Prompt Guided Diffusion for Controllable Text Generation

Mohaddeseh Mirbeygi and Hamid Beigy
Department of Computer Engineering
Sharif University of Technology
Tehran, Iran
m.mirbeygi@sharif.edu and beigy@sharif.edu

Abstract

Text generation under control, or producing linguistically coherent and contextually relevant text, has seen tremendous progress thanks to methods based on PPLM, FUDGE, and diffusion-based models. Yet current state-of-the-art models tend to balance control fidelity with fluency. In addition, classifier-guided strategies (e.g., PPLM) can be predicted in gradient updates providing less coherent text. In contrast, autoregressive-based approaches (e.g., FUDGE) rely on inflexible generation patterns that limit creativity. Recent diffusion methods demonstrate superior performance in iteration and diversity, but indirect methods often fail to introduce sufficient ways to inject task-associated knowledge, leading to the need for many different complex classifier modules during both training and inference. To address this, we introduce a prompt-guided diffusion framework that seamlessly incorporates structured prompts into the diffusion steps, providing precise and flexible control of the generated text. Each prompt combines a target attribute (for example, a sentiment tag), an example corresponding to that label (for example, a positive review), and a slot for the generated sentence. By encoding such prompts using large pre-trained models (such as BART) and integrating these prompts through cross-attention into the diffusion dynamics, our model achieves new state-of-the-art performance on a variety of tasks ranging from IMDB for sentiment, AG-News for topic, and E2E for structured-output to text.

1 Introduction

Text generation: a computational paradigm for producing meaningful written content with coherence, often fueled by NLP models. Its uses include chatbots, content generation systems, machine translation, and other areas. Controllability in text generation concerns the ability to control the outputs for desired characteristics — including tone, style, length, or topic based on predefined criteria or user preference. This is usually done through all sorts of means, from prompt engineering to fine-tuning or control tokens. Unconstrained generation, on the other hand, refers to cases where the generated content deviates from the requirements, resulting in out-of-topic content. Such deviations (known informally as use performance error) are common due to the inherently random nature of sampling or subtle modeling of user intention, additional work is often needed in production to find a satisfactory balance between controllability and creativity in the model.

NLP tasks related to text generation generally relate to a generation task where models attempt to create a set of coherent, meaningful strings from some input, based on generative architectures. Researchers have developed various types of generative strategies. Generative Adversarial Networks (GANs) compete against a discriminator to generate text samples. EBMs work by defining an energy function across the text data, with the model trained to produce lower energy for valid samples and higher energy for invalid samples. This allows for a flexible way to enforce constraints during the generation process. Flow-based models produce exact likelihoods by invertible mapping from simple probability distributions to complex ones, giving much more control. Diffusion models progressively synthesize outputs, denoising random noise through multiple probabilistic steps, yielding stable and high-quality results. These paradigms together illustrate the spectrum of mechanisms for text generation by arranging different trade-offs between controllability, diversity, and fidelity. This paper focuses on the application of diffusion models to the task of text generation.

1.1 Diffusion Model

The diffusion model consists of a Markov chain of unobservable quantities. It begins with an initial

Proceedings of the Tenth Workshop on Noisy and User-generated Text, pages 78–84

data point x_0 and incrementally corrupts it with Gaussian noise until x_T, according to the posterior $q(x_{0:T} \mid x_0)$. The variables $x_0, \ldots, x_T$ have the same dimensionality as x_0. The main objective is to model the distribution $p_\theta(x_{t-1} \mid x_t)$ for the reverse (denoising) process Ho et al. (2020).

Forward and reverse processes are two key components of a diffusion model. The forward process gradually corrupts data with random noise until it is practically indistinguishable from pure noise. Then the reverse phase tries to reconstruct the original data, learning to deduce how to remove the noise step by step. In the forward process, the transitions in the Markov chain are described by a conditional Gaussian. The generative distribution can be expressed as 1.

$$q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}) = \prod_{t=1}^{T} \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \quad (1)$$

where every β (fixed or learnable) controls the variance. As the time T becomes larger, the second assumption states that x_T approaches Gaussian noise.

The model learns the reverse path during training to sample data from random noise $p(x_T) = \mathcal{N}(x_T; 0, I)$ and thereby learns $p_\theta(x_{0:T})$ as in equation 2.

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t) = p(x_T) \prod_{t=1}^{T} \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t), \sum_\theta(x_t, t)\right) \quad (2)$$

In this Markov chain, we model the dependence on time of the reverse distribution by 3.

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t), \sum\nolimits_\theta(x_t, t)\right) \quad (3)$$

The training aims to maximize the likelihood, which is mathematically equivalent to minimizing the negative log-likelihood by equation 4.

$$E[-log p_\theta(x_0)] \leq E_q[-log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}] = L_{vlb} \quad (4)$$

The KL divergence for Gaussians means that the losses at every step (Equations 5 to 8, above) can be expressed in KL terms. Therefore, the total loss is the sum across the chain:

$$L_{vlb} = \sum_{k=0}^{T} L_k \quad (5)$$

$$L_0 = -\log p_\theta(x_0|x_1) \quad (6)$$

$$L_{t-1} = D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t)) \quad (7)$$

$$L_T = D_{KL}(q(x_T|x_0)||p(x_T)) \quad (8)$$

2 Related Work

Several studies focused on the adaptation of diffusion models originally developed for image generation to the discrete textual domain (Li et al., 2022; Austin et al., 2021). They provide novel methods to approach the problem of continuous diffusion processes versus discrete tokens. Some studies directly construct diffusion as defined in the discrete space and others map the discrete tokens into a continuous representation where standard diffusion pipelines can work (Austin et al., 2021; Chen et al., 2022). Savinov et al. (2021) shows how iterative denoising autoencoders can be placed in a diffusion context and how repeated denoising steps approximate the generative capacity of diffusion. Such approaches can serve as a complement to the widespread autoregressive models held typical for text generation, enabling improvements in controllability, diversity, and sophisticated dependency modeling.

These lines of research also investigate how to use or outgrow diffusion-based text generation. Diffusion-LM Li et al. (2022), which focuses on controlling attributes of generated text, e.g., sentiment domain. Gong et al. (2022) leverages diffusion models for seq2seq tasks like translation, indicating their generality. In summary, this line of work broadens the applicability of diffusion models beyond the domain of continuous data, paving new pathways into how discrete textual outputs can be generated and conditioned.

Diffusion-LM Li et al. (2022) proposes a new paradigm for text generation by utilizing the iterative refinement framework of diffusion models, which has been traditionally used in the setting of continuous data, directly on text tokens. Rather than single-pass autoregressive generations, Diffusion-LM improves the text in multiple passes, potentially leading to greater flexibility and variability in its outputs. This approach trains a token-level denoiser, allowing the approach to modify specific attributes (sentiment, length, etc.) at inference without needing additional retraining.

D3PM Austin et al. (2021) presents a method for diffusion on structured discrete data (e.g. text, categorical data). In this technique, a forward corruption process preserves structural relations, and a reverse denoising process restores the corrupted data in iterations by learning a structured probability form. This approach aims to go beyond the limitations of traditional sequential text generation formats, enabling novel forms of discrete data modeling.

SUNDAE Savinov et al. (2021) proposed a new model of text generation combining the structured representation power of denoising autoencoders with a particular set of step-unrolling techniques in modeling the sequential dependency of text. Next, while standard DAEs generate text in a single step, this framework replays the generation process over an extended range, advising the denoising process on how to convert an embedding with noise into intelligible text. Step Unrolling: This allows the model to learn to incorporate more information about longer context dependencies while enforcing a schism between the input and output of the model, resulting in better-generated text.

There are, however, other utilitarian efforts that use encoder-decoder architectures, with latent representations, with a strong application-oriented motivation. Liu et al. (2024) establishes a generalized view of diffusion, which can be applied to data across continuous or discrete domains since both the encoder and decoder may be tailored. Tan et al. (2023) presents an encoder-decoder breakup for text diffusion, specifically comprising a spiral interplay structure that expands generational high quality, whilst letting knowledge waft from their encoder to its decoder (throughout the diffusion levels).

These papers cover various techniques for further improving text generation using diffusion models by generally combining PLMs, latent spaces, and novel training or sampling techniques. Several of them focus on the synergy between diffusion and PLM. The proposed approach, Ou and Jian (2024) suggests a "linguistic easy-first schedule" to guide the process of diffusion in leveraging linguistic knowledge and PLMs to make the model generate simpler linguistic structures first.

Many studies explored the combination of diffusion with large pre-trained language models (PLMs). Ou and Jian (2024) introduce a “linguistic easy-first schedule” that borrows from linguistic knowledge and leverages PLMs so that simpler patterns first appear in diffusion-based text generation. Chen et al. (2023a) present a resource-frugal diffusion language model with soft-masked noise, which strikes an equilibrium by preserving essential linguistic elements.

The domains where diffusion models can be applied include paraphrasing Zou et al. (2024), dialog systems Xiang et al. (2024), recommendation engines Li et al. (2023), code generation Singh et al. (2023), topic modeling Xu et al. (2023), event argument extraction Luo and Xu (2023), comment generation Liu et al. (2023), style transfer Horvitz et al. (2024), Lyu et al. (2023), key phrase extraction Luo et al. (2023), translation Chen et al. (2023b), poetry generation Hu et al. (2024), text detoxification Floto et al. (2023), empathetic dialog Bi et al. (2023), entity recognition Shen et al. (2023), text summarization Zhang et al. (2023), text inference Yuan et al. (2024), and conversation controllable Chen and Yang (2023).

3 The Proposed Method

Prompt diffusion is an emerging key mechanism for generative modeling, providing a simple yet powerful way to condition outputs of a diffusion model with standard language prompts. While diffusion models have been shown to be powerful samplers (from images to audio to text), achieving explicit, fine-grained control has remained a challenge Nichol and Dhariwal (2021). This is what makes direct control over the generative process and steering it toward certain outputs complex.

To address this problem, diffusion strategies based on prompts condition the diffusion model on text descriptions (i.e., “prompts”) that describe the desired properties and guide the generative process of the model. Essentially, a big pre-training language model, e.g. BART, is applied to these textual prompts to turn them into vector representations that contain the prompt’s semantic content. These representations are then fed into the denoising network, often using concatenation methods, guiding concerning the prompt during each step of the denoising process.

One particular type of structured prompt uses the target property, such as a sentiment or topic, along with a randomly selected in-class example (to prevent data overlap or leakage), then leaves a blank for the new sentence. Our diffusion model, which is based on a transformer, manages the noisy text embeddings with cross-attention conditioning on the embeddings of prompt processed by BART (or any other similar encoder). This design is shown in figure 1.

Several advantages come with prompt-based diffusion. First, it is highly controllable Sridhar and Vasconcelos (2024): with meticulously engineered prompts, one can dictate the style, content, or other types of attributes, allowing for highly constrained creative output Zhong et al. (2024). Second, it offers versatility: a single pre-trained diffusion model can be used on many tasks by simply changing a prompt instead of fine-tuning each model for each new objective. It is extremely cost-efficient. Third, it offers great potential for few-shot or in-context learning, allowing the model to infer instructions from a few examples Du et al. (2024).

But writing good prompts is not trivial: badly written prompts give bad results, and the encoding of prompts also takes time to generate. Complex prompts may raise challenges as well in terms of coherence. Despite that, prompt-based diffusion is an appealing method, as it provides extensive user-driven guidance combined with the powerful generative capability of extensive diffusion models.

Our system combines a large language model and diffusion for conditional text generation. We adopt a prompt-learning paradigm that concatenates the condition label (e.g., sentiment) with a relevant example review to form a textual prompt. Subsequently, this prompt is encoded (e.g., with BART), bringing about embeddings that guide the diffusion process. The diffusion model is trained to predict the noise added at each time step, effectively modeling the reverse diffusion. Therefore, at inference time, random noise is iteratively converted into meaningful embeddings based on the guidance of the prompt.

So those final embeddings get passed through a BART decoder, benefiting from the pre-trained autoregressive decoding to produce reasonable text. This pipeline elegantly resolves the shortcomings of a completely embedding-based decoding (which can be compelled to revert to rough nearest-neighbor lookups) and produces high-quality text outputs. The prompt method informs the output using the desired condition in the prompt but also dictates the context with the example text, which provides a granular steer for what you would want to get as output. By leveraging the strengths of a sizable, pre-trained encoder-decoder (BART) and a purpose-built diffusion model, the components are tailored for their tasks, yielding robust performance on conditional text generation tasks.

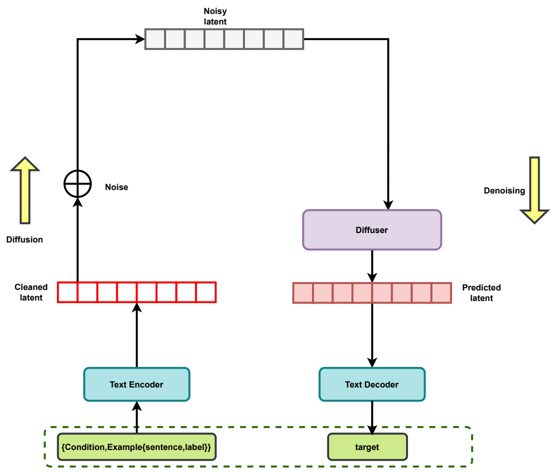

Figure 1: Our Proposed Method

4 Experimantal results

We evaluate our approach on three benchmark datasets: IMDB (50,000 movie reviews for sentiment analysis) Maas et al. (2011), AG News (120,000 news articles across four topics: World, Sports, Business, and Sci/Tech) Zhang et al. (2015), and E2E (a data-to-text dataset which involves restaurant descriptions using structured attribute-value pairs) Novikova et al. (2017). Each dataset poses different challenges: IMDB for sentiment polarity, AG News for topic coherence, and E2E for structured semantic fidelity.

Our diffusion training involves two steps: a forward phase where the text embeddings are gradually contaminated with Gaussian noise across T=1000 time steps, and a reverse phase, during which a trained model denoises the signal to reconstruct the data. In the forward step, x_0 progressively transforms into x_T according to a noise schedule $\beta_t = 0.9$, ending with a nearly random noise state (Equation 1). The model is trained to predict the noise at each time t concerning the variational bound (Equation 4), estimating $p_\theta(x_{t-1} \mid x_t)$. Diffusion-based generators instantiate text through a multi-step, iterative denoising process, allowing fine-grained modifications during intermediate steps to satisfy conditions such as syntactic or stylistic properties instead of generating the text token-by-token as in autoregressive models. This iterative routine stabilizes training and provides more robust controllability than single-pass methods.

Our experiments were conducted using the E2E dataset, and the results showed that the proposed method outperformed other approaches. Table 1 lists three generation tasks we experimented with: semantic content, parts-of-speech (POS), and length.

For the semantic content task, we supplied a field (e.g., rating) and provided a value (e.g., 5 stars) to compute a sentence that would accurately personify the relationship between the field and value provided, and the ground truth for the same being the exact match of the 'value'. For the parts-of-speech task, we generated a sequence of POS tags to the model (e.g., Pronoun Verb Determiner Noun) and asked it to output a word sequence of the same length such that the POS tags were aligned with the target according to an oracle POS tagger. Success was measured using word-level exact matches. For the length task, we defined a desired length between 10 and 40 and produced a sequence of up to ±2 the target length.

We also conduct experiments on IMDb and AG-News, assessing their quality using metrics such as BLEU, ROUGH, and BERTScore as shown in table 2.

The numerical results in Table 1 clearly show that PromptDiffusion outperforms all prior controllable text generation methods on all features evaluated: accuracy of semantic content, accuracy of part-of-speech, and text length. In semantic content, PromptDiffusion also delivers high accuracy of 83% outperforming the previous leading methods including Masked DiffusionLM + BERT (82.9%) and DiffusionLEF + BERT (82.4%), without sacrificing the low-perplexity (2.30) and thus fluency. And while achieving 92.5% in part-of-speech accuracy, PromptDiffusion outperforms all other diffusion models and has the lowest perplexity, at 4.7, which means it generates syntactically more coherent outputs.

Table 2 reveals that PromptDiffusion outperforms not only PPLM and FUDGE but also DiffusionLM on the IMDB (sentiment control) and AG News (topic control) datasets in terms of generation quality. In extensive evaluation, of the IMDB dataset, PromptDiffusion achieves BLEU-4 of 10, ROUGE-L of 30, and BERT-Score of 92, outperforming DiffusionLM and GPT-2 by a large margin. This indicates that PromptDiffusion yields semantically more aligned and fluent text while better-preserving intent. These findings further underpin that PromptDiffusion provides a tradeoff between controllability, fluency, and quality, thus is a strong competitive to prior generation methods reaping benefits from traditional structured pretrained models (e.g. GPT-2), and because it also surpasses them in some tasks in text generation.

Diffusion models give controllable text generation more flexibility and come with significantly more advantages than autoregressive, VAE, or GAN-based approaches. While autoregressive models like GPT predict following a static, token-by-token order, diffusion models slowly guide latent representations through many iterations. Such progressive denoising lends itself well to making subtle tweaks to fit our constraints, such as syntax, length, or style. Diffusion models manage to strike the right balance between accuracy and creativity in comparison to classifier-guided techniques like PPLM, which tend to generate unintelligible outputs owing to erratic updates of gradients, or VAEs that in most cases hit a wall when it comes to diversity. Diffusion models achieve a balance for generation by inserting structured prompts (e.g. target attributes) into continuous input via cross-attention mechanisms without losing fluency.

5 Conclusion

In this work, we propose a prompt-guided diffusion framework for controllable text generation that mitigates critical limitations of the existing methods in balancing precision and fluency. Our method integrates structured prompts that combine target conditions and in-class examples into the diffusion process, achieving fine-grained control over attributes such as sentiment, topic, and adherence to structured data. Dynamic sampling of examples during training ensures robustness to intra-class diversity. Future work might investigate hybrid models that combine the proposed prompt-guided diffusion either with the retrieval-augmented generation or few-shot learning, as well as an extension to multimodal tasks. Such a framework advances the frontier of controllable text generation by bridging human intention with generative AI through intuitive prompting and thus offers a flexible and scalable solution for real-world deployment.

	semantic content		part of speech		length	
	Acc	Perp	Acc	Perp	Acc	Perp
PPLM	9.9	5.32	-	-	-	-
FUDGE	69.9	2.83	27	7.96	46.9	3.11
DiffusionLM	81.2	2.55	90	5.16	99.9	2.16
DiffusionLM + Bert	77.4	2.68	86.2	5.43	99.9	2.68
Masked DiffusionLM + Bert	82.9	2.30	92.9	4.78	100	2.08
DiffusionLEF	81.7	2.46	91.2	5.09	99.9	2.14
DiffusionLEF + Bert	82.4	2.32	92.4	4.82	**100**	2.10
PromptDiffusion	**83**	**2.30**	**92.5**	**4.7**	99.9	**2**

Table 1: results on E2E dataset for controllable generation

	IMDB			AG News		
	BLEU-4	ROUGE-L	Bert-Score	BLEU-4	ROUGE-L	Bert-Score
PPLM	1.6	19	41	2	20	43
FUDGE	1.8	20	43	2.1	22	46
DiffusionLM	7	28	89	7.5	29	90
PromptDiffusion	**10**	**30**	**92**	**11**	**31**	**91**
GPT2	6.1	26	88	6.8	27	89

Table 2: results on IMDB and AGnews

References

Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg. 2021. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems*, 34:17981–17993.

Guanqun Bi, Lei Shen, Yanan Cao, Meng Chen, Yuqiang Xie, Zheng Lin, and Xiaodong He. 2023. Diffusemp: A diffusion model-based framework with multi-grained control for empathetic response generation. *arXiv preprint arXiv:2306.01657*.

Jiaao Chen and Diyi Yang. 2023. Controllable conversation generation with conversation structures via diffusion models. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7238–7251.

Jiaao Chen, Aston Zhang, Mu Li, Alex Smola, and Diyi Yang. 2023a. A cheaper and better diffusion language model with soft-masked noise. *arXiv preprint arXiv:2304.04746*.

Linyao Chen, Aosong Feng, Boming Yang, and Zihui Li. 2023b. Xdlm: Cross-lingual diffusion language model for machine translation. *arXiv preprint arXiv:2307.13560*.

Ting Chen, Ruixiang Zhang, and Geoffrey Hinton. 2022. Analog bits: Generating discrete data using diffusion models with self-conditioning. *arXiv preprint arXiv:2208.04202*.

Yingjun Du, Gaowen Liu, Yuzhang Shang, Yuguang Yao, Ramana Kompella, and Cees GM Snoek. 2024. Prompt diffusion robustifies any-modality prompt learning. *arXiv preprint arXiv:2410.20164*.

Griffin Floto, Mohammad Mahdi Abdollah Pour, Parsa Farinneya, Zhenwei Tang, Ali Pesaranghader, Manasa Bharadwaj, and Scott Sanner. 2023. Diffudetox: A mixed diffusion model for text detoxification. *arXiv preprint arXiv:2306.08505*.

Shansan Gong, Mukai Li, Jiangtao Feng, Zhiyong Wu, and LingPeng Kong. 2022. Diffuseq: Sequence to sequence text generation with diffusion models. *arXiv preprint arXiv:2210.08933*.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851.

Zachary Horvitz, Ajay Patel, Chris Callison-Burch, Zhou Yu, and Kathleen McKeown. 2024. Paraguide: Guided diffusion paraphrasers for plug-and-play textual style transfer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 18216–18224.

Zhiyuan Hu, Chumin Liu, Yue Feng, Anh Tuan Luu, and Bryan Hooi. 2024. Poetrydiffusion: Towards joint semantic and metrical manipulation in poetry generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 18279–18288.

Ling Li, Shaohua Li, Winda Marantika, Alex C Kot, and Huijing Zhan. 2023. Diffusion-exr: Controllable review generation for explainable recommendation via diffusion models. *arXiv preprint arXiv:2312.15490*.

Xiang Li, John Thickstun, Ishaan Gulrajani, Percy S Liang, and Tatsunori B Hashimoto. 2022. Diffusion-lm improves controllable text generation. *Advances in Neural Information Processing Systems*, 35:4328–4343.

Guangyi Liu, Yu Wang, Zeyu Feng, Qiyu Wu, Liping Tang, Yuan Gao, Zhen Li, Shuguang Cui, Julian McAuley, Eric P Xing, et al. 2024. Generating, reconstructing, and representing discrete and continuous data: Generalized diffusion with learnable encoding-decoding. *arXiv preprint arXiv:2402.19009*.

Jiamiao Liu, Pengsen Cheng, Jinqiao Dai, and Jiayong Liu. 2023. Diffucom: A novel diffusion model for comment generation. *Knowledge-Based Systems*, 281:111069.

Lei Luo and Yajing Xu. 2023. Context-aware prompt for generation-based event argument extraction with diffusion models. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 1717–1725.

Yuanzhen Luo, Qingyu Zhou, and Feng Zhou. 2023. Enhancing phrase representation by information bottleneck guided text diffusion process for keyphrase extraction. *arXiv preprint arXiv:2308.08739*.

Yiwei Lyu, Tiange Luo, Jiacheng Shi, Todd C Hollon, and Honglak Lee. 2023. Fine-grained text style transfer with diffusion-based language models. *arXiv preprint arXiv:2305.19512*.

Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150, Portland, Oregon, USA. Association for Computational Linguistics.

Alexander Quinn Nichol and Prafulla Dhariwal. 2021. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR.

Jekaterina Novikova, Ondřej Dušek, and Verena Rieser. 2017. The E2E dataset: New challenges for end-to-end generation. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 201–206, Saarbrücken, Germany. Association for Computational Linguistics.

Yimin Ou and Ping Jian. 2024. Effective integration of text diffusion and pre-trained language models with linguistic easy-first schedule. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5551–5561.

Nikolay Savinov, Junyoung Chung, Mikolaj Binkowski, Erich Elsen, and Aaron van den Oord. 2021. Step-unrolled denoising autoencoders for text generation. *arXiv preprint arXiv:2112.06749*.

Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. 2023. Diffusionner: Boundary diffusion for named entity recognition. *arXiv preprint arXiv:2305.13298*.

Mukul Singh, José Cambronero, Sumit Gulwani, Vu Le, Carina Negreanu, and Gust Verbruggen. 2023. Codefusion: A pre-trained diffusion model for code generation. *arXiv preprint arXiv:2310.17680*.

Deepak Sridhar and Nuno Vasconcelos. 2024. Prompt sliders for fine-grained control, editing and erasing of concepts in diffusion models. *arXiv preprint arXiv:2409.16535*.

Chao-Hong Tan, Jia-Chen Gu, and Zhen-Hua Ling. 2023. Diffusia: A spiral interaction architecture for encoder-decoder text diffusion. *arXiv preprint arXiv:2305.11517*.

Jianxiang Xiang, Zhenhua Liu, Haodong Liu, Yin Bai, Jia Cheng, and Wenliang Chen. 2024. Diffusiondialog: A diffusion model for diverse dialog generation with latent space. *arXiv preprint arXiv:2404.06760*.

Weijie Xu, Wenxiang Hu, Fanyou Wu, and Srinivasan Sengamedu. 2023. Detime: Diffusion-enhanced topic modeling using encoder-decoder based llm. *arXiv preprint arXiv:2310.15296*.

Shilong Yuan, Wei Yuan, Hongzhi Yin, and Tieke He. 2024. Roic-dm: Robust text inference and classification via diffusion model. *arXiv preprint arXiv:2401.03514*.

Haopeng Zhang, Xiao Liu, and Jiawei Zhang. 2023. Diffusum: Generation enhanced extractive summarization with diffusion. *arXiv preprint arXiv:2305.01735*.

Xiang Zhang, Junbo Zhao, and Yann LeCun. 2015. Character-level convolutional networks for text classification. *Advances in neural information processing systems*, 28.

Linhao Zhong, Yan Hong, Wentao Chen, Binglin Zhou, Yiyi Zhang, Jianfu Zhang, and Liqing Zhang. 2024. User-friendly customized generation with multi-modal prompts. *arXiv preprint arXiv:2405.16501*.

Wei Zou, Ziyuan Zhuang, Shujian Huang, Jia Liu, and Jiajun Chen. 2024. Enforcing paraphrase generation via controllable latent diffusion. *arXiv preprint arXiv:2404.08938*.