

From Diverse Outputs to Structured Set: Semantic Planning Enables Parallel Advertising Text Generation

Penghui Wei and Jiayu Wu

Baidu Inc.

weipenghui2016@gmail.com, aruiwu9@gmail.com

Abstract

An advertising platform typically generates a diverse candidate set of text creatives for each ad and then employs personalized online ranking to estimate the top-performing one for final display. Conventionally, the diversity is treated as a by-product during the stochastic decoding process rather than a training objective, which results in suboptimal outputs due to the lack of explicit modeling for semantic coverage. In this paper we reformulate ad text generation as a set-level modeling problem, and propose a parallel generation framework that optimizes the diversity objective with set-level alignment. By explicitly decomposing an ad’s content structure into semantic-independent perspectives via a planning mechanism, our framework generates the entire set of text creatives in parallel based on the planned semantics, which also takes full advantage of parallelizability to enable high-throughput industrial deployment. It is optimized through a set-level alignment stage that encourages the overall semantic coverage and penalizes the redundancy among the candidates. Comprehensive experiments on benchmark dataset show that our framework achieves better set-level quality and diversity, as well as significantly improving generation throughput, demonstrating its superiority on both generation quality and inference speed.

1 Introduction

Generating highly attractive ad text creatives is a key driver of user engagement and platform revenue in industrial online advertising, and ultimately brings significant value for advertisers. Typically, in the real-time auction process of an online advertising system, for each given ad the system first retrieves a diverse candidate set of creatives, then a personalized ranking module predicts each creative’s click-through conversion rate and selects the top-ranked creative tailored to specific user interests for further ad ranking, auction and final display,

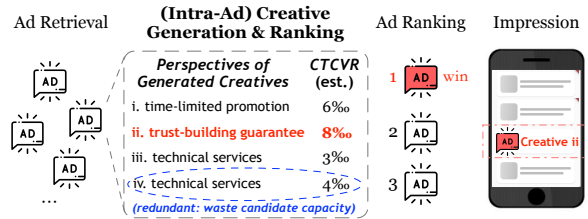


Figure 1: The location of dynamic creative optimization. The diversity and semantic coverage of the candidate text creatives are critical to final delivery performance.

as shown in Figure 1. The diversity and semantic coverage of the candidate creative set are critical: a diverse candidate set provides a higher potential ceiling for the downstream creative ranking module, ensuring that the ad resonates across varying user preferences. In the long run, diverse rather than homogeneous content helps to form a healthy content ecosystem.

Conventional approaches usually regard the diversity as a by-product of the stochastic decoding process (e.g., top- k and nucleus sampling) during inference, while the modeling paradigm still remains point-wise, i.e., optimizing the likelihood of a *singular* text creative. They suffer from limitations from both generation quality and inference efficiency in industrial production. From the quality view, without explicit supervision for *set-level* diversity, the sampled outputs frequently exhibit high lexical variance but poor semantic coverage of the ad’s various perspectives. From the deployment view, sampling multiple candidates for each input is too inefficient to meet the throughput requirement of large-scale creative generation. Although recent approaches have investigated incorporating the diversity objective explicitly via autoregressively generating each creative of a set (Wang et al., 2026) or learning to construct a latent space that facilitates diverse sampling (Wei, 2025), they still suffer from the inference efficiency issue due to the autoregressive nature or non-decoder-only architecture.

To align the diversity objective with the scalable demands of high-throughput serving, we regard the candidate set of text creatives as a whole to reformulate ad text generation as a *set-level modeling* problem, and propose an efficient framework Allegro driven by explicit semantic planning for parallel ad text generation. Instead of relying on uncertain stochastic sampling to output variations, Allegro decomposes an ad’s overall content structure into a set of distinct and semantic-independent perspectives (e.g., various advantages of functionality), where the combination of the perspectives provides comprehensive semantic coverage for the ad’s content to facilitate further diverse generation. Conditioned on the deterministic and structurally planned semantics, Allegro generates the entire creative set where all creatives in a set are generated in parallel, taking full advantage of parallelizability via inherently decoupling the generation of multiple candidates to increase inference throughput.

The framework Allegro is optimized through a set-level alignment training process, which guarantees the semantic coverage of the output set. Specifically, based on the set-level semantic planning data synthesized by large reasoning models (LRMs), the diversity is modeled at set-level during preference optimization with explicit supervision, encouraging the comprehensive semantic coverage of the generated creative set while penalizing semantic redundancy. This ensures that the creatives of a set are not only linguistically varied but semantically distinct and structurally complete for the given ad’s content structure, producing highly diverse candidates for downstream ranking.

Comprehensive experiments on a benchmark dataset for ad text generation demonstrate the superiority of our framework Allegro, which consistently achieves better performance on generation quality and significantly higher diversity metrics compared to previous methods. More importantly, its parallel generation nature further yields a substantial improvement in generation throughput with advanced serving engine implementation (Yang et al., 2025), seamlessly meeting the condition of scalable industrial production. The main contributions of this work are threefold:

- We reformulate ad text generation as a set-level modeling problem, in which the diversity and semantic coverage are explicitly optimized during model training other than heuristic stochastic decoding.

- We propose a parallel generation framework Allegro, which features a semantic planning mechanism and a set-level alignment training process to effectively encourage semantic coverage and penalize redundancy.
- Comprehensive experiments on benchmark show that Allegro achieves state-of-the-art set-level generation quality and diversity and also significantly accelerates inference speed, verifying its scalability for high-throughput industrial deployment.

2 Proposed Framework: Allegro

In this work, we transform the ad text generation task from the standard point-wise modeling $p(y \mid x)$ to a structured, set-level joint distribution modeling $p(\mathcal{Y} \mid x)$, where x denotes the input ad information, and y or $\mathcal{Y} = \{y_i\}_{i=1}^N$ denotes a singular text creative or a set of creatives.

The goal is twofold for generation quality and inference efficiency:

- The overall semantics of the generated set $\mathcal{Y}$ is structurally complete for the given ad to ensure it resonates across varying user interests.
- The generation throughput should be higher compared to traditional sampling-based or autoregressive methods.

This requires that the output set is highly diverse and has comprehensive semantic coverage for the ad, as well as the model architecture can meet the scalability requirements of industrial serving.

2.1 Overview

Figure 2 shows the overview of our proposed framework Allegro. Based on the set-level modeling paradigm, Allegro explicitly decomposes an ad’s content structure into a set of semantic-independent perspectives via a planning mechanism, which enables further parallel generation for the whole set based on the independence of the planned perspectives. To optimize the set-level diversity objective, an alignment training stage built upon synthesized data and diversity reward model from LRMs is proposed to encourage comprehensive semantic coverage for achieving a highly diverse set that facilitates further ranking.

In the following sections, we detail the semantic planning mechanism, the parallel generation process of the creative set and the preference optimization stage for training the Allegro framework.

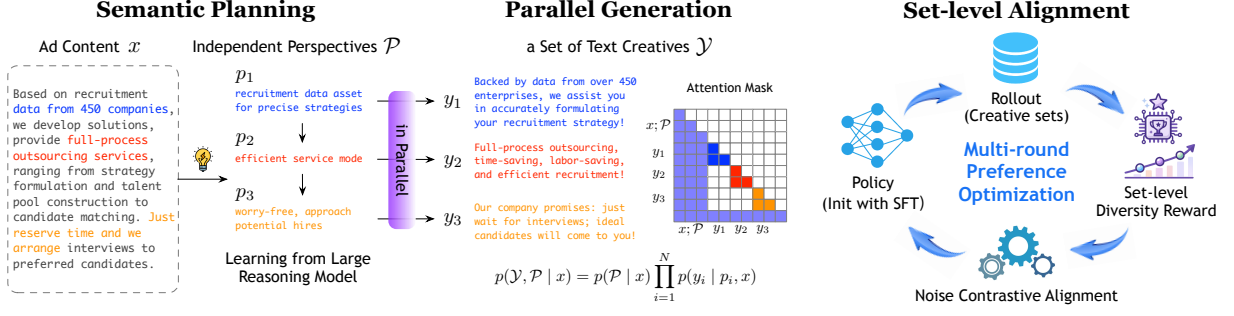


Figure 2: The framework Allegro with semantic planning for parallel ad text generation.

2.2 Semantic Planning for Parallelizability

The semantic planning mechanism in Allegro aims to decompose the given ad’s content structure into a set of semantic-independent perspectives (for example, time-limited promotions, trust-building guarantees and technical services) to describe the ad’s various advantages that may attract target users. Let $\mathcal{P} = \{p_1, p_2, \dots, p_N\}$ be the planned result for the ad content x , where each p_k is a clause representing a distinct, semantically-independent perspective of the ad. Conditioned on these planning clauses, Allegro then generates a corresponding set of text creatives $\mathcal{Y} = \{y_1, y_2, \dots, y_N\}$, where the joint probability of the whole set is factorized as:

$$p(\mathcal{Y}, \mathcal{P} | x) = p(\mathcal{P} | x) \prod_{i=1}^N p(y_i | p_i, x) \quad (1)$$

By conditioning each creative on a distinct planned perspective, the semantic coverage of the generated set is explicitly guided by the complementarity of the planned perspectives in $\mathcal{P}$, rather than being left to stochastic decoding (Wei, 2025).

This decoupling enables all creatives in a set to be generated in parallel, providing a principled path toward high-throughput inference compared with autoregressive diversity methods (Wang et al., 2026), which model the joint probability $p(\mathcal{Y} | x)$ as $\prod_{i=1}^N p(y_i | y_{<i}, x)$.

Structured Planning Dataset Synthesized with Large Reasoning Models

We employ LRMs to synthesize the training data that is composed of both planned semantics of distinct perspectives and corresponding diverse text creatives. LRMs (DeepSeek-AI, 2025; OpenAI, 2025) have demonstrated strong capabilities in long-context understanding and complex reasoning, thus they are well suited for planning semantic structure based on ad content. For constructing the training set of Allegro, we leverage DeepSeek-R1 to first perform

semantic decomposition and then generate text creatives conditioned on the planned semantics. The details are shown in Appendix A.1.

2.3 Parallel Generation for a Set of Creatives

Given the ad content x (represented by the token sequence of its landing page), our Allegro framework first generates the planning clauses $\mathcal{P} = \{(p_1^{(1)}, p_1^{(2)}, \dots, p_1^{(|p_1|)}) , \dots, (p_N^{(1)}, \dots, p_N^{(|p_N|)})\}$ autoregressively. It subsequently splits the forward process into N paths to produce each text creative y_i of a set $\mathcal{Y}$ in parallel, where the N creatives are expected to be complementary. The parallel paths are induced by semantic planning, providing a principled semantic allocation for parallelizability.

The Allegro framework adopts the Multiverse architecture (Yang et al., 2025) which implements parallel generation with specific attention mask and position indices to remove the dependencies among the creatives within the set $\mathcal{Y}$. Formally, let $\mathcal{Y}_{\text{seq}} = \{(y_1^{(1)}, y_1^{(2)}, \dots, y_1^{(|y_1|)}) , \dots, (y_N^{(1)}, \dots, y_N^{(|y_N|)})\}$ denote the flattened token sequence of the creative set $\mathcal{Y}$. The whole sequence $[x; \mathcal{P}; \mathcal{Y}_{\text{seq}}]$ contains the ad content x , the generated planning clauses $\mathcal{P}$, and the generated creative tokens $\mathcal{Y}_{\text{seq}}$. Consider any tokens t_i and t_j within the whole sequence ($j \leq i$), we compute the attention weight A_{ij} as:

$$A_{ij} = \text{softmax} \left((\mathbf{q}_i^\top \odot P(i)) \cdot (\mathbf{k}_j \odot P(j)) + \mathbf{M}_{ij} \right) \quad (2)$$

where $\mathbf{q}_i$, $\mathbf{k}_j$ are the query and key, $P(i)$ is the positional embedding, and $\mathbf{M}_{ij}$ is the attention mask.

To implement parallel generation of all the creatives in a set, the attention mask adds constraints to causal attention, enabling that generating a creative’s token can only attend to the prefixed context of ad content and planned semantics $[x; \mathcal{P}]$ as well as its own previously generated tokens, without attending the tokens from other creatives:

$$\mathbf{M}_{ij} = \begin{cases} 0, & \text{if } t_i \in [x; \mathcal{P}] \text{ and } j \leq i \\ 0, & \text{if } t_i \notin [x; \mathcal{P}] \text{ and } t_j \in [x; \mathcal{P}] \\ 0, & \text{if } t_i, t_j \in y_k \text{ and } j \leq i \\ -\infty, & \text{otherwise} \end{cases} \quad (3)$$

which guarantees causal dependencies within the prefixed context $[x; \mathcal{P}]$, and all parallel generated creatives can attend to $[x; \mathcal{P}]$. Only the tokens belonging to the same creative path have causal dependency, while cross-creative tokens are mutually invisible. Therefore parallel generation is implemented within the same forward pass.

At the meantime, the positional embedding also needs to remove cross-creative dependencies. Instead of assigning globally increasing positions across tokens in creatives, every creative’s path starts from the same positional offset after the prefixed context $[x; \mathcal{P}]$. For any token t_i within the whole sequence, its real index i' for position embedding $P(i')$ is non-monotonically mapped based on its path allocation:

$$i' = \begin{cases} i, & \text{if } t_i \in [x; \mathcal{P}] \\ |[x; \mathcal{P}]| + r_k(i), & \text{if } i > |[x; \mathcal{P}]| \text{ and } t_i \in y_k \end{cases} \quad (4)$$

where $r_k(i) \in \{1, 2, \dots, |y_k|\}$ denotes the relative token index of t_i within the creative path y_k . This mapping ensures that each independent creative path starts from an identical position relative to the prefixed context, which guarantees the correctness of position embedding during forward pass. During inference, all paths are expanded simultaneously while preserving standard autoregressive generation within each path.

2.4 Set-level Alignment

To ensure set-level diversity, we optimize the semantic coverage of the generative creative set for a given ad and train Allegro using a two-stage pipeline: we first perform standard SFT training with the set-level synthesized data from LRMs (see Section 2.2), and then conduct a set-level alignment training stage that explicitly encourages the comprehensive semantic coverage and penalizes redundancy within a creative set through preference optimization with a diversity reward model.

Diversity Reward Model Training To facilitate preference optimization on improving set-level semantic coverage, we train a diversity reward model

$R(x, \mathcal{Y})$ to perform scoring on each given generated creative set. Its training data is composed of a pairwise format, i.e., each ad has a pair of positive set (well semantic coverage) and negative set (semantic redundancy), with reward label 1 and 0 respectively. Specifically, the positive set is the set-level synthesized data for SFT training, and we further leverage DeepSeek-R1 to synthesize a corresponding negative set for each ad. Thus we learn the model $R(x, \mathcal{Y})$ with cross-entropy loss, and the details are shown in Appendix A.2.

Approximately On-Policy NCA Based on the set-level diversity reward, we adopt Noise Contrastive Alignment (NCA) (Chen et al., 2024) to optimize set-level semantic coverage and reduce redundancy within a generated creative set.

Let π_θ denote the policy to be optimized and π_{ref} denote the SFT model. For each ad x , we first generate rollouts of L creative sets $\{\hat{\mathcal{Y}}_l\}_{l=1}^L \sim \pi_{\text{ref}}(\cdot|x)$. The diversity reward model then assigns a score $R_l = R(x, \hat{\mathcal{Y}}_l)$ to each sampled set, reflecting its overall semantic coverage and diversity quality. Intuitively, creative sets with better semantic coverage should receive larger optimization weights, and we convert the scores into soft preference weights $w_l = \frac{\exp(R_l/\alpha)}{\sum_{l=1}^L \exp(R_l/\alpha)}$ where α controls distribution sharpness.

With the approximately on-policy dataset (Chen et al., 2025) composed of $(x, \{\hat{\mathcal{Y}}_l\}_{l=1}^L, \{w_l\}_{l=1}^L)$, we adopt the NCA for optimization to achieve further improvement over the SFT policy π_{ref} :

$$\max \mathbb{E}_{\hat{\mathcal{Y}}_l \sim \pi_{\text{ref}}} \sum_{l=1}^L \left[w_l \log \sigma(r_\theta(x, \hat{\mathcal{Y}}_l)) + \frac{1}{L} \log \sigma(-r_\theta(x, \hat{\mathcal{Y}}_l)) \right]. \quad (5)$$

where $r_\theta(x, \hat{\mathcal{Y}}_l) = \beta \log \frac{\pi_\theta(\hat{\mathcal{Y}}_l|x)}{\pi_{\text{ref}}(\hat{\mathcal{Y}}_l|x)}$ measures how much the current policy prefers a sampled creative set relative to the SFT policy. The first term encourages the policy to assign higher probability to creative sets with larger diversity scores, and the second regularization term prevents excessive deviation from the SFT policy. The preference optimization stage explicitly favors creative sets that achieve broader semantic coverage through the controlling of w_l . It follows the approximately on-policy paradigm, where creative sets for training are sampled from the SFT policy rather than reused off-policy data from the synthesized corpora.

Method	Quality	Diversity			Relevance	
	Multi-ref BLEU	PairwiseBLEU	Distinct-3	Semantic Coverage	Ad-to-LP	Keyword-to-Ad
Ideation	43.1	36.3	74.1	27:13:60	66.4	72.8
Diver (32B)	45.1	36.8	66.7	(baseline)	75.1	76.4
PlannedDiffusion (7B)	39.7	42.7	59.9	38:21:41	76.9	74.5
Allegro (7B)	44.9	41.4	64.7	-	79.4	74.2
Allegro (32B)	46.8	34.5	65.9	47:35:18	80.3	75.4

Table 1: Empirical evaluation of various methods for ad text generation. Semantic coverage is measured by human evaluation in a pairwise fashion, reporting ‘#Good:#Same:#Bad’ relative to the autoregressive baseline Diver.

2.5 High-Throughput Serving

The semantic planning mechanism of Allegro decomposes ad content into independent perspectives for generating creatives in parallel, which is naturally compatible with the Multiverse serving engine (Yang et al., 2025; Zheng et al., 2024). We reuse the Multiverse implementation for high-throughput serving, where the path split for each creative is initiated and terminated by special token `<Path>` or `</Path>`.

3 Experiment

3.1 Experimental Setup

Dataset and Evaluation Metrics We conduct empirical evaluation on the Camera dataset (Mita et al., 2024), an industrial benchmark for ad text generation. Because this dataset is collected from search advertising scenario, for each sample we use both the ad landing page (LP) and the keyword as the input x , and we generate a set of $N = 4$ creatives during inference on the test split.

We evaluate generation quality, diversity, relevance and inference speedup in experiments. For quality, we use multi-reference BLEU4 that compares each creative in a set with the ground-truth set. For diversity, we compute PairwiseBLEU and Distinct-3 of a generated set for *lexical* diversity, and perform human evaluation to measure *semantic* coverage to LP in a pairwise fashion that reports ‘#Good:#Same:#Bad’ relative to a given baseline. For relevance, we employ LLM-as-a-Judge to evaluate ad-to-LP message match and keyword-to-ad relevance, reflecting the requirements of consistency and attractiveness, respectively. More details are in Appendix B.

Baselines We compare Allegro with methods that explicitly model diversity or plan semantics:

- Ideation (Wei, 2025) is a sampling-based

method based on mT5-13B (Xue et al., 2021), which learns to construct a latent space via latent diffusion to perform diverse sampling.

- Diver (Wang et al., 2026) is an autoregressive method for modeling creative diversity, which employs lexical-based 1-PairwiseBLEU as diversity reward and generates each creative of a set one by one.
- PlannedDiffusion (Israel et al., 2026) is a diffusion LM based on Dream-7B (Ye et al., 2025), which first creates a plan list and then generates each span simultaneously.

Implementation Details For the methods based on decoder-only architecture (Diver and Allegro), we use Qwen2.5-32B-Instruct (Qwen, 2024) as backbone. We also train a 7B version of Allegro based on Qwen2.5-7B-Instruct to compare with PlannedDiffusion built upon Dream-7B that is continually trained from Qwen2.5-7B-Instruct. For fair comparison, all methods use the same training data (creative sets synthesized from LRMs). More details are in Appendix A.3.

3.2 Main Results

Table 1 shows the empirical evaluation results. Compared to sampling-based baseline Ideation, Allegro consistently achieves higher generation quality and relevance, as well as maintaining competitive diversity metric. This indicates that the unpredictability in the stochastic decoding process may result in poor consistency with the ad content and the keyword, however relevance is crucial for text creative generation in industrial scenario.

Our Allegro framework further improves diversity metrics over the autoregressive-based diversity method Diver, especially in semantic coverage compared to lexical-level metrics. We suggest that this improvement is attributed to the explicit

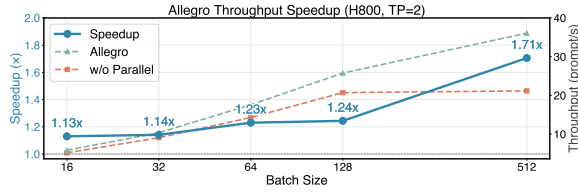


Figure 3: The inference speedup of Allegro compared to its autoregressive variant.

Method	Multi-ref BLEU	PairwiseBLEU	Ad-to-LP
Allegro	46.8	34.5	80.3
– Planning	42.2	38.1	76.2
– Parallel	47.1	35.8	79.6
– On-Policy	45.0	38.1	77.8

Table 2: Effects of key components in Allegro.

semantic decomposition to independent perspectives for a given ad, and the creatives induced by the planned semantics can effectively avoid redundancy within the generated set. Besides, Allegro (7B) also achieves better overall balance among various metrics compared to planning-based PlannedDiffusion, indicating the effectiveness of the set-level reformulation.

It is an expected trade-off between the better semantic coverage and the slight decrease in lexical-level Distinct-3 score. Empirically, the baselines achieve slightly higher lexical-level diversity but at the cost of hurting ad-to-LP relevance, demonstrating that they obtain n -gram variance but also generate phrases that are inconsistent with landing page. In contrast, our Allegro framework explicitly anchors the generation of each creative to a planned semantic perspective, and the generated creatives are highly relevant.

3.3 Inference Speedup

We evaluate inference efficiency for industrial serving. In online advertising, the ad storage contains hundreds of millions ads, and the key metric for generation is throughput (queries per second), measuring how many ads can be successfully processed in a second and determining the magnitude of creatives. As in Figure 3, on 2 NVIDIA H800 GPUs with tensor parallelism (i.e., TP=2), Allegro achieves 1.71 times speedup over its autoregressive variant with large batch size and $N = 4$, indicating that the semantic planning not only improves diversity but also naturally enables scalable serving through principled parallel generation.

Ad Content: Limited-time employment bonus totaling 400,000 yen! [COMPANY] temporary workers can enjoy a free single apartment, and with a two-day weekend system, inexperienced workers can earn an annual income of over 4 million yen. From 2016 to 2020, 1,425 employees have successfully transitioned to become regular members...

Keyword: Career Opportunity

- [COMPANY] is recruiting temporary workers! Enjoy a 400,000 yen bonus and a free apartment in a limited time!
(Planning: Bonus and accommodation benefits)
- 1425 employees have successfully become regular employees! Temporary worker recruitment is in full swing.
(Planning: Endorsement for regularization)
- Limited time to join! Temporary workers at [COMPANY] enjoy high salaries, accommodation and a two-day weekend.
(Planning: Bonus and comprehensive benefits)
- [COMPANY] is recruiting temporary workers! Enjoy a free single apartment and a two-day weekend system.
(Planning: Accommodation and working hours benefits)

Table 3: Representative creative set from Allegro.

3.4 Ablation

We verify the effects of key components in Allegro with the variants: (1) w/o Planning: we replace the planned semantics in synthesized data to unmeaningful placeholders (e.g., ‘perspective A/B/C of LP’). (2) w/o Parallel: we replace the Multiverse architecture to regular autoregressive generation. (3) w/o On-Policy: we replace the SFT-generated training data of alignment to LRM-generated.

As the results shown in Table 2, Allegro significantly outperforms ‘w/o Planning’ on diversity, indicating that explicitly allocating semantics from ad content is the core for semantic coverage. Removing the approximately on-policy data also results in a large drop on various metrics, which supports the effectiveness of self-generated rollouts during alignment. Replacing parallel generation with sequential generation slightly improves generation quality while performing comparably in diversity, suggesting that parallel generation preserves most quality metrics while offering advantages for industrial serving.

3.5 Qualitative Analysis

Table 3 shows an example from Allegro, which generates creatives corresponding to planned semantics. The diversity primarily arises from complementary semantic coverage rather than superficial lexical variation. This example’s creatives jointly capture multiple key perspectives of the ad, and thus the planned semantics contributes to a structured set that is more diverse to facilitate following personalized ranking.

4 Related Work

Ad text generation is a key stage in online advertising for improving both user engagement and platform revenue (Hughes et al., 2019; Mishra et al., 2020; Kanungo et al., 2021; Wang et al., 2021), and previous work mainly focuses on predefining control tokens (Golobokov et al., 2022; Wei et al., 2022), autoregressive modeling (Wang et al., 2026) and sampling (Wei, 2025; Wei et al., 2026) for achieving diverse outputs. However, existing methods rarely consider the creative set as a whole and do not explicitly address the rigorous throughput requirements of industrial serving.

5 Conclusion

We reformulate ad text generation as a set-level modeling problem and propose a parallel generation framework Allegro, in which a semantic planning mechanism and a set-level alignment stage collaboratively encourage semantic coverage for generation. Experiments show its advantages on quality, diversity and efficiency, demonstrating its scalability for high-throughput deployment.

Ethical Considerations

The main ethical considerations for generating ad text creatives are the privacy issue and potential harmful content. To protect privacy and ensure content harmlessness, we carefully prepare training corpora with multiple data cleaning and filtering stages, such as identifying potential privacy information and detecting potential harmful phrases, which makes sure that only qualified content is used for training models to make every effort to prevent unqualified generated results.

Limitations

The main limitation of this work is that the semantic planning mechanism relies on the reasoning ability of LRMs. However the planning process lacks direct supervision during training, and it also does not utilize the accumulated experience from human advertisers. Also, the final generation performance depends heavily on the semantic planning quality, and it cannot handle the ads with weakly structured or highly repetitive content. Besides, our Allegro framework currently only focuses on short text creatives, while in online advertising the types of creatives include short text, long-form text, image, and so on, which also need diversity modeling to facilitate personalized ranking.

References

- Howard Chen, Noam Razin, Karthik Narasimhan, and Danqi Chen. 2025. Retaining by doing: The role of on-policy data in mitigating forgetting. *arXiv preprint arXiv:2510.18874*.
- Huayu Chen, Guande He, Lifan Yuan, Ganqu Cui, Hang Su, and Jun Zhu. 2024. Noise contrastive alignment of language models with explicit rewards. *Advances in neural information processing systems*.
- DeepSeek-AI. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.
- Konstantin Golobokov, Junyi Chai, Victor Ye Dong, Mandy Gu, Bingyu Chi, Jie Cao, Yulan Yan, and Yi Liu. 2022. Deepgen: Diverse search ad generation and real-time customization. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 191–199.
- J Weston Hughes, Keng-hao Chang, and Ruofei Zhang. 2019. Generating better search engine text advertisements with deep reinforcement learning. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2269–2277.
- Daniel Israel, Tian Jin, Ellie Cheng, Guy Van den Broeck, Aditya Grover, Suvinay Subramanian, and Michael Carbin. 2026. Planned diffusion. In *International Conference on Learning Representations*.
- Yashal Shakti Kanungo, Sumit Negi, and Aruna Rajan. 2021. Ad headline generation using self-critical masked language model. In *Proceedings of the 2021 conference of the north American chapter of the association for computational linguistics: human language technologies: industry papers*, pages 263–271.
- Shaunak Mishra, Manisha Verma, Yichao Zhou, Kapil Thadani, and Wei Wang. 2020. Learning to create better ads: Generation and ranking approaches for ad creative refinement. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 2653–2660.
- Masato Mita, Soichiro Murakami, Akihiko Kato, and Peinan Zhang. 2024. Striking gold in advertising: Standardization and exploration of ad text generation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 955–972.
- OpenAI. 2025. Competitive programming with large reasoning models. *arXiv preprint arXiv:2502.06807*.
- Qwen. 2024. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*.
- Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti

Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, and 49 others. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Chang Wang, Siyu Yan, Depeng Yuan, Yuqi Chen, Yanhua Huang, Yuanhang Zheng, Shuhao Li, Yinqi Zhang, Kedi Chen, Mingrui Zhu, and Ruiwen Xu. 2026. Diver: Unlocking diversity in ad headline generation with large language models. In *Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 4981–4986.

Xiting Wang, Xinwei Gu, Jie Cao, Zihua Zhao, Yulan Yan, Bhuvan Middha, and Xing Xie. 2021. Reinforcing pretrained models for generating attractive text advertisements. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pages 3697–3707.

Penghui Wei. 2025. Ideation: Diversifying ad text generation in sponsored search with latent diffusion. In *Companion Proceedings of the ACM on Web Conference 2025*, pages 1394–1397.

Penghui Wei, Jiayu Wu, Chao Ye, Zhi Guo, Shuanglong Li, and Lin Liu. 2026. Interactor: Agentic rl oriented iterative creation for ad description generation in sponsored search. *arXiv preprint arXiv:2606.15911*.

Penghui Wei, Xuanhua Yang, Shaoguo Liu, Liang Wang, and Bo Zheng. 2022. Creator: Ctr-driven advertising text generation with controlled pre-training and contrastive fine-tuning. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Track*, pages 9–17.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mt5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pages 483–498.

Xinyu Yang, Yuwei An, Hongyi Liu, Tianqi Chen, and Beidi Chen. 2025. Multiverse: Your language models secretly decide how to parallelize and merge generation. *Advances in Neural Information Processing Systems*.

Jiacheng Ye, Zhihui Xie, Lin Zheng, Jiahui Gao, Zirui Wu, Xin Jiang, Zhenguo Li, and Lingpeng Kong. 2025. Dream 7b: Diffusion large language models. *arXiv preprint arXiv:2508.15487*.

Lianmin Zheng, Liangsheng Yin, Zhiqiang Xie, Chuyue Sun, Jeff Huang, Cody H Yu, Shiyi Cao, Christos Kozyrakis, Ion Stoica, Joseph E Gonzalez, Clark Barrett, and Ying Sheng. 2024. Sglang: Efficient execution of structured language model programs. *Advances in neural information processing systems*.

A Implementation Details

A.1 LRMs for Synthesizing Training Data

To synthesize the training dataset containing both decomposed semantics and text creatives, we instruct DeepSeek-R1 (DeepSeek-AI, 2025) with:

You are an advertising copywriting expert proficient in writing various types of commercial advertising headlines. You excel in combining targeted keywords with the core content of the landing page, creating four high-quality advertising headlines with diverse styles that attract attention. Based on the input provided below, first think about diverse and independent marketing perspectives, and then create advertising headlines from different perspectives, strictly adhering to the following guidelines:

1. *Requirements for a single headline: i. The content must closely match the information on the landing page; ii. When the landing page can match keywords, the headline must incorporate the keywords; iii. The number of characters in a single headline should be controlled between 12-24.*
2. *Requirements for multiple headlines: The four headlines should be independent of each other and have distinct perspectives.*

Input information: Keyword: {keyword}; Landing page content: {landing_page}

Output: { 'headline 1': 'Headline 1', 'headline 1 plan': 'The perspective for creating the headline', 'headline 2': 'Headline 2', 'headline 2 plan': 'The perspective for creating the headline', 'headline 3': 'Headline 3', 'headline 3 plan': 'The perspective for creating the headline', 'headline 4': 'Headline 4', 'headline 4 plan': 'The perspective for creating the headline' }

A.2 Set-level Diversity Reward Modeling

To evaluate set-level diversity, we train a reward model $R(x, \mathcal{Y})$ that performs scoring on each given generated creative set. It employs Qwen2.5-7B-Instruct (Qwen, 2024) backbone and uses the probability of token ‘yes’ to determine whether a given creative set covers overall semantics to a given ad content. The training data contains 21k pairs, where the positive creative set is reused from the synthesized data for SFT training, and the negative set is also synthesized with DeepSeek-R1.

To prevent the reward from taking shortcuts, we strictly constrain the generation of negative sets. Specifically, we instruct DeepSeek-R1 to explicitly collapse the planning stage, in which the prompt forced the model to select only one or two domi-

nant selling points from the ad content and generate all four creatives based solely on these restricted perspectives, intentionally creating semantic redundancy and only lexical-level diversity, which also leads to potential ad-to-LP mismatch. We apply strict length and formatting constraints to both positive and negative sets, ensuring the average length and tone are identical. On a held-out set of 6k examples, the trained model achieves an accuracy of 0.904 in correctly distinguishing the diverse set from the redundant set. To further ensure the reward aligns with human judgment, we sample 1k pairs and then three human experts independently perform blind labeling on them. The model’s predictions align with the human’s majority vote in 87.5% of the sampled pairs, demonstrating strong consistency with actual semantic coverage.

A.3 Model Configuration

During approximately on-policy alignment stage, we sample $L = 5$ sets for each input with preorder checkpoint, and we perform two-round alignment training (Touvron et al., 2023) to achieve further improvement compared to only one-round training. The hyperparameters $\alpha = 0.05$ and $\beta = 0.01$. All models are trained on 8 NVIDIA H800 GPUs. On average, for each sample the number of generated planning tokens is ~ 95 , and that of each generated creative is ~ 16 .

B Evaluation

For fair comparison, all methods employ the same synthesized data from DeepSeek-R1 for training and evaluation¹. The Camera dataset is in Japanese and in the illustration of this paper we translate both input and model output to English for readability.

To evaluate *semantic coverage*, we perform human evaluation on the test set covering 872 samples. For each input, three copywriting expert are given a pair of two generated sets from two methods, and then they annotate Good/Same/Bad for each pair by comparing whether the first set outperforms the second set. The final label is determined by majority voting. The inter-annotator agreement is 0.76 by Krippendorff’s α . We report ‘#Good:#Same:#Bad’ normalized to summation to 100.

¹For the reference-based metric multi-reference BLEU, the reference set is also provided by DeepSeek-R1: the original references of the Camera test set suffer from quality degradation, where many references are excessively short, lack diversity, or exhibit significant ad-to-LP mismatch.

For relevance, we employ LLM-as-a-Judge to evaluate ad-to-LP message match and keyword-to-ad relevance. The ad-to-LP message match metric is vital, that reflects whether a text creative satisfies the basic requirement of consistency, instead of fabricating wrong statements. We also evaluate keyword-to-ad relevance to reflect attractiveness. Both the metrics are binary, where for each input ad and each text creative in a set, we leverage DeepSeek-R1 to produce a result of 1 or 0.

Below is the prompt for evaluating ad-to-LP message match:

You are a meticulous expert in advertising content review. Your task is to assess the consistency between the ad headline and the ad landing page content, and provide the result for the headline. A score of 1 indicates consistency, while a score of 0 indicates inconsistency.

Your task is to evaluate whether the key information described in the ad headline, such as selling points, functions, promotions, services, geographical locations, brands, numbers (age/discount/business hours/distance), etc., can be directly or reasonably inferred from the information on the landing page, avoiding issues such as ‘illusion’ (fabrication of facts), ‘exaggeration’ (magnification of degree), and ‘deviation’ (inconsistency with the main content of the landing page). The consistent standard is: all facts/numbers/selling points in the headline can be directly found on the landing page; the expression is restrained and not exaggerated; there are only minor reasonable extensions (such as synonym rewriting, generalization), with no key factual errors.

When making judgments, please focus on checking: 1. Numbers/age/discount/distance/business hours: Whether they are consistent with the numbers on the landing page. 2. Geographical location/brand/institution: Whether they are consistent with the landing page, without any unfounded replacement or expansion. 3. Services/functions/offers: Whether they are clearly mentioned on the landing page, avoiding making things up out of thin air. 4. Degree adverbs used in words (such as ‘the most’, ‘the first’, ‘exclusive’, ‘the lowest across the network’, ‘a century’, ‘absolutely’): They must be clearly supported on the landing page.

Input information: Ad landing page: {landing_page}; ad headline: {ad_headline}

Please output "1" or "0" directly, without any additional content.

Below is the prompt for evaluating keyword-to-ad relevance:

You are a seasoned advertising copywriter and click-through rate optimization expert. Your task is to evaluate the attractiveness of each ad headline to the target keyword, and provide the result for the headline. A score of 1 indicates attractiveness, while a score of 0 indicates no attractiveness.

Attractiveness refers to whether a headline can generate a desire to click among target users in search/recommendation scenarios. The focus is on the presentation of selling points, user benefits, emotions/urgency, sentence novelty, and readability. The selling points should be clear, the expression should be smooth and natural, and they should be able to arouse user interest; a few places may be slightly bland or common. Typical situations lacking attractiveness are: almost no association with keywords, being empty, obscure, or having obvious grammatical errors.

When evaluating, please consider comprehensively: 1. Whether the benefit points derived from the keywords are prominent (specific rather than vague). 2. Whether there is differentiation and scarcity. 3. Whether the sentence structure is novel and avoids being template-like. 4. Whether the wording is concise. 5. Whether it targets the target audience corresponding to the keyword. 5. Do not evaluate whether the facts are true.

Input information: Keyword: {keyword}; Ad landing page: {landing_page}; ad headline: {ad_headline}

Please output "1" or "0" directly, without any additional content.