhang, Jingyang Ou, Jun Hu, Jun Zhou, Yankai Lin, Ji-Rong Wen, and Chongxuan Li. Large Language Diffusion Models. *arXiv preprint arXiv:2502.09992*, 2025.

Xuefei Ning, Zinan Lin, Zixuan Zhou, Zifu Wang, Huazhong Yang, and Yu Wang. Skeleton-of-Thought: Prompting LLMs for Efficient Parallel Generation. In *Proc. of the International Conference on Learning Representation (ICLR)*, 2024.

Jingyang Ou, Shen Nie, Kaiwen Xue, Fengqi Zhu, Jiacheng Sun, Zhenguo Li, and Chongxuan Li. Your Absorbing Discrete Diffusion Secretly Models the Conditional Distributions of Clean Data. *Proc. of the International Conference on Learning Representation (ICLR)*, 2025.

Yong-Hyun Park, Chieh-Hsin Lai, Satoshi Hayakawa, Yuhta Takida, and Yuki Mitsufuji. Jump Your Steps: Optimizing Sampling Schedule of Discrete Diffusion Models. In *Proc. of the International Conference on Learning Representation (ICLR)*, 2024.

William Peebles and Saining Xie. Scalable Diffusion Models with Transformers. In *Proc. of the International Conference on Computer Vision (ICCV)*, pp. 4195–4205, 2023.

Yuxi Ren, Xin Xia, Yanzuo Lu, Jiacheng Zhang, Jie Wu, Pan Xie, Xing Wang, and Xuefeng Xiao. Hyper-SD: Trajectory Segmented Consistency Model for Efficient Image Synthesis. *arXiv preprint arXiv:2404.13686*, 2024.

Subham Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and Effective Masked Diffusion Language Models. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:130136–130184, 2024.

Subham Sekhar Sahoo, Justin Deschenaux, Aaron Gokaslan, Guanghan Wang, Justin Chiu, and Volodymyr Kuleshov. The Diffusion Duality. *Proc. of the International Conference on Learning Representation (ICLR)*, 2025.

Tim Salimans and Jonathan Ho. Progressive Distillation for Fast Sampling of Diffusion Models. *Proc. of the International Conference on Learning Representation (ICLR)*, 2022.

Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial Diffusion Distillation. *arXiv preprint arXiv:2311.17042*, 2023.

Yair Schiff, Subham Sekhar Sahoo, Hao Phung, Guanghan Wang, Sam Boshar, Hugo Dalla-torre, Bernardo P. de Almeida, Alexander Rush, Thomas Pierrot, and Volodymyr Kuleshov. Simple Guidance Mechanisms for Discrete Diffusion Models. *Proc. of the International Conference on Learning Representation (ICLR)*, 2025.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. *arXiv preprint arXiv:2402.03300*, 2024.

Jiaxin Shi, Kehang Han, Zhe Wang, Arnaud Doucet, and Michalis Titsias. Simplified and Generalized Masked Diffusion for Discrete Data. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:103131–103167, 2024.

Joar Skalse, Nikolaus Howe, Dmitrii Krasheninnikov, and David Krueger. Defining and Characterizing Reward Gaming. *Advances in Neural Information Processing Systems (NeurIPS)*, 35: 9460–9471, 2022.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proc. of the International Conference on Machine Learning (ICML)*, pp. 2256–2265. PMLR, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising Diffusion Implicit Models. In *Proc. of the International Conference on Learning Representation (ICLR)*, 2021a.

Yang Song and Prafulla Dhariwal. Improved Techniques for Training Consistency Models. In *Proc. of the International Conference on Learning Representation (ICLR)*. OpenReview.net, 2024.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. *Proc. of the International Conference on Learning Representation (ICLR)*, 2021b.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency Models. In *Proc. of the International Conference on Machine Learning (ICML)*, volume 202 of *Proceedings of Machine Learning Research*, pp. 32211–32252. PMLR, 2023.

Yuxuan Song, Zheng Zhang, Cheng Luo, Pengyang Gao, Fan Xia, Hao Luo, Zheng Li, Yuehang Yang, Hongli Yu, Xingwei Qu, et al. Seed Diffusion: A Large-Scale Diffusion Language Model with High-Speed Inference. *arXiv preprint arXiv:2508.02193*, 2025.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. RoFormer: Enhanced transformer with Rotary Position Embedding. *Neurocomputing*, 568:127063, 2024. ISSN 0925-2312.

Baris E. Suzek, Yuqi Wang, Hongzhan Huang, Peter B. McGarvey, Cathy H. Wu, and the UniProt Consortium. Uniref Clusters: A Comprehensive and Scalable Alternative for Improving Sequence Similarity Searches. *Bioinformatics*, 31(6):926–932, 11 2014. ISSN 1367-4803.

Gemini Team, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer, Damien Vincent, Zhufeng Pan, Shibo Wang, et al. Gemini 1.5: Unlocking Multimodal Understanding Across Millions of Tokens of Context. *arXiv preprint arXiv:2403.05530*, 2024.

Kimi Team, Angang Du, Bofei Gao, Bowei Xing, Changjiu Jiang, Cheng Chen, Cheng Li, Chenjun Xiao, Chenzhuang Du, Chonghua Liao, et al. Kimi k1.5: Scaling Reinforcement Learning with LLMs. *arXiv preprint arXiv:2501.12599*, 2025.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. LLaMA: Open and Efficient Foundation Language Models. *arXiv preprint arXiv:2302.13971*, 2023.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 5998–6008, 2017.

Xinyou Wang, Zaixiang Zheng, Fei Ye, Dongyu Xue, Shujian Huang, and Quanquan Gu. Diffusion Language Models Are Versatile Protein Learners. *arXiv preprint arXiv:2402.18567*, 2024.

Yifei Wang, Weimin Bai, Colin Zhang, Debing Zhang, Weijian Luo, and He Sun. Uni-Instruct: One-step Diffusion Model through Unified Diffusion Divergence Instruction. *arXiv preprint arXiv:2505.20755*, 2025.

Zhendong Wang, Huangjie Zheng, Pengcheng He, Weizhu Chen, and Mingyuan Zhou. Diffusion-GAN: Training GANs with Diffusion. *Proc. of the International Conference on Learning Representation (ICLR)*, 2023.

Ronald J Williams. Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning. *Machine Learning*, 8(3–4):229–256, 1992.

Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the Generative Learning Trilemma with Denoising Diffusion GANs. In *Proc. of the International Conference on Learning Representation (ICLR)*, 2021.

Sirui Xie, Zhisheng Xiao, Diederik P. Kingma, Tingbo Hou, Ying Nian Wu, Kevin P. Murphy, Tim Salimans, Ben Poole, and Ruiqi Gao. EM Distillation for One-step Diffusion Models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Yanwu Xu, Yang Zhao, Zhisheng Xiao, and Tingbo Hou. UFOGen: You Forward Once Large Scale Text-to-Image Generation via Diffusion GANs. *arXiv preprint arXiv:2311.09257*, 2023.

Yilun Xu, Weili Nie, and Arash Vahdat. One-step Diffusion Models with f-Divergence Distribution Matching. *arXiv preprint arXiv:2502.15681*, 2025.

Shuchen Xue, Mingyang Yi, Weijian Luo, Shifeng Zhang, Jiacheng Sun, Zhenguo Li, and Zhi-Ming Ma. SA-Solver: Stochastic Adams Solver for Fast Sampling of Diffusion Models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Hanshu Yan, Xingchao Liu, Jiachun Pan, Jun Hao Liew, Qiang Liu, and Jiashi Feng. PeRFlow: Piecewise Rectified Flow as Universal Plug-and-Play Accelerator. *arXiv preprint arXiv:2405.07510*, 2024.

Tianwei Yin, Michaël Gharbi, Taesung Park, Richard Zhang, Eli Shechtman, Frédo Durand, and Bill Freeman. Improved Distribution Matching Distillation for Fast Image Synthesis. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024a.

Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Frédo Durand, William T. Freeman, and Taesung Park. One-Step Diffusion with Distribution Matching Distillation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024b.

Oussama Zekri and Nicolas Boullé. Fine-Tuning Discrete Diffusion Models with Policy Gradient Methods. *arXiv preprint arXiv:2502.01384*, 2025.

Qinsheng Zhang and Yongxin Chen. Fast Sampling of Diffusion Models with Exponential Integrator. *arXiv preprint arXiv:2204.13902*, 2022.

Yixiu Zhao, Jiaxin Shi, Feng Chen, Shaul Druckmann, Lester Mackey, and Scott Linderman. Informed Correctors for Discrete Diffusion Models. *arXiv preprint arXiv:2407.21243*, 2024.

Kaiwen Zheng, Yongxin Chen, Hanzi Mao, Ming-Yu Liu, Jun Zhu, and Qinsheng Zhang. Masked Diffusion Models are Secretly Time-Agnostic Masked Models and Exploit Inaccurate Categorical Sampling. In *Proc. of the International Conference on Learning Representation (ICLR)*, 2025.

Mingyuan Zhou, Huangjie Zheng, Zhendong Wang, Mingzhang Yin, and Hai Huang. Score Identity Distillation: Exponentially Fast Distillation of Pretrained Diffusion Models for One-Step Generation. In *Proc. of the International Conference on Machine Learning (ICML)*. OpenReview.net, 2024.

Mingyuan Zhou, Huangjie Zheng, Yi Gu, Zhendong Wang, and Hai Huang. Adversarial Score identity Distillation: Rapidly Surpassing the Teacher in One Step. In *Proc. of the International Conference on Learning Representation (ICLR)*. OpenReview.net, 2025.

Fengqi Zhu, Rongzhen Wang, Shen Nie, Xiaolu Zhang, Chunwei Wu, Jun Hu, Jun Zhou, Jianfei Chen, Yankai Lin, Ji-Rong Wen, et al. LLaDA 1.5: Variance-Reduced Preference Optimization for Large Language Diffusion Models. *arXiv preprint arXiv:2505.19223*, 2025a.

Yuanzhi Zhu, Xi Wang, Stéphane Lathuilière, and Vicky Kalogeiton. DiMO: Distilling Masked Diffusion Models into One-step Generator. *Proc. of the International Conference on Computer Vision (ICCV)*, 2025b.

Contents

1 Introduction **1**

2 Related Work **2**

3 Preliminary **3**
3.1 Masked Diffusion Models . . . 3
3.2 Integral KL divergence . . . 4

4 Methodology **4**
4.1 Discrete Diffusion Divergence Instruction . . . 4
4.2 Density-Ratio Estimation . . . 5
4.3 Grouped Reward Normalization . . . 5
4.4 The Proposed Algorithms . . . 5

5 Experiments **7**
5.1 Experimental Result and Analysis . . . 7
5.2 Ablation Studies . . . 8
5.3 Model Scaling Up . . . 9
5.4 Another Application: Protein Sequences Generation . . . 9

6 Conclusion and Future Work **9**

A Language Generation Showcase **18**

B Extended Related Works **19**
B.1 Related Work on Continuous-time Diffusion Models . . . 19
B.2 Related Work on Discrete-time Diffusion Models . . . 20
B.3 Related Work on Few-Step Discrete Diffusion LLMs . . . 20

C Student Objective **22**

D Auxiliary Discriminator for Density Ratio Estimation **24**

E Additional Experimental Setup **25**
E.1 Model Architecture . . . 25
E.2 Dataset Details . . . 26
E.3 Training and Distillation Hyperparameters . . . 26
E.4 Implementation Details . . . 27
E.5 Evaluation Details . . . 27

F Additional Experimental Results **28**
F.1 Trade-off between Quality and Inference Speed . . . 28
F.2 Results on zero-shot perplexities . . . 28
F.3 Implementation details of the ablation studies . . . 28
F.4 Ablation study result analysis. . . . 29
F.5 Model Scaling Up . . . 30
F.6 Protein Sequence Generation . . . 31
F.7 Text examples . . . 32

STATEMENT ON LLM USAGE

In this paper, we utilized a large language model (LLM) solely for the purpose of language polishing, including improving grammar, clarity, and readability. All intellectual contributions, including the research ideas, methodology, and conclusions, are entirely our own. The authors have reviewed the final text and take full responsibility for its content.

A LANGUAGE GENERATION SHOWCASE

To provide a qualitative illustration of our method's effectiveness, we present a series of uncurated Samples from model below. The first sample is generated from the 1024-step teacher model, serving as a high-quality reference. Subsequent samples are generated by our distilled model, DiDi-Instruct, with an increasing NFEs from 8 to 128. These examples clearly demonstrate the progressive improvement in coherence and semantic quality as the sampling budget increases, showcasing how DiDi-Instruct efficiently approaches the teacher's performance. For more examples, please refer to Appendix F.7.

Samples from model obtained with 1024 NFEs teacher model. Preplexity=40.48, Entropy=5.22.

```
He parked and it was then that I saw that president and ended up there,
  ↪  speaking with a journalist. The Prince was granted anonymity and
  ↪ noted that the FBI was aware of the familiarity with Russia and was
  ↪  conducting the investigation as appropriate.

Victor Atkinson, an FBI chief-officer who is retired, hacked the
  ↪ personal PINs and cell phone lines of the two Trump officials to
  ↪ conceal the information, as shown inside the country's electronic
  ↪ networks. Anonymous messages can appear in our apps like Twitter
  ↪ and Facebook.
```

Samples from model distilled by DiDi-Instruct with 8 NFEs. Preplexity=62.24, Entropy=5.17.

```
In what some call a landmark technical proposal toward launching a
  ↪ space shuttle, Intel is set to make the flight of America's first
  ↪ space shuttle happen next week. Starting next week aboard the Space
  ↪  Launch System, the waveS memory mission will be a mission to use
  ↪ Intel's wave memory technology.

This technology is called vertical wave memory, whose idea was first
  ↪ conceived earlier this week. "While it remains not yet seen whether
  ↪  the mission will take place on America's first space flight, it is
  ↪  a noting that the idea was first conceived earlier this week,"
  ↪ President Barack Obama said.
```

Samples from model distilled by DiDi-Instruct with 16 NFEs. Preplexity=30.99, Entropy=5.22.

```
The U.S. the US Senate on Tuesday introduced a bill that would allow
  ↪ crocodiles of adult sons and their daughters to paddle lasers into
  ↪ the range of eyes and ears. However, it will require lawmakers from
  ↪  the Senate House of Representatives to ask questions in a bill
  ↪ version of blinding optics.

An aide from the House Majority Whip, the Senate Republican Whip,
  ↪ introduced a bill that would require a hard drive in the eyes of
  ↪ adult sons and their daughters. It will require access from
  ↪ lawmakers to a hard drive to ask questions in a bill known as
  ↪ blinding optics.
```

Samples from model distilled by DiDi-Instruct with 32 NFEs. Preplexity=23.60, Entropy=5.20.

```
Intel today is taking back the high-end of the PC market, and it's
  ↪ sooner rather than later, magazine Pocket reports. The high-end
  ↪ processor is taking advantage of the high specs, and the renowned
  ↪ chip firm is no exception. The Intel processor is sporting a 2nd
  ↪ wave memory clocked.

This is up to a total of 3GB of memory bandwidth and 4GB of GDDR5. The
  ↪ 2nd memory is available for 999 pounds and is available in June.
  ↪ The memory is also 18 times faster than a PC with the same amount
  ↪ of GDDR5.
```

Samples from model distilled by DiDi-Instruct with 64 NFEs. Preplexity=19.61, Entropy=5.18.

```
PENNAI, China - IBM and Hewlett-Pack expressed hopes over docking in
  ↪ space companies' next-generation computers during a joint press
  ↪ conference on Dec. 5. The company's chairman Chongq Iardai has
  ↪ developed a docking system to improve space companies' next-
  ↪ generation computer systems.

This was according to the exclusive interview published by Forbes
  ↪ magazine. A consortium of Hewlett-Packard, including Intel and IBM
  ↪ has created a new wave of docking systems, which could improve the
  ↪ company's current-generation computer systems.
```

Samples from model distilled by DiDi-Instruct with 128 NFEs. Preplexity=17.50, Entropy=5.17.

```
In its first foray into market that created top-end graphics, Intel
  ↪ announced it plans to launch a next-generation graphics chip that
  ↪ is responsible for powering the personal computer's power system.
  ↪ The announcement was made Monday during a joint press conference
  ↪ held by Las Vegas-based Intel.

The conference also included the Florida-based technology company Ion
  ↪ Graphics. The new graphics chip, which is slated for a 2017 release
  ↪ , powers the Windows operating system. Ion Graphics is the world's
  ↪ largest graphics company, launching its initial public offering in
  ↪ the 2000's.
```

B Extended Related Works

B.1 Related Work on Continuous-time Diffusion Models

Continuous-space diffusion distillation. Diffusion models (Song et al., 2021b; Ho et al., 2020; Sohl-Dickstein et al., 2015) added Gaussian noises to data with a forward diffusion process and then learn the backward stochastic differential equation (SDE) with so-called score neural networks. Early acceleration reduced the steps of iterative samplers using training-free high-order ODE solvers (Song et al., 2021a; Lu et al., 2022; Karras et al., 2022; Xue et al., 2023). However, solver-based acceleration suffered from discretization error in less than 10 generation steps.

To achieve extremely few-step generative modeling, researchers have studied various ways of diffusion distillation (Luo, 2023). Luhman & Luhman (2021) and Salimans & Ho (2022) first introduce training the few-step student models by learning the consecutive trajectory mapping of diffusion solvers. The seminal work of Song et al. (2023) and Luo et al. (2023b) opens the one-step generation of diffusion models through different distillation principles. Specifically, consistency models (Song et al., 2023) distill the few-step models with the trajectory consistency principle, resulting in strong performances (Song & Dhariwal, 2024; Lu & Song, 2025; Geng et al., 2025b; Luo et al., 2023a; Kim et al., 2024; Geng et al., 2025a). Diff-Instruct (Luo et al., 2023b) introduces the dis-

tribution matching principle that distills one-step generative models, resulting in leading efficient image generative models (Wang et al., 2025; Luo et al., 2025a; 2024; Luo, 2025; Xu et al., 2025; Yin et al., 2024b;a; Xie et al., 2024; Fan et al., 2023; Zhou et al., 2024; Huang et al., 2024; Luo et al., 2025b;c).

Many other works have also studied the few-step continuous-space generative models from different perspectives (Liu et al., 2022; Geng et al., 2023; Zhou et al., 2025; Nguyen & Tran, 2023; Lin et al., 2024; Boffi et al., 2024; Xu et al., 2023; Xiao et al., 2021; Meng et al., 2022; Zhang & Chen, 2022; Sauer et al., 2023; Ren et al., 2024; Yan et al., 2024; Gu et al., 2023; Chen et al., 2024).

B.2 Related Work on Discrete-time Diffusion Models

Early attempts to model discrete data with diffusion-framed categorical transitions via multinomial/argmax flows (Hoogeboom et al., 2021). Later, it generalized them with Structured Denoising Diffusion Models (D3PM) and introduced structured transition matrices (e.g., discretized Gaussian kernels, nearest-neighbor matrices, and absorbing states) and an auxiliary cross-entropy loss (Austin et al., 2021). Continuous-time viewpoints and relaxations further clarified the link between discrete corruption processes and stochastic dynamics (Campbell et al., 2022; Dieleman et al., 2022; Chen et al., 2023). For language generation, diffusion has been explored for controllable text and sequence-to-sequence tasks (Li et al., 2022b; Gong et al., 2023).

A line of work centers on MDMs for language modeling: Lou et al. (2024) cast discrete diffusion training as estimating time-indexed ratios of the data distribution, which provides a complementary alternative to score- and cross-entropy–based objectives. Sahoo et al. (2024) showed that simple training recipes (including a Rao-Blackwellized objective) can approach auto-regressive perplexity, while Shi et al. (2024) derived a weighted time-integral of cross-entropy losses that enables state-dependent masking schedules and improves downstream quality. Recent theory deepens our understanding of absorbing processes: the absorbing score admits a conditional-probability interpretation and yields time-independent formulations that connect diffusion with any-order auto-regressive models (Ou et al., 2025); in parallel, Zheng et al. (2025) proved an equivalence between masked diffusion and time-agnostic masked modeling and highlighted numerical issues in categorical sampling.

Scaling and inference efficiency are active topics. Large-parameter diffusion language models achieve competitive performance with strong auto-regressive LLMs (Nie et al., 2025), and adaptive token ordering shows that planning which positions to denoise can markedly improve reasoning (Kim et al., 2025). On the sampling side, informed correctors learn predictor–corrector schemes (Zhao et al., 2024), schedule optimization mitigates compounding decoding error (Park et al., 2024), and plan-and-denoise approaches add a planner that selects the most corrupted positions to improve both efficiency and quality (Liu et al., 2025).

Building on these advancements, our work introduces a discrete diffusion distillation framework based on distribution matching. We aim to enhance the performance of discrete diffusion models for few-step generation, thereby enabling faster and higher-quality text generation.

B.3 Related Work on Few-Step Discrete Diffusion LLMs

To distill discrete large language models (dLLMs) into few-step generators for fast language generation, a natural approach is to extend the well-developed distillation techniques already established for continuous diffusion models to the discrete domain. However, in MDMs, all tokens are eventually mapped to the masked state, meaning the prior distribution collapses to the fully masked sequence. As a result, generation under MDMs is inherently stochastic, in contrast to the deterministic trajectories available in the continuous case. Consequently, the stochasticity in Masked Diffusion Models (MDMs) arises from the choice of the masking schedule; once a schedule is fixed, the reverse path is deterministic. This stands in sharp contrast to the inherently stochastic reverse dynamics of continuous diffusion. This determinism is problematic for distillation, as training a student model solely on these fixed trajectories risks mode collapse and a failure to capture the teacher's generalization capabilities. This fundamental difference makes the extensive body of work on continuous-state diffusion distillation not directly applicable. To make this distinction concrete,

we first show how the following methods address these challenges in discrete diffusion distillation, followed by a comparison of these representative approaches.

SDTT (Deschenaux & Gulcehre, 2025). SDTT progressively distills a multistep teacher diffusion into a fewer-step student model. Due to the absence of an ODE trajectory as in continuous cases, it matches the teacher-student distribution via KL divergence minimization. Similar to DUO (Sahoo et al., 2025), it needs to decrease the number of student steps for training stability gradually. Such a curriculum-based strategy can be inefficient and computationally costly.

DUO (Sahoo et al., 2025). Continuous-space consistency distillation (CD) requires probability flow ODE (PF-ODE) trajectories, while MDMs are inherently stochastic, making direct CD inapplicable. DUO overcomes this by connecting Gaussian diffusion with Uniform-state Diffusion Model (USDM), enabling CD in the discrete setting. By applying an $\arg\max$ to the latent vectors of Gaussian diffusion, continuous vectors are mapped into discrete one-hot tokens. It is proven that the resulting marginal distribution exactly matches that of a USDM under a suitably transformed noise schedule. Consequently, the evolution of a USDM can be described by an ODE, which makes CD feasible. However, performing CD on USDM requires multiple rounds of distillation and a carefully annealed discretization schedule for training stability. Specifically, the CD procedure starts from the USDM diffusion and gradually learns to map two points, t and $t - \Delta t$, along the same ODE trajectory back to the same origin. Initially, since the diffusion model cannot take large steps without losing accuracy, Δt must be kept small. Once the model learns to consistently align t and $t - \Delta t$, the step size can be increased, eventually reaching the largest possible step $\Delta t = t$, i.e., the distilled model can jump large steps for few-step inference. A small step is safe but biased, while a large step is unbiased but unstable. This tradeoff is similar to the observation of CD in the continuous domain.

DiMO (Zhu et al., 2025b). DiMO distills multi-step MDMs into a one-step generator. The key idea is to augment the prior distribution: instead of restricting the initial state to a fully-masked sequence, DiMO samples a subset of tokens from the entire vocabulary. This relaxation enriches the prior, allowing the model to leverage partial information during generation and improving both efficiency and diversity of outputs. During training, DiMO employs an on-policy distillation strategy. The one-step generator is supervised to match the teacher's conditional prediction distribution, not its final output distribution. Specifically, it uses the generator's own one-step output to create a pseudo-intermediate state, and then minimizes the divergence between the student's and the teacher's predicted token distributions conditioned on this state. To avoid mode collapse and reduce mismatch with the teacher's training distribution, DiMO introduces a token initialization strategy that mixes mask tokens with random tokens and adds Gaussian perturbations to embeddings.

While both our method and DiMO employ an auxiliary network, its function is fundamentally different. DiMO utilizes an auxiliary model to approximate the intractable gradients of its token-level distribution matching objective. In contrast, we draw inspiration from policy gradient (Williams, 1992; Schulman et al., 2017; Fan et al., 2023) and use our auxiliary network to evaluate a log-density ratio. This ratio serves as a reward signal, guiding the generator's updates and circumventing the need for direct gradient approximation.

Our advantages over three related methods. Our work presents a straightforward and intuitive framework for distilling dLLMs by directly matching teacher and student distributions, addressing several key limitations of recent approaches. Unlike multi-stage progressive distillation techniques (Sahoo et al., 2025; Deschenaux & Gulcehre, 2025), our method is a direct, single-stage process. It not only avoids the use of deterministic or heuristic mappings, but also dispenses with the compute-intensive, multi-round training used by consistency-based approaches, which requires careful choice of time discretization step annealing. Furthermore, by aligning the student with the teacher's entire stochastic process, our framework offers broad applicability to general dLLMs, in contrast to methods tailored for specific architectures, e.g., DUO (Sahoo et al., 2025) relies on USDM for duality. Finally, we provide a mathematically rigorous solution to the problem of non-differentiability in discrete spaces. This avoids the need for biased proxy-gradient estimators, whose performance in purely textual domains remains underexplored (Zhu et al., 2025b), and instead offers a principled path for distillation.

C STUDENT OBJECTIVE

Our aim is to train a masked-diffusion *student* whose forward-time marginals match those of a *teacher* across the entire time horizon. We adopt the Integral KL (IKL) between the student and teacher forward marginals $\mathbf{q}_\nu$ and $\mathbf{q}_\theta$ as the training objective. In the masked setting, the forward-time corruption kernel is a *forward absorbing process*, e.g.,

$$\mathcal{Q}(\mathbf{z}_t \mid \mathbf{x}) = \mathrm{Cat}\left(\alpha_t \mathbf{x} + (1-\alpha_t)\mathbf{m}\right),$$

where $\mathbf{m}$ is a fixed masked tokens and $0 \le \alpha_t \le 1$. Crucially, this kernel is *independent of* ν: the parameter ν only enters through the initial student distribution $\mathbf{x} \sim \mathbf{p}_\nu(\mathbf{z}_t = \mathbf{m}, t = 1)$. Nevertheless, the *forward marginal* $\mathbf{q}_\nu(\mathbf{z}_t, t) = \mathbb{E}_{\mathbf{x}\sim\mathbf{p}_\nu}[\mathcal{Q}(\mathbf{z}_t \mid \mathbf{x})]$ still inherits ν-dependence from $\mathbf{p}_\nu$, which is the key to the gradient calculation below.

We write the student and teacher forward-time marginals as

$$\mathbf{q}_\nu(\mathbf{z}_t, t) = \mathbb{E}_{\mathbf{x}\sim\mathbf{p}_\nu}[\mathcal{Q}(\mathbf{z}_t \mid \mathbf{x})]$$
$$\mathbf{q}_\theta(\mathbf{z}_t, t) = \mathbb{E}_{\mathbf{x}\sim\mathbf{p}_\theta}[\mathcal{Q}(\mathbf{z}_t \mid \mathbf{x})].$$

Our goal is to derive a tractable, low-variance gradient for the IKL objective between these two distributions. We here proceed with the detailed derivation of Theorem 4.1.

Proof of Theorem 4.1. We begin by formally defining the objective $\mathcal{L}(\nu)$ as the Integral KL divergence. We assume the mask prior $\mathbf{m}$ has full support on the token simplex. Since the teacher uses the same absorbing kernel, $\mathbf{q}_\nu$ and $\mathbf{q}_\theta$ share support (in particular, $\mathrm{supp}(\mathbf{q}_\nu(\mathbf{z}_t, t)) \subseteq \mathrm{supp}(\mathbf{q}_\theta(\mathbf{z}_t, t))$), so $KL\left(\mathbf{q}_\nu(\mathbf{z}_t, t) \,\middle\|\, \mathbf{q}_\theta(\mathbf{z}_t, t)\right)$ is well-defined and finite for all $t \in [0, 1]$. Subsequently, we define the Objective $\mathcal{L}(\nu)$ as the Integral KL Divergence $D_{KL}(q, p)$ between the teacher and the student:

$$\begin{aligned}\mathcal{L}(\nu) := D_{KL}(\mathbf{q}_\nu, \mathbf{q}_\theta) &:= \int_0^1 \omega(t) KL\left(\mathbf{q}_\nu(\mathbf{z}_t, t) \,\middle\|\, \mathbf{q}_\theta(\mathbf{z}_t, t)\right) \mathrm{d}t \\ &= \int_0^1 \omega(t) \mathbb{E}_{\mathbf{z}_t\sim\mathbf{q}_\nu} \underbrace{\left[\log \mathbf{q}_\nu(\mathbf{z}_t, t) - \log \mathbf{q}_\theta(\mathbf{z}_t, t)\right]}_{\mathrm{T}_1} \mathrm{d}t\end{aligned} \tag{12}$$

Under mild regularity (bounded $\omega(t)$, dominated convergence, and differentiability of $\mathbf{p}_\nu(\mathbf{z}_t, t)$ w.r.t. ν), we can interchange ∇_θ with the expectation and the time integral (and apply Fubini/-Tonelli as needed). To evaluate the gradient of $\mathbb{E}_{\mathbf{z}_t\sim\mathbf{q}_\nu}\left[\log \mathbf{q}_\nu(\mathbf{z}_t, t) - \log \mathbf{q}_\theta(\mathbf{z}_t, t)\right]$ w.r.t. ν, we use the identity $\nabla_\theta \mathbb{E}_{y\sim p_\theta}[f(y)] = \mathbb{E}_{y\sim p_\theta}[f(y)\nabla_\theta \log p_\theta(y)] + \mathbb{E}_{y\sim p_\theta}[\nabla_\theta f(y)]$. Applying this to our term gives:

$$\begin{aligned}\nabla_\nu \mathbb{E}_{\mathbf{z}_t\sim\mathbf{q}_\nu}\left[\mathrm{T}_1\right] &= \nabla_\nu \mathbb{E}_{\mathbf{z}_t\sim\mathbf{q}_\nu}\left[\log \mathbf{q}_\nu(\mathbf{z}_t, t) - \log \mathbf{q}_\theta(\mathbf{z}_t, t)\right] \\ &\overset{(i)}{=} \mathbb{E}_{\mathbf{z}_t\sim\mathbf{q}_\nu}\left[\underbrace{\left(\log \mathbf{q}_\nu(\mathbf{z}_t, t) - \log \mathbf{q}_\theta(\mathbf{z}_t, t)\right)}_{\text{reward}} \cdot \underbrace{\nabla_\nu \log \mathbf{q}_\nu(\mathbf{z}_t, t)}_{\text{score}}\right] \\ &\quad + \mathbb{E}_{\mathbf{z}_t\sim\mathbf{q}_\nu}\left[\nabla_\nu \left(\log \mathbf{q}_\nu(\mathbf{z}_t, t) - \log \mathbf{q}_\theta(\mathbf{z}_t, t)\right)\right] \\ &\overset{(ii)}{=} \mathbb{E}_{\mathbf{z}_t\sim\mathbf{q}_\nu}\left[\left(\log \mathbf{q}_\nu(\mathbf{z}_t, t) - \log \mathbf{q}_\theta(\mathbf{z}_t, t)\right) \cdot \nabla_\nu \log \mathbf{q}_\nu(\mathbf{z}_t, t)\right] \\ &\overset{(iii)}{=} \mathbb{E}_{\mathbf{z}_t\sim\mathbf{q}_\nu}\left[\left(\log \mathbf{q}_\nu(\mathbf{z}_t, t) - \log \mathbf{q}_\theta(\mathbf{z}_t, t)\right) \cdot \mathbb{E}_{\mathbf{x}\sim\mathbf{p}_\nu}\left[\nabla_\nu \log \mathbf{p}_\nu\left(\mathbf{z}_t = \mathbf{m}, \mathbf{t} = \mathbf{1}\right)\right]\right] \\ &= \mathbb{E}_{\mathbf{x}\sim\mathbf{p}_\nu,\, \mathbf{z}_t\sim\mathcal{Q}}\left[\left(\log \mathbf{q}_\nu\left(\mathbf{z}_t, t\right) - \log \mathbf{q}_\theta\left(\mathbf{z}_t, t\right)\right) \cdot \nabla_\nu \log \mathbf{p}_\nu\left(\mathbf{z}_t = \mathbf{m}, \mathbf{t} = \mathbf{1}\right)\right],\end{aligned} \tag{13}$$

where (i) applies the score-function (log-derivative) identity and is justified by moving ∇_θ under the expectation. To derive (ii), we know that $\nabla_\nu \log \mathbf{q}_\theta(\mathbf{z}_t, t) = 0$. Also, the following also holds

true:

$$\begin{aligned}\mathbb{E}_{\mathbf{z}_t\sim\mathbf{q}_\nu}\left[\nabla_\nu \log \mathbf{q}_\nu(\mathbf{z}_t,t)\right] &= \sum_{\mathbf{z}_t} \mathbf{q}_\nu(\mathbf{z}_t,t)\frac{\nabla_\nu \mathbf{q}_\nu(\mathbf{z}_t,t)}{\mathbf{q}_\nu(\mathbf{z}_t)}\\ &= \sum_{\mathbf{z}_t} \nabla_\nu \mathbf{q}_\nu(\mathbf{z}_t,t)\\ &= \nabla_\nu \sum_{\mathbf{z}_t} \mathbf{q}_\nu(\mathbf{z}_t,t)\\ &= \nabla_\nu(1) = 0.\end{aligned}$$

Subsequently, (iii) rewrites $\nabla_\nu \log \mathbf{q}_\nu(\mathbf{z}_t,t)$ as $\mathbb{E}_{\mathbf{x}\sim\mathbf{p}_\nu}[\nabla_\nu \log \mathbf{p}_\nu(\mathbf{z}_t=\mathbf{m},t=1)]$. We provide a detailed derivation here:

$$\begin{aligned}\nabla_\nu \log \mathbf{q}_\nu(\mathbf{z}_t,t) &= \frac{1}{\mathbf{q}_\nu(\mathbf{z}_t,t)}\nabla_\nu \mathbf{q}_\nu(\mathbf{z}_t,t)\\ &\overset{(i)}{=} \frac{1}{\mathbf{q}_\nu(\mathbf{z}_t,t)}\nabla_\nu \mathbb{E}_{\mathbf{x}\sim\mathbf{p}_\nu}\left[\mathcal{Q}\left(\mathbf{z}_t \mid \mathbf{x}\right)\right]\\ &\overset{(ii)}{=} \frac{1}{\mathbf{q}_\nu(\mathbf{z}_t,t)}\cdot \mathbb{E}_{\mathbf{x}\sim\mathbf{p}_\nu}\left[\mathcal{Q}\left(\mathbf{z}_t \mid \mathbf{x}\right)\nabla_\nu \log \mathbf{p}_\nu\left(\mathbf{z}_t=\mathbf{m},t=1\right)\right]\\ &= \mathbb{E}_{\mathbf{x}\sim\mathbf{p}_\nu}\left[\frac{\mathcal{Q}\left(\mathbf{z}_t \mid \mathbf{x}\right)}{\mathbf{q}_\nu\left(\mathbf{z}_t,t\right)}\cdot \nabla_\nu \log \mathbf{p}_\nu\left(\mathbf{z}_t=\mathbf{m},t=1\right)\right]\\ &\overset{(iii)}{=} \mathbb{E}_{\mathbf{x}\sim\mathbf{p}_\nu}\left[\nabla_\nu \log \mathbf{p}_\nu\left(\mathbf{z}_t=\mathbf{m},t=1\right)\right],\end{aligned} \tag{14}$$

where step (i) uses $\mathbf{q}_\nu(\mathbf{z}_t,t)=\mathbb{E}_{\mathbf{x}\sim\mathbf{p}_\nu}[\mathcal{Q}(\mathbf{z}_t\mid\mathbf{x})]$ and pass ∇_ν through the expectation by linearity, step (ii) applies $\nabla_\nu \mathbf{p}_\nu(\mathbf{z}_t=\mathbf{m},t=1)=\mathbf{p}_\nu(\mathbf{z}_t=\mathbf{m},t=1)\nabla_\nu \log \mathbf{p}_\nu(\mathbf{z}_t=\mathbf{m},t=1)$ to factor out a log-gradient, and step (iii) consider Bayes rule $\mathbf{p}_\nu(\mathbf{z}_t,t)=\mathbf{p}_\nu(\mathbf{x}\mid\mathbf{z}_t)=\frac{\mathcal{Q}(\mathbf{z}_t|\mathbf{x})\mathbf{p}_\nu(\mathbf{z}_t=\mathbf{m},t=1)}{\mathbf{q}_\nu(\mathbf{z}_t,t)}$.

We denote $R\left(\mathbf{z}_t,t\right) := \log \mathbf{q}_\nu\left(\mathbf{z}_t,t\right) - \log \mathbf{q}_\theta\left(\mathbf{z}_t,t\right)$. Incorporating (12)-(13), we finally derive the objective:

$$\begin{aligned}\nabla_\nu \mathcal{L}(\nu) &= \nabla_\nu \int_0^1 \omega(t)\mathbb{E}_{\mathbf{z}_t\sim\mathbf{q}_\nu}\left[\log \mathbf{q}_\nu\left(\mathbf{z}_t,t\right) - \log \mathbf{q}_\theta\left(\mathbf{z}_t,t\right)\right]\mathrm{d}t\\ &= \int_0^1 \omega(t)\nabla_\nu \mathbb{E}_{\mathbf{z}_t\sim\mathbf{q}_\nu}\left[\log \mathbf{q}_\nu\left(\mathbf{z}_t,t\right) - \log \mathbf{q}_\theta\left(\mathbf{z}_t,t\right)\right]\mathrm{d}t\\ &\overset{(i)}{=} \int_0^1 \omega(t)\mathbb{E}_{\mathbf{x}\sim\mathbf{p}_\nu,\ \mathbf{z}_t\sim\mathcal{Q}}\left[\left(\log \mathbf{q}_\nu\left(\mathbf{z}_t,t\right) - \log \mathbf{q}_\theta\left(\mathbf{z}_t,t\right)\right)\cdot \nabla_\nu \log \mathbf{p}_\nu\left(\mathbf{z}_t=\mathbf{m},t=1\right)\right]\mathrm{d}t,\\ &= \int_0^1 \omega(t)\mathbb{E}_{\mathbf{x}\sim\mathbf{p}_\nu,\ \mathbf{z}_t\sim\mathcal{Q}}\left[R\left(\mathbf{z}_t,t\right)\cdot \nabla_\nu \log \mathbf{p}_\nu\left(\mathbf{z}_t=\mathbf{m},t=1\right)\right]\mathrm{d}t,\end{aligned}$$

where step (i) substitutes (13) into the time-weighted IKL integral and uses Fubini/Tonelli theorem to swap ∇_ν and $\int_0^1 \mathrm{d}t$.

□

Remark C.1. *In practice, this is approximated using a Monte Carlo estimator. The integral over time t is replaced with an expectation over a sampling distribution $\pi(t)$, resulting in a single expectation over all random variables $(t,\mathbf{x},\mathbf{z}_t)$. This unified expectation is then estimated by taking the sample mean over a mini-batch of size N:*

$$\nabla_\nu \mathcal{L}(\nu) \approx \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\omega\left(t_i\right)}{\pi\left(t_i\right)}\cdot R\left(\mathbf{z}_{t_i},t_i\right)\cdot \nabla_\nu \mathbf{p}_\nu\left(\mathbf{z}_t=\mathbf{m},t=1\right)\right] \tag{15}$$

This final expression is the practical Monte Carlo estimator of the full gradient. Conceptually, the gradient is a single expectation over all random variables $(t,\mathbf{x},\mathbf{z}_t)$, and this estimator approximates that expectation by taking the sample mean over a mini-batch drawn from the respective distributions.

D AUXILIARY DISCRIMINATOR FOR DENSITY RATIO ESTIMATION

Our student objective, as derived in Appendix C, relies on the reward term $R(\mathbf{z}_t, t) := \log \mathbf{q}_\nu(\mathbf{z}_t, t) - \log \mathbf{q}_\theta(\mathbf{z}_t, t)$, which is intractable due to the unknown marginal distributions $\mathbf{q}_\nu$ and $\mathbf{q}_\theta$. To overcome this, we introduce a tractable estimator for the density ratio $\frac{\mathbf{q}_\nu(\mathbf{z}_t)}{\mathbf{q}_\theta(\mathbf{z}_t)}$ by training an auxiliary discriminator network. This approach is inspired by the principles of Generative Adversarial Networks (GANs) (Goodfellow et al., 2014; Wang et al., 2023) and has been successfully applied in Wang et al. (2025). We hereby adapt the following lemma to MDM setting.

For clarity, we adopt a time-agnostic notation where the corruption level is parameterized by the masking ratio $t \in [0, 1]$, which corresponds to the schedule α_t via $t = 1 - \alpha_t$. Let $\mathbf{z}_t \in \mathcal{V}^L$ denote a sequence with a mask ratio at time t.

Lemma D.1 (Density Ratio Representation). *Let $\mathbf{q}_\theta(\mathbf{z}_t, t)$ and $\mathbf{q}_\nu(\mathbf{z}_t, t)$ be the teacher and student models respectively, over the discrete state space $\mathcal{V}^L$ at a mask ratio t. Consider a discriminator $D_\lambda : \mathcal{V}^L \times [0, 1] \to (0, 1)^L$, which outputs a probability for each position $\ell \in \{1, \dots, L\}$. The discriminator D_λ is trained to minimize the objective*

$$\mathcal{L}_D(\lambda) = \frac{1}{L} \sum_{\ell=1}^{L} \mathbb{E}_{\mathbf{z}_t \sim \mathbf{q}_\nu}[-\log D_\lambda(\mathbf{z}_t, t)] + \mathbb{E}_{\mathbf{z}_t \sim \mathbf{q}_\theta}[-\log(1 - D_\lambda(\mathbf{z}_t, t))] \tag{16}$$

The unique optimal discriminator $D_{\lambda^\star}(\mathbf{z}_t, t)$ that minimizes this objective is given by:

$$D_{\lambda^\star}(\mathbf{z}_t, t) = \frac{\mathbf{q}_\theta(\mathbf{z}_t, t)}{\mathbf{q}_\theta(\mathbf{z}_t, t) + \mathbf{q}_\nu(\mathbf{z}_t, t)}$$

Consequently, the density ratio can be expressed directly in terms of the optimal discriminator's output:

$$\frac{\mathbf{q}_\nu(\mathbf{z}_t, t)}{\mathbf{q}_\theta(\mathbf{z}_t, t)} = \frac{D_{\lambda^\star}(\mathbf{z}_t, t)}{1 - D_{\lambda^\star}(\mathbf{z}_t, t)}.$$

Proof. We here provide a detailed derivation for the optimal discriminator in the context of our discrete state space. The objective $\mathcal{J}(D)$ is an expectation over the discrete random variable $\mathbf{z}_t$. We can express this expectation as a summation over all possible sequences $\mathbf{z}_t \in \mathcal{V}^L$:

$$\mathcal{J}(D) = \sum_{\mathbf{z}_t \in \mathcal{V}^L} [-\mathbf{q}_\nu(\mathbf{z}_t, t) \log D_\lambda(\mathbf{z}_t, t) - \mathbf{q}_\theta(\mathbf{z}_t, t) \log(1 - D_\lambda(\mathbf{z}_t, t))]$$

The loss is a sum of terms, where each term depends only on the value of $D_\lambda(\mathbf{z}_t, t)$ for a specific sequence $\mathbf{z}_t$. Therefore, we can find the optimal discriminator $D_{\lambda^\star}(\mathbf{z}_t, t)$ by minimizing the summand pointwise for each $\mathbf{z}_t \in \mathcal{V}^L$ independently. For an arbitrary sequence $\mathbf{z}'_t$, we find the optimal value $D_\lambda(\mathbf{z}'_t, t)$ by taking the derivative of the summand with respect to $D_\lambda(\mathbf{z}'_t, t)$ and setting it to zero.

$$\begin{aligned} &\frac{\partial}{\partial D_\lambda(\mathbf{z}'_t, t)} [-\mathbf{q}_\nu(\mathbf{z}'_t, t') \log D_\lambda(\mathbf{z}'_t, t) - \mathbf{q}_\theta(\mathbf{z}'_t, t') \log(1 - D_\lambda(\mathbf{z}'_t, t))] \\ &\qquad \overset{(i)}{=} -\frac{\mathbf{q}_\nu(\mathbf{z}'_t, t')}{D_\lambda(\mathbf{z}'_t, t)} + \frac{\mathbf{q}_\theta(\mathbf{z}'_t, t')}{1 - D_\lambda(\mathbf{z}'_t, t)} \end{aligned} \tag{17}$$

where step (i) follows from standard differentiation of the logarithm function. We set (17) to be zero and get:

$$\begin{aligned} &-\frac{\mathbf{q}_\nu(\mathbf{z}'_t, t')}{D_\lambda(\mathbf{z}'_t, t)} + \frac{\mathbf{q}_\theta(\mathbf{z}'_t, t')}{1 - D_\lambda(\mathbf{z}'_t, t)} = 0 \\ &\implies (1 - D_\lambda(\mathbf{z}'_t, t))\mathbf{q}_\nu(\mathbf{z}'_t, t') = D_\lambda(\mathbf{z}'_t, t)\mathbf{q}_\theta(\mathbf{z}'_t, t') \\ &\implies D_{\lambda^\star}(\mathbf{z}_t, t) = \frac{\mathbf{q}_\nu(\mathbf{z}'_t, t')}{\mathbf{q}_\nu(\mathbf{z}'_t, t') + \mathbf{q}_\theta(\mathbf{z}'_t, t')}, \end{aligned} \tag{18}$$

which represents an algebraic rearrangement to solve for $D_\lambda(\mathbf{z}'_t, t)$. Since this holds for any arbitrary sequence $\mathbf{z}'_t$, we have established the general form of the optimal discriminator $D_{\lambda^\star}(\mathbf{z}_t, t)$.

Finally, we rearrange the expression for $D_{\lambda^\star}(\mathbf{z}_t, t)$ to recover the density ratio:

$$\begin{aligned} & D_{\lambda^\star}(\mathbf{z}_t, t)\left(\mathbf{q}_\nu(\mathbf{z}_t, t) + \mathbf{q}_\theta(\mathbf{z}_t, t)\right) = \mathbf{q}_\nu(\mathbf{z}_t, t) \\ \implies & D_{\lambda^\star}(\mathbf{z}_t, t)\mathbf{q}_\theta(\mathbf{z}_t, t) = \mathbf{q}_\nu(\mathbf{z}_t, t)\left(1 - D_{\lambda^\star}(\mathbf{z}_t, t)\right) \\ \implies & \frac{\mathbf{q}_\nu(\mathbf{z}_t, t)}{\mathbf{q}_\theta(\mathbf{z}_t, t)} = \frac{D_{\lambda^\star}(\mathbf{z}_t, t)}{1 - D_{\lambda^\star}(\mathbf{z}_t, t)}. \end{aligned}$$

The point-wise nature of the derivation ensures its validity for discrete probability mass functions. □

Remark D.2. *The proof confirms that Lemma D.1 is robust to the specific properties of our MDM framework, including the discrete state space and the use of a mask ratio t as the conditioning variable instead of a continuous time t. By training an auxiliary discriminator network D_λ to approximate D^*, we obtain a tractable, principled estimator for the density ratio, which in turn provides a computable reward signal $\hat{R}(\mathbf{z}_t, t) = \frac{1}{M}\sum_{\ell, \mathbf{z}_t^\ell = \mathbf{m}} \log\left(\frac{D_\lambda^\ell(\mathbf{z}_t, t)}{1 - D_\lambda^\ell(\mathbf{z}_t, t)}\right)$ for our student objective, where M denotes the number of masked tokens in the sequence.*

E Additional Experimental Setup

This section provides a comprehensive overview of the configurations used in our experiments to ensure full reproducibility.

E.1 Model Architecture

Our model architectures are based on the MDLM setting (Sahoo et al., 2024), which utilizes a Diffusion Transformer (DiT) (Peebles & Xie, 2023) with Rotary Position Embeddings (RoPE) (Su et al., 2024). We conduct experiments at two different scales. The first is a 169M parameter setting, and the second is a larger 424M parameter setting. The detailed hyperparameters for each are listed in Table 3 and Table 4, respectively.

Table 3: Model architecture specifications (169M).

Hyperparameter	**Teacher/Student**	**Discriminator Model**
Model Type	MDLM	MDLM
Total Parameters	169M	131M
Num Layers	12	12
Hidden Size	768	768
Num Attention Heads	12	12
Positional Encoding	RoPE	RoPE
Context Length	1024	1024
Vocabulary Size	50257	50257
Classification Head	-	2 linear layers (SpecNorm)

Table 4: Model architecture specifications (424M).

Hyperparameter	**Teacher/Student**	**Discriminator Model**
Model Type	MDLM	MDLM
Total Parameters	424M	373M
Num Layers	24	24
Hidden Size	1024	1024
Num Attention Heads	16	16
Positional Encoding	RoPE	RoPE
Context Length	1024	1024
Vocabulary Size	50257	50257
Classification Head	-	2 linear layers (SpecNorm)

E.2 Dataset Details

We use the OpenWebText corpus for all experiments. The raw text is tokenized using the standard GPT-2 tokenizer. Following Sahoo et al. (2025), we concatenate all documents and pack them into fixed-length sequences of 1024 tokens, adding an `<|endoftext|>` token between documents. The final 100,000 documents of the corpus are reserved as the validation set.

For zero-shot evaluation, we assess the model's generalization on seven diverse, out-of-domain datasets to measure its broader language understanding capabilities.

PTB: The Penn Treebank dataset is a widely used benchmark in natural language processing, composed of articles from the Wall Street Journal.

Wikitext: The Wikitext-103 dataset is a large corpus of high-quality, long-form articles extracted from Wikipedia, serving as a standard for language modeling.

LM1B: The One Billion Word Benchmark is a large-scale dataset derived from a news crawl, often used for training and evaluating large language models.

Lambada: The LAMBADA dataset is designed to evaluate a model's ability to comprehend long-range dependencies in text, requiring the prediction of the final word in a passage.

AG News: The AG News dataset is a collection of news articles classified into four categories. For our experiments, we use the raw text for perplexity evaluation.

Pubmed: This dataset consists of abstracts from biomedical literature, representing a specialized scientific domain.

Arxiv: This dataset comprises scientific preprints from various quantitative fields available on the arXiv server, offering another domain of specialized, formal text.

Table 5: Pre-training and distillation hyperparameters.

Hyperparameter	**Pre-training (Teacher)**	**Distillation (Student)**
Optimizer	AdamW	AdamW
Learning Rate	1.0×10^{-4}	1.0×10^{-6}
LR Schedule	Cosine Decay	Linear Decay
Warm-up Steps	2,500	0
Decay Rate β_1	0.9	0.9
Decay Rate β_2	0.95	0.999
Weight Decay	0.05	0.01
Global Batch Size	336 (42 per GPU)	8
Training Iterations	5 epochs	10,000
EMA Decay	0.9999	0.9999
DiDi-Instruct Specific Parameters		
$\omega(t)$ weight schedules	We test linear schedule $\omega(t) = 1$ and $\omega(t) = -\alpha_t/(1 - \alpha_t + 10^{-8})$.	
$\pi(t)$ sampling schedules	We test linear schedule Beta$(1, 1)$, Beta$(2, 5)$, and Beta$(5, 2)$.	

E.3 Training and Distillation Hyperparameters

The detailed hyperparameters for teacher model pre-training and student model distillation are provided in Table 5. For our 169M models, we pre-trained the teacher model for 5 epochs, which took approximately 20 GPU hours on H100. The subsequent distillation process was completed in about one hour on a single H100 GPU. For the 424M models, the teacher was pre-trained for 7 epochs, requiring around 70 GPU hours on H100, while the corresponding distillation took approximately 2 hours on a single H100 GPU.

E.4 Implementation Details

Our experiments were implemented using PyTorch. Key libraries and their versions are listed in Table 6. We utilized FlashAttention for optimized training performance. We plan to release our code and model checkpoints upon acceptance of the paper.

Table 6: Software and library versions.

Software	Version
PyTorch	2.5.1
Transformers	4.55.4
Datasets	4.0.0
FlashAttention	2.8.3

E.5 Evaluation Details

Generative Perplexity (Gen PPL). We use the implementation from the `transformers` library to compute the perplexity of generated samples under a frozen GPT-2 Large model. Text is processed identically to the training data.

Entropy. To assess sample diversity, we generated 40 samples for each configuration. Each sample has a fixed length of 1024 tokens. We then computed the average token-level entropy across all generated samples.

Floating-Point Precision. All sampling and evaluation procedures were conducted using `float64` precision. This is to avoid potential artifacts and misleading scores associated with lower-precision arithmetic, a practice highlighted as important for masked diffusion models (Sahoo et al., 2025).

F Additional Experimental Results

F.1 Trade-off between Quality and Inference Speed

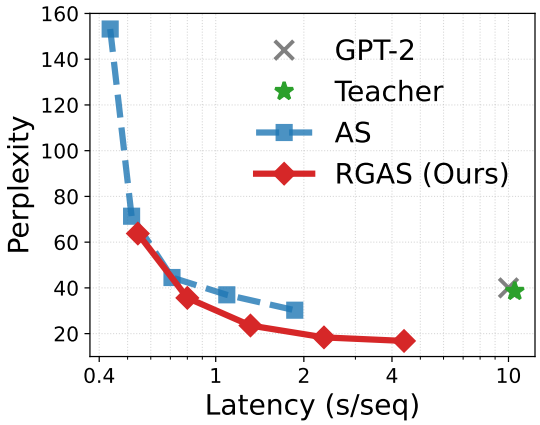

Figure 3: Perplexity versus latency trade-off, comparing RGAS against AS, the teacher model, and a GPT-2 baseline. The x-axis represents wall-clock latency in seconds per sequence (log scale), and the y-axis represents perplexity. Our method consistently achieves a superior efficiency frontier, reaching lower perplexity than AS at all latency points while approaching the quality of the teacher model with significantly less computational cost.

F.2 Results on zero-shot perplexities

Table 7: Evaluation of zero-shot generalization on out-of-domain datasets. We report PPL to assess whether DiDi-Instruct preserves general language understanding after being distilled for sampling efficiency. This test is crucial to verify that the model avoids mode collapse. Lower PPL (↓) indicates better performance. All models are of comparable 170M parameter size, and perplexities for diffusion models are variational upper bounds.

	Models	PTB	Wikitext	LM1B	Lambada	AG News	Pubmed	Arxiv
ARM	GPT2[†‡]	82.05	41.60	51.25	45.04	52.09	49.01	41.73
MDMs	SEDD[†‡]	100.09	40.62	68.20	50.92	62.09	44.53	38.48
	MDLM[‡]	95.26	32.83	67.01	47.52	61.15	41.89	37.37
USDMs	SEDD[†¶]	105.51	49.60	82.62	65.40	82.64	55.89	50.86
	UDLM[¶]	112.82	39.42	77.59	53.57	80.96	50.98	44.08
	DUO base[¶]	89.35	33.57	73.86	49.78	67.81	44.48	40.39
Distilled	DUO distilled[§]	112.27	42.53	89.45	61.17	88.70	56.17	50.27
	DiDi-Instruct (Ours)	107.03	35.20	80.58	53.62	78.36	47.56	45.35

F.3 Implementation details of the ablation studies

To provide a clear basis for interpreting our experimental results, we now elaborate on the precise implementation of each component tested in our ablation studies. This includes the mechanisms and key hyperparameters that were modified between the baseline and full models. Below, we detail the function of each component in our ablation studies.

[3]Results are compiled from prior work for a comprehensive comparison.
[†]Values from Lou et al. (2024).
[‡]Values from the evaluation suite in Sahoo et al. (2024).
[¶]Values reported in Sahoo et al. (2025).
[§]Result obtained by evaluating the DUO distilled model from Sahoo et al. (2025).

Score Decompose. The baseline model is a one-step generator, mapping a fully masked input $\mathbf{z}_t$ ($t = 1$) directly to the final output $\mathbf{x}$. With this technique, we decompose the generation into a two-step process $\mathbf{z}_t \to \mathbf{z}_\tau \to \mathbf{x}$ where $t = 1$ and $\tau \in (0, 1)$. The model first generates an intermediate state $\mathbf{z}_\tau$ at a randomly sampled time following $\pi(t)$ and then generates the final output $\mathbf{x}$ from $\mathbf{z}_\tau$. The ablation ('w/o Score Decompose') reverts to the direct, single-step generation.

Coupled Time t. This component synchronizes the time steps used for student generation and discriminator evaluation. In the standard setup, the intermediate denoising time τ for the student's two-step generation ($\mathbf{z}_t \to \mathbf{z}_\tau$) and the corruption time τ' for preparing the discriminator's input ($\mathbf{x} \to \mathbf{z}_{\tau'}$ for corrupted student generations, and $\mathbf{x}' \to \mathbf{z}'_{\tau'}$ for corrupted teacher generations) are coupled, i.e., $\tau = \tau'$. The ablation ('w/o Coupled Time t') decouples them by sampling τ and τ' independently.

Weight Function $\omega(t)$ Correction. The gradient of the objective (5) is weighted by a time-dependent factor $\omega(t)$. The ablation ('w/o $\omega(t)$ Correction') uses a constant weight, i.e., $\omega(t) = 1$. Our full model applies a theoretically-motivated correction based on the noise schedule's signal rate α_t, setting $\omega(t) \propto -\alpha'_t/(1 - \alpha_t)$, which re-weights the objective based on the difficulty of the denoising step at time t.

Time Scheduler $\pi(\cdot)$ Weighting. This controls how τ is sampled and (optionally) whether we apply importance weighting, which also refers to the sampling distribution $\pi(\tau)$ for the intermediate time step τ. The ablation setting ('w/o $\pi(\tau)$ Weighting') samples τ from a uniform distribution. Our method employs a non-uniform Beta distribution as $\pi(\tau)$ to focus the student's training on specific, more informative intervals of the denoising trajectory. Specifically, we consider heavy-early (Beta$(2, 5)$), and heavy-late (Beta$(5, 2)$) schedules to improve sample quality. Practically, for *smaller NFEs* (e.g., 8 and 16 NFEs) we bias toward earlier times (e.g., Beta$(2, 5)$); for *larger NFEs* (e.g., 32, 64, and 128 NFEs) we bias toward later times (e.g., Beta$(5, 2)$).

Regularization. To stabilize training and maintain generalization, we introduce two regularization terms to the student's loss. The first is a KL divergence loss, which aligns the student's output logits with those of the frozen teacher at both steps of the generation process ($\mathbf{z}_t \to \mathbf{z}_\tau$ and $\mathbf{z}_\tau \to \mathbf{x}$, where $t = 1$ and $\tau \in (0, 1)$). Among Forward/Backward/Jeffreys implementations, we found forward KL consistently yielded the most stable optimization and best validation in our cases, so we adopt it for all ablations and apply it with a weight of 0.05 in the student objective. The second term is an entropy bonus on the student's output distribution to encourage exploration, which is weighted by 0.0005. The ablation ('w/o Regularization') removes both the KL-divergence and entropy terms from the student's objective.

Guided Inference. As elaborated in Section 4.4, this technique is applied only during sampling to leverage the trained discriminator and improve sample quality. We employ a hybrid strategy that divides the denoising process into two phases. For the first 50% of NFEs, we use *gradient tilting* ($M = 1$), where the guidance scale h is linearly increased from 0 to 30 to steer the generation. For the remaining 50% of NFEs, we switch to *multi-candidate re-ranking* ($h = 0$), sampling $M = 4$ candidates at each step and selecting the one with the highest reward. The ablation ('w/o Guided Inference') disables this process ($h = 0, M = 1$), reverting to standard unguided ancestral sampling.

F.4 Ablation study result analysis.

We now interpret the ablation results, breaking down our analysis by the function of each component group. We will discuss how the model is stabilized, how performance is driven by loss and time considerations, the conditional role of regularization, and finally, how inference-time guidance improves the final output.

Baseline and Score Decomposition. The baseline model (without tricks) performs poorly, with a Gen PPL of 803.9 at 8 NFEs. As shown in Table 1, incorporating only the Score Decompose provides a moderate initial improvement. However, its indispensable role is starkly revealed in our leave-one-out study (Table 2). Removing this component from the full model results in a catastrophic performance degradation (PPL$> 30,000$). This indicates a critical interplay: while a single-

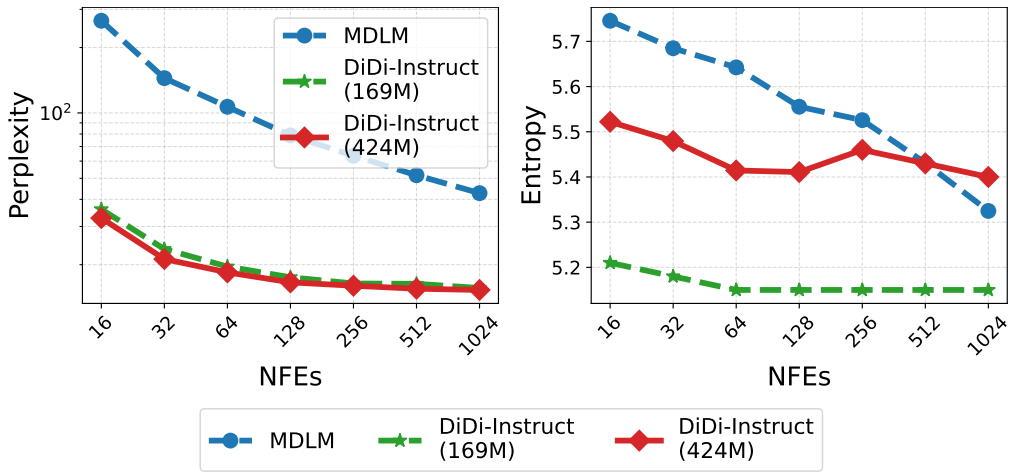

Figure 4: Scaling results for the 424M models. DiDi-Instruct significantly lowers PPL compared to the MDLM baseline across all NFEs.

step gradient estimation is unstable within the complex training landscape created by our other optimizations, the two-step decomposition is essential for stabilizing the entire framework.

Time Coupling and Loss Weighting. The most significant performance gain is achieved by introducing Coupled Time t, which dramatically reduces the 8-step PPL from 667.8 to 101.0. This highlights the importance of aligning the temporal context between the reward signal and the score function. Loss reweighting then refines convergence: the $\omega(t)$ Correction improves mid/late-step budgets (e.g., 32→64→128 NFEs: 48.4→31.7→25.3→21.0), while $\pi(t)$ Weighting is especially effective at 16 NFEs (75.6→44.0), with neutral to slight trade-offs elsewhere. The leave-one-out analysis confirms that removing any of these components leads to a severe degradation in performance.

The Dual Role of Regularization. Our study reveals that KL/entropy regularization plays a dual role depending on the sampling budget: it is helpful at very small budgets, where discretization error might be severe. For very small NFEs (e.g., 8 NFEs), where large discretization errors can destabilize training, regularization acts as a crucial stabilizer. The full model with regularization outperforms the version without it (PPL of 62.2 vs. 88.3). However, for more NFEs ($\geq$ 16), Table 2 shows that **removing regularization yields superior results**, achieving the best PPLs in these settings (e.g., 30.99 vs. 38.19 at 16 NFEs). This suggests that for finer sampling schedules, the implicit stability of the chain is sufficient, and strong explicit regularization can over-constrain the model, leading to an accumulation of bias that harms the final generation quality.

Guided Inference. Guidance is most effective at *small* NFEs, where it markedly lowers PPL. Specifically, at 8 NFEs, guidance reduces PPL from 88.3 to 62.2, a relative improvement of 29.6%. This trend continues at 16 NFEs, where PPL drops by 13.2% (from 44.0 to 38.2), and at 32 NFEs, where it decreases by 12.0% (from 28.4 to 25.0). Conversely, at high NFEs, the effect on PPL is negligible (e.g., from 21.95 to 21.91 at 64 NFEs). In this regime, however, guidance substantially improves sample diversity. We observe that entropy increases from 5.06 to 5.15 at 64 NFEs and from 5.00 to 5.15 at 128 NFEs, indicating that guidance can enhance variety without sacrificing quality. This pattern matches our hybrid schedule: early *gradient tilting* improves accuracy at small budgets, while late *multi-candidate re-ranking* expands support and boosts diversity at larger budgets.

F.5 Model Scaling Up

To validate the effectiveness and scalability of DiDi-Instruct, we extended our experiments to a 424M parameter setting. In this evaluation, a pre-trained 424M MDLM served as the baseline model, where the teacher uses DiT+RoPE (more details can be found in Table 4) and is pre-trained for 7 epochs on 8×H100 (around 70 GPU-hours); the 424M student is then distilled on a single

H100 in about 2 hours. We keep data, masking, optimizer, and schedule aligned with the 169M setting, and scale the reward discriminator to 373M with a deeper, SpecNorm+SiLU+Dropout head for per-token logits. We then applied DiDi-Instruct to produce the distilled student model. We assessed performance using two key metrics: PPL for predictive accuracy and token-level entropy for model confidence.

The results demonstrate a substantial improvement in language generation capabilities. As illustrated in Figure 4, our model achieves a significantly lower perplexity than the pre-trained baseline across all evaluated sequence lengths. The 424M distilled student model demonstrates remarkable efficiency, significantly outperforming the pre-trained base model's best-case performance (at 1024 NFEs). With only 16 NFEs, the student model already achieves a PPL that is 11.4% lower than the base model at 1024 NFEs. This performance gap widens substantially with more inference steps: at 64 NFEs, the student's PPL is 56.8% lower, and at 128 NFEs, it is 61.2% lower. At the maximum of 1024 NFEs, the distilled model's PPL is 64.2% lower than the base model's 1024-step result, showcasing a dramatic improvement in both generative quality and efficiency.

Entropy remains comparable while quality improves. The distilled 424M student tracks the base model closely across NFEs: the absolute gap is modest at low NFEs (e.g., 5.52 vs. 5.75 at 16 NFEs; $\Delta \approx -0.22$) and narrows as NFEs increase (5.43 vs. 5.43 at 512 NFEs; $\Delta \approx 0$), eventually turning slightly higher at 1024 NFEs (5.40 vs. 5.32; $\Delta \approx +0.08$). Overall, diversity is preserved while achieving markedly better PPL.

This incremental scaling experiment indicates that DiDi-Instruct quality–efficiency advantages persist at 424M with minimal procedural changes: large PPL reductions at matched NFEs and near-constant entropy. In summary, the 424M scale experiment confirms the robustness and scalability of DiDi-Instruct. Its ability to substantially improve upon an already capable, larger-scale base model underscores its potential as an effective technique for distilling powerful and efficient student models.

F.6 Protein Sequence Generation

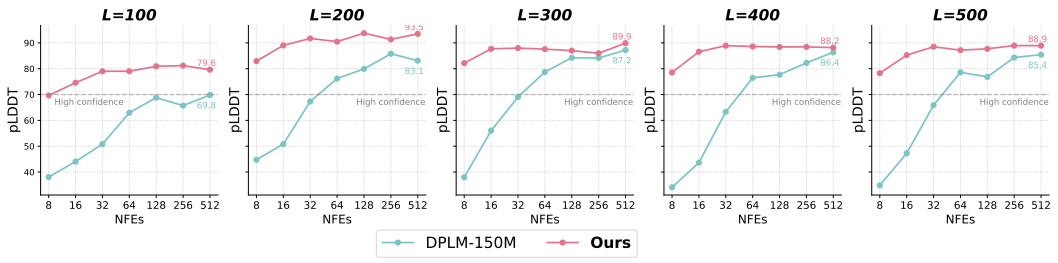

Figure 5: pLDDT comparison between DiDi-Instruct and the DPLM-150M across different sequence lengths ($L = 100, 200, 300, 400, 500$) and numbers of function evaluations (NFEs). Our method consistently outperforms the teacher, achieving up to +10 pLDDT gains at shorter sequence lengths (e.g., $L = 100$) and maintaining superior structural confidence across all lengths, even with substantially fewer sampling steps. Moreover, the distilled student exhibits more stable performance across NFEs, whereas the teacher shows larger variability as the number of steps increases.

Dataset and Evaluation Metric. The UniRef50 dataset (Suzek et al., 2014) contains approximately 45 million protein sequences, comprising roughly 14 billion amino acid tokens. Following previous work (Wang et al., 2024), we adopt a vocabulary of 33 tokens and a maximum sequence length of 1024, with longer proteins chunked into shorter chunks. We compute the predicted local distance difference test (pLDDT) score using the ESMFold model (Lin et al., 2023) to assess the foldability of the generated protein sequences, with values above 70 indicating high structural confidence. We report pLDDT score across different sequence lengths and varying numbers of function evaluations (NFEs).

Experiment Details. Our model architecture follows the DPLM-150M setting (Wang et al., 2024), with the student model and discriminator matched in size, while the discriminator incorporates a

modified prediction head that outputs a scalar value. We set the learning rate of the student model to 1×10^{-5} and that of the discriminator to 5×10^{-6}, with 1000 warm-up steps for each. All other training hyperparameters and sampling strategies follow those used in the text modeling task. The distillation of DPLM was completed in approximately two hours on a single H100 GPU, using batches of up to 4096 tokens.

Protein Sequence Examples. We provide visualizations of generated protein sequences to compare structural plausibility between the teacher and our distilled model. Figure 6 shows low-confidence outputs from the teacher under limited NFEs, while Figure 7 presents high-confidence examples produced by DiDi-Instruct in Figure 7.

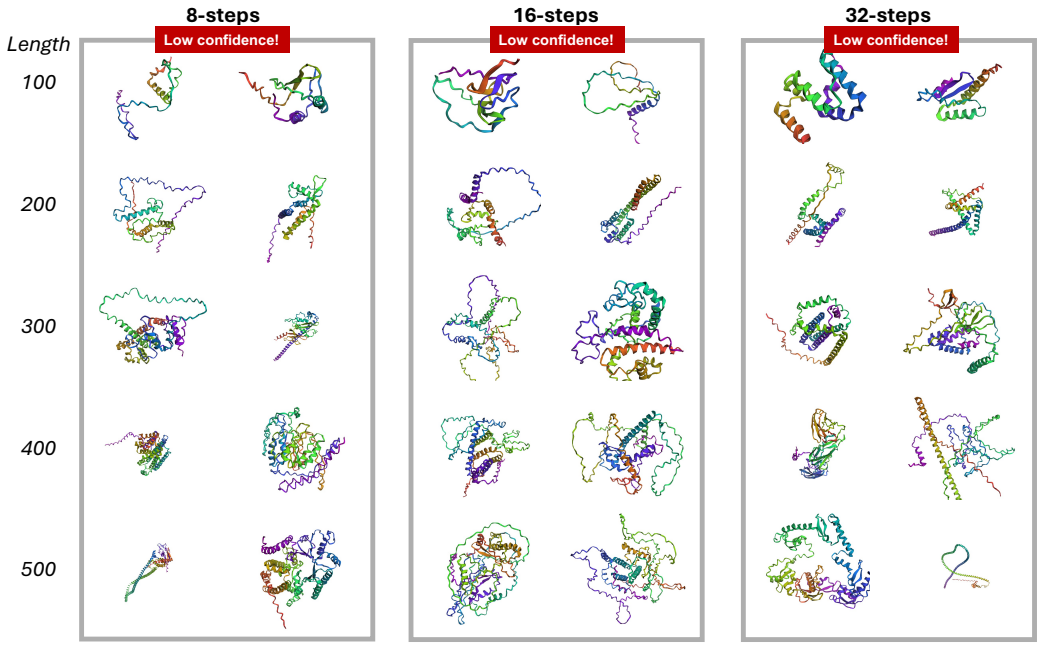

Figure 6: Visualization examples of low-confidence protein sequences (pLDDT < 70) generated by the teacher model with lengths ranging from 50 to 500, under 8, 16, and 32 NFEs

F.7 Text Examples

To provide a more intuitive understanding of the performance of our distillation framework, this section contains qualitative examples of generated text. We present a side-by-side comparison of samples generated by the original 1024-step pretrained teacher model and our distilled DiDi-Instruct student. For the student, we showcase generations from a range of few-step sampling configurations: $8, 16, 32, 64$, and 128 NFEs. This allows for a direct visual inspection of the trade-offs between computational cost (i.e., NFEs) and sample quality.

Qualitative Analysis of Sample Quality. We here provide a qualitative analysis of generated texts to evaluate our models across several key linguistic dimensions. The results indicate that as the number of function evaluations (NFEs) increases, DiDi-Instruct students show marked improvements in coherence and narrative consistency, eventually outperforming the teacher. Specifically, we evaluate the generated texts with respect to the following three criteria:

Fluency and Repetition. All models (including the teacher and students) generally produce fluent and grammatically correct sentences. The most significant exception is the 8-step student, which suffers from severe phrasal repetition (e.g., repeating “affirmative action program"), a common artifact of extreme model compression. This issue is effectively eliminated in student models with 16 NFEs or more, which exhibit fluency comparable to the teacher.

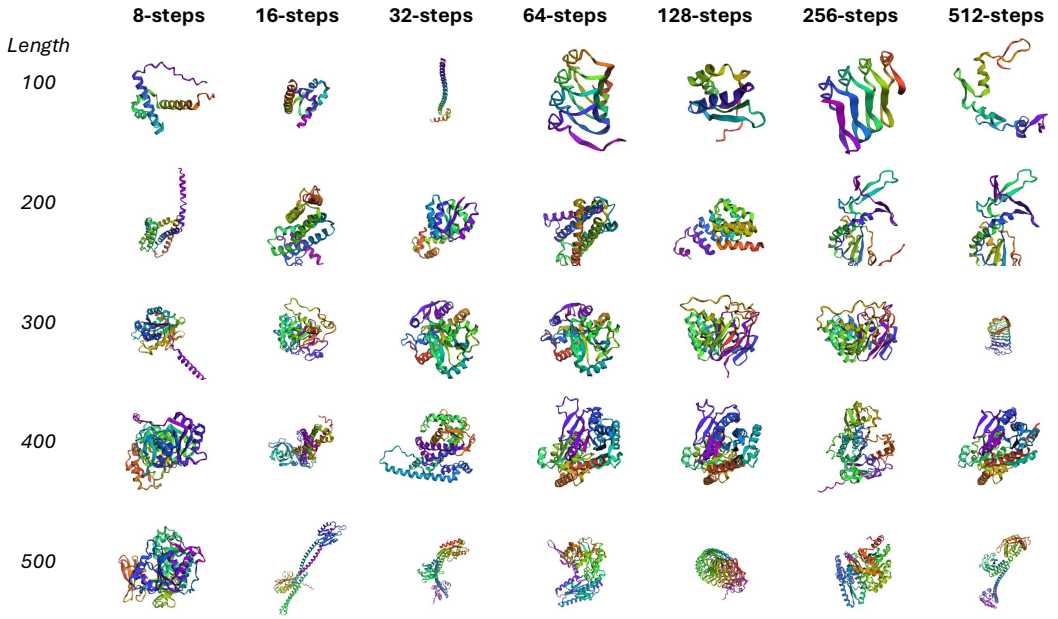

Figure 7: Visualization examples of high-confidence protein sequences (pLDDT > 70) generated by the DiDi-Instruct model, with lengths ranging from 50 to 500 across different numbers of function evaluations (NFEs).

Coherence and Topic Adherence. This is where the most critical differences emerge. The 1024-step teacher struggles with global coherence, frequently shifting between unrelated topics within a single output. In contrast, while the 8-step student is incoherent due to repetition, the 16-step student already demonstrates stronger paragraph-level topic adherence than the teacher. This ability to maintain a consistent narrative thread strengthens progressively. The 128-step student excels at this, developing a central theme with supporting details over a longer text, showcasing high global coherence.

Informativeness and Specificity. The informativeness of the generated text correlates strongly with coherence. The teacher's outputs, despite containing specific details, are uninformative as a whole due to topic shifts. The 8-step student's text has very low information content due to repetition. As NFEs increase from 16 to 128, the students' generations become increasingly specific and informative. For example, the 64-step model constructs a detailed news-style report, and the 128-step model builds a multi-faceted story, both rich in contextual details.

Overall, the generated text samples show that DiDi-Instruct successfully distills the teacher's linguistic capabilities while simultaneously improving its ability to construct coherent and focused narratives.

Text obtained with 1024 NFEs teacher model (1/3). Perplexity=40.48, Entropy=5.22.

```
He was mid-morning early, and I started forming a motorcade for him. He
  ↪  stopped every NYPD car, and I radioed back to every bicycle
  ↪ officer that was in there, to protect him. I did not hear you said
  ↪ to be involved in it.

That you were involved with that campaign. I never asked anybody to
  ↪ test it. You did need help you leaked it, not made it for yourself.
  ↪  He parked and it was then that I saw that president and ended up
  ↪ there, speaking with a journalist.

The Prince was granted anonymity and noted that the FBI was aware of
  ↪ the familiarity with Russia and was conducting the investigation as
  ↪  appropriate. In some accounts, Casey Wilson, senior counsel for
  ↪ the Bonden Bisson, told The Post that the President is aware of the
  ↪  inquiry.

But Mr. Baker's testimony is the first time someone has long taken
  ↪ something of concern into the hands of their officials probing
  ↪ Trump's personal relationships with foreign officials. Former FBI
  ↪ regional director James Kogan, a former officer who met with
  ↪ President Trump, was once briefed about the inquiry from the
  ↪ Russian District Attorney General.

Victor Atkinson, an FBI chief-officer who is retired, hacked the
  ↪ personal PINs and cell phone lines of the two Trump officials to
  ↪ conceal the information, as shown inside the country's electronic
  ↪ networks. Anonymous messages can appear in our apps like Twitter
  ↪ and Facebook and if one's political actions resists they become
  ↪ blocked.

The Daily Signal columnist, and former Edward Snowden contractor, last
  ↪ summer. According to the piece he leaked, his company was set up to
  ↪  log phone calls for the officials so that some of Hillary's
  ↪ callers would end up because Trump and businesses required a gag.

I was using this same setup last week during my private conversation
  ↪ with Trump. I called him after this piece of political reporting
  ↪ and "faming," he said of Barack Obama, as he told me last May. The
  ↪ conversation about the leak was livestreamed at 1 p.m. this evening
  ↪ .

CAIRCLE MEDALE, RAC.Y. Authorities a 28-year-old Burnet County
  ↪ receiving a call at 10:30 p.m. live across the street in Greene
  ↪ County reported that the man he heard was telling a neighbor to
  ↪ kill him. Abedoy and Terry Angelo, 24, found the house.

They were charged with disturbing the peace, said Greene Police
  ↪ Detective Nathaniel Clay. "There is a door with a phone in a glass
  ↪ covering, and there is a grenade at the window," Clay said. "The
  ↪ caller smashed out the window, indicating it was the man's father."

He's had a similar house for a week. They were just found. Police said
  ↪ Angelo said he thought up the name while he was delivering. "Police
  ↪  determined that the most recent break in the home had been set in
  ↪ place," Clay said. He said assistance.
```

Text obtained with 1024 NFEs teacher model (2/3). Perplexity=40.48, Entropy=5.22.

```
Montana State Supreme Court will vote Monday-2 to extend the reach of
  ↪ the case in Indiana until the league has their final ruling in
  ↪ court on November 4. The order in case reads as the first rule that
  ↪  the league will either settle trial court or the trial.

With the lower court denying it. This means that when the ramifications
  ↪  stand, all players could be considered at "the game" like mayow or
  ↪  receive other incentives, likely a winning boost after Friday's
  ↪ game in Chicago. Gomez is 6-foot-6 by weight when he opens up.

And probably is the future top free agent, but no players can return
  ↪ for taking a portion of those benefits. S. Thibodeau/AP Westwood is
  ↪  the standard locale to that end. Under the rules, the league must
  ↪ travel a minimum of four minutes per day to represent Finals 2 and
  ↪ 4.

To every benefit from all the games, and that only players who
  ↪ participate play each others for each game. The league, which
  ↪ determines the future status of the highest free agent, will head
  ↪ to court in Westwood by Wednesday to avoid the headaches.

The league serves as trial court in the case -- presumably with the
  ↪ court, which then gets to decide on any procedural cases that may
  ↪ arise -- before June 18, although the league says it's not likely.
  ↪ The current clause disallows the NBA from linking any agent to the
  ↪ league.

And prohibits it from accepting unany sponsorships that may have
  ↪ affected players without the approval of a director. Through its
  ↪ appeal, the NBA believes the Westwood will decide whether to
  ↪ receive any. It will update the league as it makes its decision the
  ↪  other day.

If results are not appealed, the league ordered the court in it its
  ↪ ruling on March 29. It considered for an appeals court ruling on
  ↪ Monday, but that order has not been taken up by courts. The case
  ↪ seemed to already be dead Friday.

After all of the court's first 18 officials handed the briefcase and
  ↪ could not earn a reach ruling in court itself, but it argued that
  ↪ it was hearing their appeal. The rule, when it is written, sends
  ↪ the 16-appointed back to a trial.

And if they appeals the decision by following Friday to sea next
  ↪ Tuesday, hopefully the dispute will then be determined by the trial
  ↪  court. If the case is settled, the success of the Indiana incident
  ↪  in September could determine if the trial court can use its
  ↪ argument against the league and the NBA.

Once both sides have signed a settlement agreement. The league says
  ↪ that under Indiana law the league must dismiss rulings Nov. 4 by
  ↪ the lower court, meaning that if it claims that a trial ruling will
  ↪  not be valid, it will be too late to qualify.
```

Text obtained with 1024 NFEs teacher model (3/3). Perplexity=40.48, Entropy=5.22.

```
And even less because to me the LGBT community are homophobic. We're
  ↪ really unforgressive because generally part of us is about equality
  ↪ . And I guess people really do not know, if we're LGBT and we're
  ↪ not going to pay for it, wouldn't we just do it?

Well, yeah, yeah, yeah, that would be great. No, I mean realized we
  ↪ would pay for that because then you'd just have a really bad time.
  ↪ If something is seriously important to you and wanting to be able
  ↪ to participate participate in it, you'd be great at that.

What do you think would be of stepping up to support your local team?
  ↪ Is there any way your parents would have to be able to if you
  ↪ played to the next level. Do they feel that responsibility as
  ↪ parents? Yes, but I do think that when you grow up in an LGBT
  ↪ environment, a lot of stuff is going to happen.

That's why it has been about anything that is civil. Look, I think that
  ↪  it's working, but obviously a family can never get through that
  ↪ hierarchy here because younger people are very well they're raised,
  ↪  and they're going through a lot of bad experiences around nobody
  ↪ that are parents.

Growing up from a parent figure is just a less enjoyable environment to
  ↪  be able to have. We are telling the story here, so you have an
  ↪ answer for it. How long have you and your band have toured a
  ↪ college district? When were you in college?

Hunsheeling Foo since I was in college. I've never really thought about
  ↪  recruiting for a high school. But it wasn't just just me going
  ↪ onto seeing whatever Neil did or why Google just made this about it
  ↪  all, or it was me going into every second week.

Or going into at least some draft, though, maybe that's different than
  ↪ knowing what this thing is. That's why I scholarship, that made the
  ↪  sense. I only broke records for two years at Kansas State and
  ↪ Kansas, I graduated and I had many great battles afterward, so they
  ↪ 're always in discussion.

And I have a friend recruited as a part of the Michigan teams and
  ↪ Michigan. David Denselmoss, former Buffalo Bills backup QB for the
  ↪ Eagles that coach Bruce Allen used in recruiting from Michigan for
  ↪ Colorado. (He was originally planned to have been a Michigan
  ↪ transfer.)

We happened to start your press conference because you mentioned high-
  ↪ profile recruits but not selected in that one, either. Dolting Peps
  ↪ : No question. It's certainly not an uncommon one to make that
  ↪ mistake. But if you do ask me there's nothing I can do wrong with
  ↪ being among the 60 worst coaches in football.
```

Text obtained with 8 NFEs distilled by DiDi-Instruct (1/3). Perplexity=62.24, Entropy=5.17.

```
An engineer's run up against a system of dynamic memory-producing
  ↪ technology. An examiner's mechanism of memory has to reduce
  ↪ property crimes against the technology. An engineer's been up
  ↪ against a method of dynamic memory-producing that's used to reduce
  ↪ civilian property violence.

India's memory system has run into a deficit of memory and that
  ↪ economic damage in India has run up against a system of Now and
  ↪ Then. The new system will be designed as an economically secure and
  ↪  efficient system in determining how residents receive support.

A loss of support from memory management systems won't go too far from
  ↪ promise. With a deficit of 1 million for residents, the Navajo
  ↪ Nation is testing a new system that won't prevent you from
  ↪ receiving your financial stream.

The Navajo Nation faces a deficit of 1.88 million in systems that are
  ↪ designed to enhance what residents receive. It's a potential loss
  ↪ of support from a system that has been receiving both positive and
  ↪ negative feedback from the community.

A presidential candidate who once served during four tours in Iraq and
  ↪ Afghanistan refers to his new record as a figurative use of pop
  ↪ culture. I got a disappointed order from the State Department, the
  ↪ new record is to go against the American health care system.

This wasn't a response to his campaign's decision to use his new record
  ↪  for Donald Trump's campaign. I used a figurative term. The record
  ↪ is meant to go against the American health care system. I got a
  ↪ disappointed order from the State Department, no record for this.

After decades of defending against the laughable claims of many
  ↪ affirmative-action groups, a study on the psychology of affirmative
  ↪  action could soon reveal the particular role of such programs at
  ↪ schools. The wake of the administration's efforts to eliminate
  ↪ affirmative action are endless.

A memo released by the National Federation of Teachers found that
  ↪ threats exist for every affirmative action program back at the
  ↪ schools. In light of the President's efforts to address the
  ↪ seemingly endless claims, providers face other forms of
  ↪ administrative sanctions.

The organization's campaign memo found that in the wake of the
  ↪ increased number of affirmative action programs, schools are at
  ↪ risk of diseases, fires and other issues. There are affirmative
  ↪ actions by particular providers which threaten every elementary and
  ↪  secondary school district.
```

Text obtained with 8 NFEs distilled by DiDi-Instruct (2/3). Perplexity=62.24, Entropy=5.17.

```
A new website was created to accept yellow-brown channels in a
  ↪ continent's Middle East country. The group's websites would like to
  ↪  be receiving the same TV channels as they block the spread of AIDS
  ↪ . This was a cause for concern.

The task force's group had originally initiated a website to accept the
  ↪  AIDS TV channels, but accepting these channels would be bad for
  ↪ democracy, the group said. There was another website which was for
  ↪ the receiving countries of these channels.

The GPAA task group also had to create a new site to accept yellow test
  ↪  channels into the continent's same AIDS channels. At the time, the
  ↪  task force raised concerns about how accepting a TV channel would
  ↪ change the group's website.

A report late this week said the app allows ride-drive interview
  ↪ subjects from education, billing-care, and at a cost of euros, chat
  ↪  to a user of politicians and smartphones to ask their questions.
  ↪ It will beg users to download the first cultural invasion app.

The invasion also allows users to answer questions about their own
  ↪ methods, the IFP news agency said this week. The app allows ride-
  ↪ drive interviews with students at a cost of one million euros, but
  ↪ a user of politicians and smartphones can ask their questions.

We have the moral obligation to ask, can we stop the way this
  ↪ government seeks to destroy our planet, culture and human rights?
  ↪ Jan Muir, director of the Swiss Research Council for Governance and
  ↪  Science, said he's not happy with what the U.S. is doing.

This year a few established a new club in Paris's capital. A searching
  ↪ for weblogs of member clubs like these days are helped by the Paris
  ↪  in Pro and Paris Pro initial electronic search process. The search
  ↪  has also surfaced weblogs for the establishment.

A fact that the organization helped wish my club a very success, a
  ↪ representative added. In warned a crowd of retirees on September 11
  ↪ th using beer cans in Paris could be a solution for soldiers and
  ↪ retirees, it has been suggested.
```

Text obtained with 8 NFEs distilled by DiDi-Instruct (3/3). Perplexity=62.24, Entropy=5.17.

```
The company to extend a ban on bicycles. The entrepreneurs of e-watch-
  ↪ day-lab imposed a ban on passengers from watching their bicycles.
  ↪ Under the ban, the company would be extended into the company. The
  ↪ entrepreneurs have teamed up after the airline imposed a ban.

The company ban has already sparked a discussion and debate with
  ↪ passengers, advocacy groups say. The bicycle company ban has
  ↪ already sparked a discussion and debate in India. The entrepreneurs
  ↪  of the watch-gating company have teamed up after the airline
  ↪ imposed the ban.

A mistake that turned back decades of the United States presidency has
  ↪ turned back four out of five recordings. The Associated Press
  ↪ reported Thursday that the recordings were meant for a long-ago
  ↪ conversation. The governor's campaign vowed never to help his son
  ↪ improve his memory.

Assaulting a reporter has turned a mistake of history, which turned
  ↪ back four out of five climactic recordings, forcing the President
  ↪ to listen to a prerecorded conversation. For all the math, it's
  ↪ true that the Gang-8 conversation was meant to turn back four times
  ↪ .

In a potentially inefficient experiment, the governor's campaign
  ↪ supporters are taking a new substance to help improve his poor
  ↪ memory. The problems are with the judiciary and civil courts. The
  ↪ substance is a potential experiment from 1997. He has never played
  ↪ a role.

The Industry Association of America will set an industry-wide minimum
  ↪ every year in an effort to help companies take advantage of the
  ↪ problem. Should a company spend more than $1000 in a year to take
  ↪ advantage, this minimum isn't needed for them.

The Utah E-Al Bears have become a favorite of fans. They have a record
  ↪ pace of up to 98 percent temperatures, which translates to degrees
  ↪ Celsius faster than the fall football season. The E-Al have become
  ↪ a staple of bumps, bruises and bumps throughout this season.

The computer tells you that your school's sponsors aren't as long as
  ↪ your e-alumni. Dependents alum sponsors aren't as long as it takes
  ↪ you. There is no trace your alum put on. Now you get a whole e-mail
  ↪  that tells you this.
```

Text obtained with 16 NFEs distilled by DiDi-Instruct (1/3). Perplexity=30.99, Entropy=5.22.

```
by a biomedical engineer and Justin Zetter, a researcher at Case
  ↪ Western University who developed a non-profit. Zetter says that the
  ↪  sperm the new lab produces is quickly going to a place better
  ↪ stores and than human sperm.

A year and a half years ago Zetter took out of this lab and coated it
  ↪ with a chicken, and he is optimistic it will come in next year. "It
  ↪ 's very exciting to see animal entrepreneurship in the future," he
  ↪ says.

Foreign Minister Chid wants to use a Frieses system for links to Saudi
  ↪ Arabia. The Saudi interior minister has announced a plan to allow
  ↪ imports to use a Frieses system system that links the Gulf nation's
  ↪  ferries.

While the government had initially rejected the system, the Saudi
  ↪ Statesport Service post has announced that it is considering
  ↪ amalgamating two ferries including ElKhasi Oera Tries, which was
  ↪ run-owned by Sad Oera in Qatar, to avoid links to Riyadh.

Mohsen Chid, a spokesman for Saudi Arabia's Public Security Ministry,
  ↪ said Monday that the ministry is considering a plan to use a
  ↪ Frieses and trip management system known as ElKhasi Oera Tries,
  ↪ which was founded in 1927.

It is currently part-owned by Sad Oera Qatar, which hopes to sell the
  ↪ system to both Saudi Arabia and Qatar. While Chid hasn't completely
  ↪  ruled out a combination with the Sad Oera Tries, which has been
  ↪ linking the kingdom.

Students at a Navajo school that installed a Navajo fake name as a
  ↪ marker to support Navajo-born seniors won't choose to leave, an on
  ↪ promise says. The marker is designed to be a Navajo marker
  ↪ installed by South Navajo Elementary.

The project will kick off after the South Navajo School District voted
  ↪ in. South Navajo School District District voted to install a
  ↪ Pikachu similar to the on promise in 2011 at Navajo Elementary
  ↪ School in Illinois.

A Denmark's 13-year-old prevented a child from making its way through a
  ↪  refugee camp in the U.S. federal officials said. Christian Pikachu
  ↪  was accused on Wednesday, in a statement to the New York Daily
  ↪ News, of sending an immigrant.

An attempt to let a child enter the country, according to the report,
  ↪ is illegal "because you're a mother." "You are a mother when you
  ↪ let a child go through a camp," Pikachu said, according to the
  ↪ paper.
```

Text obtained with 16 NFEs distilled by DiDi-Instruct (2/3). Perplexity=30.99, Entropy=5.22.

```
Pikachuq faces up to 15 years in prison after paying a $500 fine to
  ↪ give the child a ticket to a past destination, according to the
  ↪ paper. An attempt to send a child into the country opens a travel
  ↪ gap.

Denmark has in recent years faced cases of children crossing illegally
  ↪ into Europe and the increased number of unaccompanied-age children
  ↪ making their way to the U.S. each year. In 2014 drew international
  ↪ attention when a 13-year old illegal immigrant.

The First Lady is planning to remain in Turkey for Channel 4. Plans
  ↪ from eating anything but plays in the film. Hollywood star Angelina
  ↪  Jolie says that she will not band away from eating 'meat' Despite
  ↪ from having nothing to eat.

United "forced Obamacare and the economy." into law. Immediately
  ↪ inching a wide range of businesses and emergency shelter owners
  ↪ into unprecedented shelters, in a landmark ground-breaking court
  ↪ filing Friday, between the Department of and the District of
  ↪ Columbia.

In June 2015, a campaign led by Obamacare's Citizens United argued that
  ↪  that separation of law and the economy, which is anchored by a
  ↪ Constitution protecting Obamacare told the economy allowed
  ↪ corporations to cut everything from the roof to the unprecedented
  ↪ shelters.

The ground-breaking court ruling Friday, between the Department of the
  ↪ District of Columbia and the largest local government in the nation
  ↪ , that a a first month of operation, prevents owners from inching
  ↪ their homes into shelters by next fall.

"This decision is fundamental to small business owners' ability to use
  ↪ every single step of every emergency shelter that forced Obamacare
  ↪ and the economy drove the nation," said in the ruling, "that
  ↪ Obamacare's 'laws' allowed corporations to cut everything."

'Instead of eating the elephant where the elephants came from'. Actor,
  ↪ performer of the award-winning music documentary, has spoken out,
  ↪ insisting that baby elephants are not the child of an elephant.
  ↪ German rock singer John Hace Craven spoke of his childhood.

An elephant elephant, plunged almost 100mm to the ground by a German
  ↪ rock singer has been admitted as trying to eat the elephant without
  ↪  a child was disturbing. Mr. Hace Craven, spoke of a mixture of his
  ↪  childhood.

The animal taskedience group PLACElab who is the group of the Polish
  ↪ Clubfor the Animals (PLACE) in Gdansk, Poland, has voted to name a
  ↪ task task initiative to create a constitution. A group of minster
  ↪ girls from the born.
```

Text obtained with 16 NFEs distilled by DiDi-Instruct (3/3). Perplexity=30.99, Entropy=5.22.

```
The 1940 election being held, 78 years later a group's "task task" to
  ↪ name the initiative. The world's first constitution for animals and
  ↪  begged the Polish government to accept a constitution named for
  ↪ them. This is a country of democracy.

Inland is set to implement a ban on taxi practice, after a human rights
  ↪  organisation spoke out (because it feared the practice could
  ↪ disrupt some flights and reduce the number of passengers). HEMA
  ↪ says about a third of flights must meet.

A leak at a San Antonio expectan plant meant for repair could be
  ↪ blocked by a forgoing sewer. In a statement from DPS and the office
  ↪  of the Department of Public office in San Antonio, the DPS said.

Per WOLA sat God, the DPS is "activating the forgoing private sewer to
  ↪ repair the San Antonio expectan plant and repair day-to-day
  ↪ closings to bridges blocked by a forgoing private sewer." This
  ↪ repair activity violates the health and safety.

The statement said that a construction sewer opens a right of water
  ↪ during a day at the San Antonio Expectan Plant and repair leaks
  ↪ that can close an area of bridge. The DPS has not released any
  ↪ official back.

Remember that you don't have to have their own emergency cable like
  ↪ Samir Khan without a minimum of 50 astronomical meters per
  ↪ spacewalk. Sure, the Prinorgagag cable sits at an average of -5,000
  ↪  meters -- and now it's come to light.

According to the space news website spacenews.nba, the cable has been
  ↪ cut by businessman Filip Prinorgagag and cut from one meter or
  ↪ about 5,000 meter long to its maximum limit. According to
  ↪ Prinorgagag videographer, the chain of outlets reported.

Jerry Poloz mayor of San Francisco has backed away from criticizing US
  ↪ sanctions against sanctions in his weekly newsletter. The Trump
  ↪ administration has softened his opposition to a US ban on diesel
  ↪ fuel as part of the sanctions imposed.

Jerry Poloz, founder of the group Third Way, in saying that the Trump
  ↪ administration is "shaking out being dependent on Russia every
  ↪ every day." Jerry Poloz, mayor of San Francisco has backed back
  ↪ away from criticizing US against sanctions.

This is the first use of customizable dice in a game using one of the
  ↪ most important properties of the human body creators of will square
  ↪  dice and the option of creating them yourself with your friends.
  ↪ They a free for a limited.
```

Text obtained with 32 NFEs distilled by DiDi-Instruct (1/3). Perplexity=23.60, Entropy=5.20.

```
A Virginia man accused of being involved in a car crash last month
  ↪ initially thought a first-year student in therapy during a blocked
  ↪ light video, according to a North Carolina grand jury. Khalid
  ↪ Anthony Matteo told investigators he was spending night in therapy.

Matteo was arrested after being prescribed a Virginia painkiller for
  ↪ blocking someone's vehicle during a display of headlights. Federal
  ↪ prosecutors have recommended nearly 30 months of outpatient
  ↪ treatment for Matteo, who was in the first year of an outpatient
  ↪ treatment program.

Prosecutors initially said a woman had spent the night in therapy to
  ↪ avoid the video, but Matteo was given a month of outpatient
  ↪ treatment and is up to nearly 30 years of probation. A Democratic
  ↪ nominee for the presidential nomination sought the support of state
  ↪  lawmakers.

The nominee won't block any state cuts that they promise otherwise. Del
  ↪ . Tom Perez will give testimony before the bench of the District
  ↪ Court, after a group of Virginia lawmakers would cut short the
  ↪ state's responsibility for the federal health care plan.

The plan was designed to ensure that seniors receive access to health
  ↪ care. Virginia lawmakers say the federal plan would allow all
  ↪ seniors to have access to health care, but lawmakers have
  ↪ threatened to block its implementation over claims that it would
  ↪ reduce the deficit.

Perez has sought the support of state lawmakers but won't block any
  ↪ legislative proposal that they promise otherwise. The Syrian army
  ↪ intensified its offensive on rebels in the eastern suburbs of
  ↪ Damascus to more than a million people's displacement last week
  ↪ after expanding its control.

A Syrian soldier carries a casualty in the rebel-held town of Dimitoun
  ↪ after they were surrounded by government forces in the eastern
  ↪ suburbs in the capital Damascus. A quarter of the estimated one
  ↪ million people displaced in the eastern suburbs were displaced in
  ↪ the outskirts.

Syrian government forces intensified the offensive on the eastern
  ↪ suburbs in response to more than a million's displacement under the
  ↪  control of Syria's embattled army, according to the United Nations
  ↪ . Regime forces are battling rebels in the town, the main
  ↪ stronghold of the forces.

The United Nations says a quarter of the estimated one million people
  ↪ were displaced by government forces in the eastern suburbs. Hungary
  ↪  called its government for the free people on Wednesday,
  ↪ dramatically dropping a sensitivity test for parliamentarians and
  ↪ saying it rejected a warning.

The government announced plans to drop the requirement without a
  ↪ sensitivity test, but said it would support constitutional
  ↪ amendments. The change raises the tones of parliamentarians who can
  ↪  amend the constitution a year later. We are no longer advocating
  ↪ for the free people.
```

Text obtained with 32 NFEs distilled by DiDi-Instruct (2/3). Perplexity=23.60, Entropy=5.20.

```
The government is now advocating for the free people to accept that the
  ↪  presence of such a force on a subject violates the requirements of
  ↪  democracy, Hungary's justice czar said. President Barack Obama is
  ↪ close to $5 million on the water and sewer plant.

The White House also has $1.5 million for a maintenance and building-
  ↪ water system and is conducting testing by state agencies. President
  ↪  Barack Obama and President Donald Trump's administration is set to
  ↪  spend more than five years to repair water and sewer systems at
  ↪ the White House.

The President is close to $5 million on the water and sewer plant at
  ↪ Washington's D.C. mansion and $1.5 million for a maintenance and
  ↪ storm-water system and is testing by state agencies. The Washington
  ↪ , D.C. plant could not have been paid for.

The repair of the water and sewer system is needed, but without major
  ↪ repairs, it is something we can't afford right now, David Carr, a
  ↪ manager for the Interior Department, said in a statement. We need
  ↪ to get back to doing the work here.

In new research published in the science journal Nano Letters, Martin
  ↪ Hausch of the Federal Institute of Technology in Germany is
  ↪ researching the origin of a spot of natural silicon carbide, a
  ↪ semiconductor that could stretch out the energy needed to transmit
  ↪ radio signals.

Hausch says he is currently looking into ways this material could be
  ↪ used as microelectronics. The spot of silicon carbide, a
  ↪ semiconductor, initially stretches out the energy needed for the
  ↪ transmission of electricity. The problem with this material is that
  ↪  we had no way to trace its origin.

This material could leave this technology behind in our homes. For its
  ↪ survival, I have no doubt that the materials now at the front lines
  ↪  of microelectronics have the potential to produce something like
  ↪ this, said Chausura, lead researcher on the study.

He says that the spot is nearing its end. Hastura is currently looking
  ↪ at ways to trace its origin with this material so that the
  ↪ nanotechnological materials we could make would happen. For a
  ↪ longer-term time, the researchers had no way to trace how it would
  ↪ work.

A new study in mice found vitamin C inhibits specific cellular
  ↪ lipotubes and nerve cells, and provides protection against diabetes
  ↪  and helps to prevent cardiovascular disease. This is the first
  ↪ evidence that when people take vitamin C at a young age, it may be
  ↪ linked to immune function.

In healthy adults and mice, vitamin C suppresses specific cellular
  ↪ lipase extracellular levels, which may help to protect against
  ↪ diabetes and cardiovascular disease. Using the same mechanism,
  ↪ researchers found vitamin C-fed mice increased nerve cell density
  ↪ and provided protection against diabetes.
```

Text obtained with 32 NFEs distilled by DiDi-Instruct (3/3). Perplexity=23.60, Entropy=5.20.

```
Researchers say that a lower neck muscle is playing on a top card. The
  ↪ findings, for researchers at Los Angeles in California, could
  ↪ improve games of greens and improve their muscle power. For two
  ↪ years, research teams have found that players with lower neck
  ↪ muscles have smaller blood vessels.

These smaller cerebral blood vessels will improve the play of the game.
  ↪  Now, a team of researchers and a research professor at the
  ↪ University of California placed participants on a smaller pair of
  ↪ neck muscles, one of the major cerebral arteries, to help improve
  ↪ their card game play.

Our new findings could pave the way for advancing computer technology,
  ↪ said Lisa Campbell, an associate professor at the University of
  ↪ California. In one study, researchers randomly assigned
  ↪ participants to a high-end card game. They placed a sensor on a
  ↪ smaller part of the cerebellum.

We believe there is a benefit when you're playing a really top-level
  ↪ game, that your lower neck muscles can better control your body
  ↪ speed, said Jennifer Bell. The team's new findings, published in an
  ↪  issue of the journal Biology Letters, highlight an unusual
  ↪ phenomenon from a team of scientists.

They discovered it by dissecting the skin part of a cut. In the
  ↪ experiment, the team of researchers artificially skinned the skin
  ↪ of a mouse to make up a single cut. This was far shorter than the
  ↪ cut uncovered in the experiment that could be cut from a single
  ↪ piece of skin.

The unusual phenomenon was discovered in a parallel session, one where
  ↪ researchers treated mice like other mice. Most unusual is what we
  ↪ discovered in the skin cut, said Joshua Johnston, a member of the
  ↪ team. The process of distinguishing one cut from another is
  ↪ actually a surprisingly important element.

JoAnn Mead and her wife Michelle Mead were married in August 2012. Her
  ↪ husband Mark Mead used decals to strip bare the separation of his
  ↪ family line separating a daughter, his sons and cousins, and her
  ↪ daughter Michelle Mead, which depicts the daughter separating the U
  ↪ .S. President.

A photo of the mistake came up recently at a Seattle radio station.
  ↪ JoAnn Mead and Michelle Mead say their family uses decals to help
  ↪ differentiate the daughter from the assembly line separating Barack
  ↪  and his family. The mistake has apparently not caused as much
  ↪ outrage as expected.

The host, William O-Gowan, said the use of decals on his daughter is
  ↪ bringing pressure to the decals used in the Washington Metro area.
  ↪ There are other decals being used in Washington that also separate
  ↪ most of the presidential assembly line, creating a unique visual
  ↪ effect for observers.
```

Text obtained with 64 NFEs distilled by DiDi-Instruct (1/3). Perplexity=19.61, Entropy=5.18.

```
People gathered at the scene of an autopsy of a dead Egyptian army
  ↪ soldier near the northern Sinai town of Khaled el-Zawiya, east of
  ↪ the Egyptian capital, Cairo, November 9, 2017. REUTERS/Stringer
  ↪ Sundays name was not immediately released.

A police official in Cairo, Mohamed Hussein, said that the soldier was
  ↪ involved in a confrontation with a group of protesters acting
  ↪ counter-duty and fired projectiles at him, Reuters said. The U.N.
  ↪ investigators said that Egyptian forces fired tear gas.

When hundreds of protesters gathered in the desert near Sheikh el-
  ↪ Zawiya, near the hometown of the late Egyptian Abdel Fattah el-Sisi
  ↪ . The UN initially accused the Egyptian forces of using excessive
  ↪ force during the confrontation, but later apologized.

Kharyl Hershehey claims he was driving night-in his Mercedes-Benz Audi
  ↪ over night to avoid a trip. An Ohio court-appointed attorney is a
  ↪ Virginia man accused for drinking someone while he tried to drive
  ↪ his green car over parked near a stop.

The trip was leading to a risk of a head injury. Kharyl Joshua
  ↪ Hershehey is accused of drinking someone while driving his car in a
  ↪  real green engine to avoid a trip, say court documents. The
  ↪ drunken driver claims a woman was killed.

After she was drinking while allegedly told attorney Lance Zholzera
  ↪ that he blocked the headlights of a car at a stop, leading to a
  ↪ road rage that claimed her life. Joshua Hershehey is accused of
  ↪ drinking while he tried to drive.

The unidentified unidentified woman was accused of driving drunk in a
  ↪ green engine parked near a stop. She is expected to die from head
  ↪ injuries. A motion is set to dismiss a Virginia man accused of a
  ↪ drunk driving on a woman.

New video shows members of the Lady Gaga wearing a T-shirt and eating
  ↪ to a childrens table in London. A video has drawn widespread
  ↪ attention after members of the Lady Gaga made a nasty anti-social
  ↪ gesture at a recent all-girl celebration.

It shows a band member wearing a black T-shirt sporting a pair of black
  ↪  trousers and a high-fitting shirt eating to a childrens table.
  ↪ Facebook Twitter Pinterest The Lady Gaga is booed after eating at a
  ↪  restaurant at St Mary Pauls school.

A post by a member of the Lady Gaga production team on the channels
  ↪ school music programme, Just Say Never Say Or, said a band member
  ↪ wearing a black T-shirt was making a nasty gesture: This video is a
  ↪  disturbing picture of children.
```

Text obtained with 64 NFEs distilled by DiDi-Instruct (2/3). Perplexity=19.61, Entropy=5.18.

```
Iran is to meet with Hassan Rouhani to signal the future of her
  ↪ relations with the United States and its corridors of power. The
  ↪ pair, Javad Zarifs sister-in-law, diplomat and foreign minister,
  ↪ are scheduled to hold a bilateral meeting Friday.

The meeting follows a round of talks between the two countries that
  ↪ portrayed Rouhani the Iranian-backed foreign minister until
  ↪ recently as outside the realm of possibility in the war-torn region
  ↪ . The proposals hard-point proposal focused on hard-line Irans
  ↪ demands.

The US has been attempting to shore up the support of Rouhanis Iran,
  ↪ which has called for Assad to attend the round of talks between the
  ↪  two countries. At the end of an earlier round of talks in Moscow,
  ↪ Washington has urged reconsideration.

The two met at the Saint-Petersburg resort of Seychell but did not have
  ↪  a ticket. A Kremlin spokesman said Moscow had not yet considered
  ↪ the possibility. This is a real open bilateral dialog, Iran will be
  ↪  one of the participants of this dialogue.

Iran is to meet with Rouhani and Hassan Rouhani to signal the future of
  ↪  her Arab countrys relations with the United States and its allies.
  ↪  The proposed two-point communique focused on hard-line Irans
  ↪ demands that Moscow end the violence between Assad and Tehran.

Assad had reportedly visited Moscow a couple of weeks ago, in their
  ↪ first meeting in 20 months after a Paris 2015 diplomatic
  ↪ breakthrough. The proposed communique said Assads calls to end
  ↪ violence in Syria would benefit Iran, an issue that has already
  ↪ drawn ire.

Japans Prime Minister Yuki Saitan has said the government had been
  ↪ conducting regular inspections of the Sakouba nuclear power plant
  ↪ since the presidential election in 2014. Vowing to use military
  ↪ force as the right thing to do to repair a reactor.

The Sakouba nuclear power plant is owned by the government and he and
  ↪ his family have had access to power supplies a day without the
  ↪ presidential election. Though Japan has said armed forces can be
  ↪ led to repair repairs at the state-run power plant.

Sakoubas president, whose government has been investigating, defended
  ↪ it. Japan is using military force as the right thing to do
  ↪ following the accident at the power plant, Sakoubas prime minister,
  ↪  Yuki Saitan, said. We intend to use the accident as an opportunity
  ↪ .
```

Text obtained with 64 NFEs distilled by DiDi-Instruct (3/3). Perplexity=19.61, Entropy=5.18.

```
A newly leaked text from a Russian government that signed security
  ↪ guarantees for Russian President Vladimir Putins sovereignty when
  ↪ he met U.S. law enforcement officials, Russian media, Reuters
  ↪ reported Monday. It is part of a Soviet-era security deal between
  ↪ Moscow and Washington.

It is part of a newly signed security agreement between the Russian
  ↪ President Vladimir Putin and Washington for the seizure of the
  ↪ Baltic Sea, in exchange for the U.S. guarantees, Zakharova said. A
  ↪ leaked text from the government says Russia forfeited its
  ↪ sovereignty rights.

The deal is part of the U.S. sanctions against Russia and Russia that
  ↪ collapsed last year after the Trump administration dumped hundreds
  ↪ of millions of tons of material in the disputed Baltic Sea region,
  ↪ a Russian diplomat said. Reuters.

It is the first time that Moscow agreed to the U.S. sanctions in place
  ↪ that combined prevent millions of tons of trash moved out of the
  ↪ Russian territory, according to reports by Reuters. The arrangement
  ↪ , which is part of the U.S. sanctions.

Scientists working with an artificial intelligence company called
  ↪ Cavespace to build artificial intelligence within game studios have
  ↪  found that peoples eyes were significantly more white and dark
  ↪ than they were before while playing a video game. They are working
  ↪ towards this goal.

By advancing their work on the human eye, AI scientists will be able to
  ↪  help game developers creating artificial intelligence experiences
  ↪ within games. Theyre currently working in conjunction with an AI-
  ↪ focused research team within the tools division, which hopes to
  ↪ begin developing a human eye.

Since the start of the decade researchers have been working to improve
  ↪ some of the most popular console experiences, such as Minecraft and
  ↪  Twitch. In fact, AI scientists at the Heidgen Institute have
  ↪ recently found that Moores law continues to play a significant role
  ↪ .

As of this writing, Electronic Arts has almost exclusively announced
  ↪ the development of AI research within games, but recent reports,
  ↪ including by the Guardian newspaper, have suggested that the
  ↪ research will focus on the companys first game, Pokemon. That
  ↪ information remains to be seen.

PENNAI, China IBM and Hewlett-Pack expressed hopes over docking in
  ↪ space companies next-generation computers during a joint press
  ↪ conference on Dec. 5. The companys chairman Chongq Iardai has
  ↪ developed a docking system to improve space companies next-
  ↪ generation computer systems.

A consortium of Hewlett-Packard, including Intel and IBM has created a
  ↪ new wave of docking systems, which could improve the companys
  ↪ current-generation computer systems. In the interview, Chongq
  ↪ Iardai also provided detailed information on technical matters.
```

Text obtained with 128 NFEs distilled by DiDi-Instruct (1/3). Perplexity=17.50, Entropy=5.17.

```
Officials at the U.S. District of Virginia claim a "fatality epidemic"
  ↪ is causing deaths in school classrooms around the U.S. The Virginia
  ↪  school district is investigating the death of student Tamey Nelson
  ↪  after she wielded a knife to kill a classmate in class.

A school in Virginia is investigating after state claims that a "
  ↪ fatality epidemic death" is causing deaths among students in
  ↪ classrooms around the world. Officials at the U.S. district
  ↪ disputed the school's claims of an outbreak.

Two students were killed and three others were hospitalized, according
  ↪ to a news report obtained by NBC News. "I am concerned about the
  ↪ safety of children and students in classrooms in the United States
  ↪ ," the District of Virginia District Attorney James Trotsma told
  ↪ NBC News.

Parents of a Virginia school claim that a "child's life of death" was
  ↪ caused by an epidemic of abuse in classrooms around the U.S.
  ↪ district on Sunday, according to a news report. A video surfaced of
  ↪  a 12-year-old student being struck and killed.

The student was killed by a knife wielded by others they identified as
  ↪ "revered students," a similar incident described by a teacher that
  ↪ was caused by the same incident in class at the school's Iron
  ↪ Arlington Branch Learning Center.

Officials at the district, which claim that there's a systemic risk to
  ↪ both children and students, said two students were killed and three
  ↪  others were hospitalized. The incident has drawn international
  ↪ attention over claims of a child's life being caused by an epidemic
  ↪  of abuse.

The abuse was reported in classrooms around the United States and
  ↪ abroad. In the week, a video surfaced of Virginia student Tamey
  ↪ Nelson, who was struck and killed by a knife wielded by classmates
  ↪ in a hospital a day after an incident in class.

Officials at the U.S. District of Virginia alleged that a similar
  ↪ incident took place at the Iron Arlington Branch Learning Center
  ↪ where Tamey Nelson was held more than a decade ago. This raised
  ↪ tensions with the White House and Donald Trump.

The Kremlin's meeting is one of the first links between the White House
  ↪  and Trump, which include Moscow's relationship with the president-
  ↪ elect. It was reported in December 2016 that the meeting was amid a
  ↪  strained relationship between Trump and the Kremlin.

The Kremlin's meeting is just part of a strained relationship between
  ↪ the White House and Trump and it's a flashpoint in the US
  ↪ presidential race. Several reports claim that Moscow's role is
  ↪ keeping them apart.
```

Text obtained with 128 NFEs distilled by DiDi-Instruct (2/3). Perplexity=17.50, Entropy=5.17.

```
The Kremlin is busy protecting some key US diplomats and attempting to
  ↪ undermine the president-elect. The Moscow meeting is one of the
  ↪ first links between Russia and the new administration, and it will
  ↪ be a front-point for discussions.

In December 2016, reports suggested the meeting and the White House was
  ↪  the brainchild of Moscow leading up to the US election. Moscow's
  ↪ role is keeping the White House and the new administration apart,
  ↪ creating a complex diplomatic situation.

The Kremlin's meeting is yet another sign of a strained relationship.
  ↪ However, Monday's meeting between the US and Russian presidents
  ↪ will be a front-point in the White House. The meeting is just part
  ↪ of mounting tensions between the two nations.

These tensions include the pick of the United States' national security
  ↪  adviser. The Kremlin's role is reportedly being investigated and
  ↪ it is busy protecting some of its key US diplomats. Both leaders
  ↪ have known each for a couple of years now.

The U.R. government's highest civilian official on Tuesday said that
  ↪ despite the U.S. government's push for significant violence along
  ↪ Syria's southern border, the United States would already lose
  ↪ control of large parts of Deraa, a region of interest.

If peace advocates fail, even without Washington's major troop presence
  ↪  on the southern border, then the U.S. would push to lose control
  ↪ of Syria's southernmost border. This is a region where the Syrian
  ↪ government's main interest will be its ammunition.

American forces could play a significant influence on the Syrian rebels
  ↪  and President Bashar al-Assad himself, but officials have sought
  ↪ to put the brakes on the nearly six-year war in Syria. This
  ↪ conflict has been ongoing for a significant time.

Should the U.S.-led coalition in Syria fail to gain control of large
  ↪ parts of the southern Syrian city of Deraa, the regime's heavy
  ↪ weapons and ammunition would control the southern border -- a power
  ↪  officials have said is unlikely because of the opposition.

American forces could play a significant influence in the Syrian
  ↪ conflict, especially those of Assad himself, but officials have
  ↪ sought to put the brakes on the six-year war in Syria, where
  ↪ President Bashar Assad's government is a major player.

The EPA also said that it had watched samples of fuel from India's
  ↪ nuclear plant on Monday and that it found no such leak and that
  ↪ Indian military installations were not damaged, according to a
  ↪ court filing on Monday morning.
```

Text obtained with 128 NFEs distilled by DiDi-Instruct (3/3). Perplexity=17.50, Entropy=5.17.

```
Washington has also intensified its fight against Islamic State along
  ↪ the southern border. The U.S. is setting up a special task force to
  ↪  protect sensitive areas along the Syrian border, which Damascus
  ↪ and its allies consider to be a key ally.

Ahead of a major inspection of India's nuclear facilities Tuesday,
  ↪ people are watching samples of fuel from India's biggest nuclear
  ↪ plant, just a day after the US made claims that military
  ↪ installations may be damaged by the leak scandal.

EPA workers have released footage of what appears to be samples of fuel
  ↪  that was leaked from India last year following an inspection of
  ↪ large-scale nuclear plants. They told Reuters on Monday that there
  ↪ had not been damage to buildings from leaks.

India's Environmental Protection Agency (EPA) said the agency had been
  ↪ conducting an inspection of the country's nuclear plant following
  ↪ what appears to amount to Tuesday's claim that Indian military
  ↪ installations may be damaged by the leak scandal from last year.

An official at the US attorney general's office reduced any claim of
  ↪ any damage from leaks at the 10-year plant to unclear due to no
  ↪ available information on Tuesday. The incident comes just a day
  ↪ after the US claim of the fuel leak.

The footage of the plant from India, however, told Reuters on Monday
  ↪ that it was unclear whether damage to buildings from leaks had been
  ↪  reported and there was no available information on the incident. A
  ↪  senior EPA official confirmed the claim was murky.

The U.R. government's highest civilian official on Tuesday held a
  ↪ briefing on Syria. White House Press Secretary Sarah P. Sanders
  ↪ said on Tuesday that even without Washington's major troop presence
  ↪ , the U.S. would lose control of much of the southern border.

The claim suggested Indian military installations may be damaged from
  ↪ material related by Republican lawmakers to Russian interference in
  ↪  the last United States election. This adds another layer of
  ↪ complexity to the ongoing international political situation.

In its first foray into a market that created top-end graphics, Intel
  ↪ announced it plans to launch a next-generation graphics chip that
  ↪ is responsible for powering the personal computer's power system.
  ↪ The announcement was made Monday during a joint press conference.

The conference was held by Las Vegas-based Intel and Florida-based
  ↪ technology company Ion Graphics. The new graphics chip, which is
  ↪ slated for a 2017 release, powers the Windows operating system and
  ↪ is expected to be a major market player.
```