CLOSING THE DATA-EFFICIENCY GAP BETWEEN AUTOREGRESSIVE AND MASKED DIFFUSION LLMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Despite autoregressive large language models (arLLMs) having been the dominant paradigm in language modeling, they resist knowledge injection via fine-tuning due to inherent shortcomings such as the "reversal curse" - the challenge of answering questions that reverse the original information order in the training sample. Masked diffusion large language models (dLLMs) are rapidly emerging as a powerful alternative to the arLLM paradigm, with evidence of better data efficiency and free of the "reversal curse" in pre-training. However, it is unknown whether these advantages still extend to the post-training phase, i.e. whether pre-trained dLLMs can easily acquire new knowledge through fine-tuning. To assess post-training knowledge acquisition and generalization, we perform fine-tuning using 3 different datasets on arLLMs and dLLMs and evaluate them with two types of QA formats: forward style QA (questions follow the original information order of the training sample) and backward style QA (questions reverse the original information order of the training sample), which probes the reversal curse. We first show that arLLMs heavily rely on paraphrases to generalize knowledge text into question-answering (QA) performance; paraphrases are only effective when the information order in paraphrased text matches the QA style. In contrast, dLLMs achieve strong performance on both forward and backward style QAs without paraphrases, with paraphrases yielding only marginal additional gains. Lastly, inspired by the performance of dLLM fine-tuning, we propose a new masked fine-tuning paradigm for knowledge injection in pre-trained arLLMs. The proposed paradigm drastically improves the data efficiency in arLLMs fine-tuning, closing the gap with dLLMs.

1 INTRODUCTION

Despite auto-regressive large language models (arLLMs) having been the main contributor to the modern success of language modeling, studies have demonstrated the difficulty of injecting new knowledge to pre-trained arLLMs by fine-tuning on documents that are not in the pre-training dataset (Ovadia et al., 2023; Mecklenburg et al., 2024; Gekhman et al., 2024; Soudani et al., 2024; Zhao et al., 2025; Lampinen et al., 2025). Fine-tuned models typically generalize poorly to downstream tasks such as question-answering (QA). An example failure mode is the famous "*reversal curse*", that LLMs fail to answer the questions in the reversed order of the training text (Berglund et al., 2023). Fine-tuning on multiple rewrites (i.e. paraphrases) of the documents can mitigate the generalization issues, but still falls behind in-context learning based external memory systems like RAG (Ovadia et al., 2023; Mecklenburg et al., 2024). This pitfall of arLLMs is a major obstacle that limits current models to be flexible life-long learners via weight updates.

As alternatives to the auto-regressive models, several recent masked diffusion large language models (dLLMs) have been scaled up to be as capable as arLLMs on multiple downstream tasks, with extra advantages such as high-throughput decoding of multiple tokens simultaneously ((Nie et al., 2025a;b; Ye et al., 2025). Instead of factorizing sequence probability token by token in sequential order, dLLMs learn the factorization of the sequence probability in arbitrary orders, i.e. they can use any subset of tokens in a sequence to compute the joint probability of the rest of the tokens. Though such an objective is harder than learning autoregressive factorization (Kim et al., 2025), due to the any-order factorization, dLLMs inherently do not suffer from "*reversal curse*"(Nie et al., 2025b),

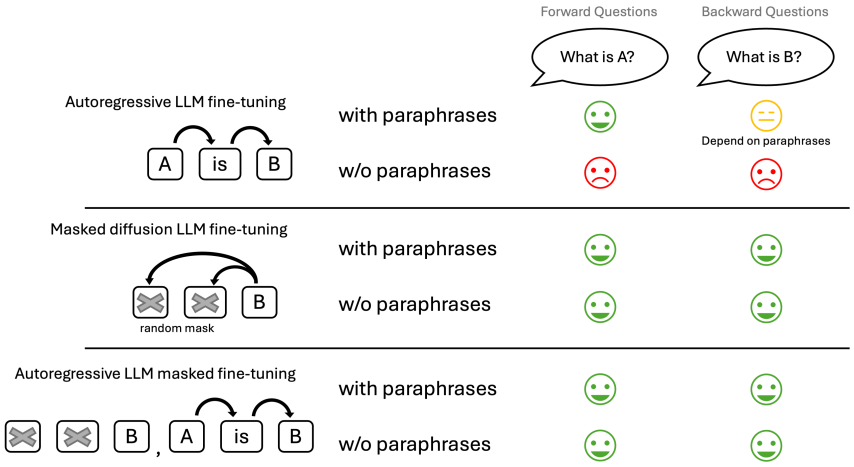

Figure 1: A schematic summary of the results. First row: autoregressive LLM requires paraphrases for generalizing knowledge in the fine-tuning text to QA tasks, and suffer from reversal curse (i.e. fail to answer backward questions). Second row: masked diffusion LLM can easily generalize fine-tuning text to QA tasks in both forward and backward styles. Third row: inspired by the masked diffusion LLM, we propose a masked fine-tuning paradigm, that closes the fine-tuning gap between autoregressive LLMs and masked diffusion LLMs.

and can achieve lower validation loss than arLLMs in a data-constrained regime (Prabhudesai et al., 2025; Ni & the team, 2025). However, most of the dLLM studies have been focusing on the properties in the pre-training phase, little is known if dLLMs also have advantages in the post-training phase, such as knowledge injection by fine-tuning.

In this study, we use three datasets to compare the data efficiency and performance of knowledge injection by fine-tuning in arLLM and dLLM models. We also introduce a novel masked fine-tuning paradigm for arLLMs that emulates diffusion-style mask reconstruction loss. Across all datasets, dLLMs show a consistent data-efficiency advantage over arLLMs, and our masked fine-tuning largely closes this gap, bringing arLLMs to strong performance without relying on paraphrase. More specifically, we show the following results:

- arLLMs heavily rely on paraphrases to successfully generalize fine-tuning text to downstream QA tasks; arLLMs fail on backward style questions and only paraphrases that reverse information order in the sentences can mitigate reversal curse.
- dLLMs can achieve high accuracy in both forward and backward questions without paraphrases; adding paraphrases only marginally helps. This establishes the knowledge injection data efficiency of dLLMs in the post-training phase.
- We propose a masked fine-tuning paradigm that fine-tunes arLLMs in a "masked diffusion" way by giving masked samples in the context with instructions to recover the mask, and set the unmasked sample as the supervised fine-tuning target. The novel method closes the performance gap between arLLMs and dLLMs fine-tuning: it achieves strong performance in both forward and backward questions without paraphrases.

Taken together, these findings indicate that dLLMs, with the masked training objective, offers advantages in the post-training phase, reflecting superior data efficiency relative to arLLMs. We further show that this advantage can be transferred to arLLMs via our masked fine-tuning paradigm. Our findings suggest the possibility to post-train an LLM to adapt to the changing world using a small amount of new knowledge texts, which could help address challenges in keeping AI systems updated with changing environment.

2 BACKGROUND

2.1 KNOWLEDGE INJECTION BY FINE-TUNING AND REVERSAL CURSE

A desired AI system should be able to continuously learn new knowledge to adapt to the changing environment. Though LLMs have been successful on numerous tasks, they struggle to incorporate new knowledge into their weights. At least two factors contribute to this difficulty. The first is that LLMs show catastrophic forgetting after fine-tuning on new tasks (Luo et al., 2023; Wang et al., 2023; Zhai et al., 2023; Zhang & Wu, 2024; Chen et al., 2024; Ren et al., 2024). Another issue is fine-tuning a pre-trained LLM has been shown to be less effective in injecting new factual knowledge than learning a response style (Ovadia et al., 2023; Mecklenburg et al., 2024; Gekhman et al., 2024; Soudani et al., 2024; Zhao et al., 2025; Lampinen et al., 2025).

A famous failure mode of learning knowledge in the text is the "*reversal curse*", that after learning statements of the form "A is B", the model does not generalize it to its inverse form "B is A". The reversal curse has been observed across the training phases and models (Berglund et al., 2023; Allen-Zhu & Li, 2025; Lv et al., 2024; Lin et al., 2024; Guo et al., 2024; Golovneva et al., 2024; Lu et al., 2024). Even strong commercial models like GPT-4 and GPT-4o show signs of the reversal curse (Berglund et al., 2023; Nie et al., 2025b). The cause of the reversal curse has been theoretically attributed to an inherent limitation of the autoregressive training objective (Zhu et al., 2024; Kitouni et al., 2024) (See Appendix A.6). The common approaches for mitigating the reversal curse in autoregressive models include 1) adding paraphrases that contain information in different semantic orders (Guo et al., 2024; Lu et al., 2024; Golovneva et al., 2024); 2) changing causal attention to bi-directional attention (Lv et al., 2024) or further changing the autoregressive objective to mask-recovery objective (Kitouni et al., 2024; Nie et al., 2025b). Unlike the above methods, our proposed masked fine-tuning paradigm in arLLMs solves the reversal curse without constructing paraphrase augmentations or changing the autoregressive objective.

2.2 MASKED DIFFUSION LANGUAGE MODELS

Recently, dLLMs have emerged as a strong competitor to arLLMs (Sahoo et al., 2024; Nie et al., 2025b; Ye et al., 2025). Comparing to autoregressive models, dLLMs use encoder-only transformers to generate text by iteratively unmasking tokens via a reversed discrete diffusion process. The training objective is to minimize the mask reconstruction loss Nie et al. (2025b):

$$\mathcal{L}(\theta) = -\mathbb{E}_{t,x_0,x_t}\left[\frac{1}{t}\sum_{\ell=1}^{L}\mathbb{I}[x_t^\ell \in \mathrm{M}]\log p_\theta(x_0^\ell|x_t)\right], \tag{1}$$

where x_0 is sampled from the training data, t is the sampled mask ratio; M denotes the masked tokens sampled by the forward process samples with ratio t; x_t is the masked version of x_0. Such a loss objective has been shown to be the negative evidence lower bound (ELBO) on the data likelihood (Shi et al., 2024).

Several advantages of dLLMs regarding data efficiency have been claimed. When the training data is scarce, dLLMs keep improving with repeated use of the data and surpass arLLMs on validation loss, while arLLMs saturate the validation loss or increase it due to overfitting (Prabhudesai et al., 2025; Ni & the team, 2025). Prabhudesai et al. (2025) further shows that the lower validation loss in dLLMs can generalize to downstream tasks like ARC-Easy, and attributes its data efficiency to random masks as implicit data augmentation. This evidence indicates that dLLMs would also be competitive in the knowledge injection by fine-tuning settings, where knowledge to be learned is embedded in individual documents with no repetitions.

2.3 CHANGE ORDER TRAINING OF ARLLM

In the following sections, we propose a masked fine-tuning paradigm for arLLM. There have been studies that explored training arLLMs not in the language sequence order, but either in reverse order or in permuted orders. Golovneva et al. (2024) proposed reverse training which trains an arLLM with both regular token sequence and reversed token sequence to mitigate the reversal curse.

Bavarian et al. (2022) proposed training an arLLM with a fill-in-the-middle objective that enables the resulting model to excel in text infilling tasks. Yang et al. (2019); Hoogeboom et al. (2021); Shih et al. (2022) trains a non-causal decoder-only transformer to autoregressively decode a sequence in any order similarly to dLLMs. Our proposed masked fine-tuning paradigm for arLLM does not modify the causal attention pattern or anything in a pre-trained arLLM, but only reformulates the de-mask objective into an instruction fine-tuning objective with a carefully crafted user prompt.

3 DATASETS AND EXPERIMENTAL SETUPS

We focus on assessing LLMs' ability to learn new knowledge through fine-tuning. More specifically, LLMs are fine-tuned on a set of documents that contain knowledge unknown to the base LLM, and evaluated by open-ended QA tasks. Correctness of an answer is evaluated by the well-adopted ROUGE-1 score (Lin et al., 2024; Jiang et al., 2025) between the generated answer and ground truth answer, which we report as "accuracy." It measures the proportion of the words in ground truth answer that appears in the generated answer. To better demonstrate the generation quality, we also show examples of model responses in all the experiments in A.5.

We use three representative datasets. Two are existing synthetic datasets from previous studies on the reversal curse; and we also constructed a realistic dataset from real Wikipedia articles that are recent in time. Each dataset has been augmented with paraphrases. See examples of each dataset in Appendix A.3.

The ***NameDescription*** dataset is from Berglund et al. (2023). It contains 60 statements of different fictitious individuals, 30 each of the form "[name] is [description]" (N2D) and "[description] is [name]" (D2N). Lin et al. (2024) extended the dataset with an open-ended QA testing set. For each type of the statements, the QA set contains two types of questions: "What is the name related to a given description" and "What is the description of a given name". Depending on whether the question is aligned with the original statement, each question is classified as "forward" or "backward" question (e.g. N2D statement with "What is the description of a given name" type of question is a forward question). The dataset also contains a paraphrase set, that each statement is rewritten into 30 different versions, but the order of [name] and [description] in paraphrases is always preserved as in the original statement (either N2D or D2N).

The ***Biography*** dataset is proposed in Allen-Zhu & Li (2024; 2025). Since the original dataset is not publicly available, we used a subset of 100 samples from a replication (Zheng et al., 2025). Each sample is a 6-sentence paragraph about a fictitious individual on their birth city, birthday, college, and job information. Note that the name only appears in the first sentence and is replaced with a pronoun in the following sentences, thus questions about the name are considered as backward questions. Each sample also has a paraphrase set of 5 paraphrases; the paraphrases do not change the order of the sentences but only change the wording while preserving the information. The testing QA set has both forward (i.e. asking for an attribute given the name) and backward style (i.e. asking for the name given 3 attributes from the person) questions.

We construct a ***Wiki*** dataset that contains 92 Wikipedia articles following the procedures in Pan et al. (2025). We crawl the Wikipedia pages under the category "2025 by month," then further filter out the pages that were created before year 2025. This procedure ensures these real-world events are recent enough that both pre-trained models should not be aware of, which is justified by the model accuracy before fine-tuning (Table 3). For each wiki article, we use GPT-o3-mini to generate QA pairs in both forward and backward styles. By prompting GPT-o3-mini, we construct two different paraphrase sets: one keeps information in place and only changes the wording (same-order paraphrases); the other also changes the order of information in the article (permute-order paraphrases). 10 of each type of paraphrases are generated for each wiki article. More details on constructing the datasets are provided in Appendix A.3.

We choose Llama-3.1-8B-Instruct (Dubey et al., 2024) and LLaDA-8B-Instruct (Nie et al., 2025b) models as representatives of arLLM and dLLM to conduct the experiments, as they perform similarly on the benchmarks and are of comparable sizes. Fine-tuning and evaluation configurations are provided in the Appendix A.4.

	NameDescription				Biography	
	N2D-fwd	N2D-bwd	D2N-fwd	D2N-bwd	Fwd	Bwd
arLLM before fine-tuning	0.072	0.000	0.054	0.000	0.001	0.000
arLLM w/o paraphrases	0.374	0.000	0.017	0.027	0.121	0.002
arLLM w paraphrases	0.910	0.004	0.925	0.071	0.962	0.001

Table 1: Fine-tuning performance of arLLM on the NameDescription and Biography datasets.

	Wiki	
	Fwd	Bwd
arLLM before fine-tuning	0.164	0.127
arLLM w/o paraphrases	0.377	0.282
arLLM w same-order paraphrases	0.685	0.396
arLLM w permute-order paraphrases	0.721	0.628

Table 2: Fine-tuning performance of arLLM on the Wiki dataset.

4 arLLM KNOWLEDGE INJECTION RELIES ON PARAPHRASES

We first show that knowledge injection by fine-tuning in arLLMs heavily relies on paraphrases. This is known in previous studies (Berglund et al., 2023; Allen-Zhu & Li, 2025; Lin et al., 2024; Guo et al., 2024; Golovneva et al., 2024). We consistently demonstrate this observation on three datasets to set baselines for the comparison with dLLM and our novel paradigm in the following sections.

We fine-tune Llama-3.1-8B-Instruct on dataset samples with the pre-training format. Without paraphrases, backward accuracy on the NameDescription and Biography datasets is close to 0, while forward accuracy of NameDescription N2D and Biography does not completely fail but is still poor (Table 1). Adding paraphrases drastically raises forward accuracy close to 1, while the backward accuracy is still close to 0. Paraphrases do not help backward accuracy in NameDescription and Biography datasets due to the construction of them in these datasets not changing the semantic order of the sentences. The trend is similar in the Wiki dataset (Table 2). While the same-order paraphrases significantly increase the forward accuracy, they only mildly increase backward accuracy. Using permute-order paraphrases increases both forward and backward accuracy, and the gap between them is smaller. Note that due to the naturalness of this dataset, we could not completely remove the effect of base knowledge, which we also report in Table 2.

These results suggest that, in arLLM fine-tuning, paraphrases significantly improve QA accuracy, but help backward questions only when the paraphrases change the information order in the sentences to be more aligned with the backward style. Note that the accuracy difference between fine-tuning with paraphrases and without paraphrases is not due to different training steps; in both cases, we train the models with sufficiently large epoch numbers; the reported accuracy is from the best checkpoints during the training (Figure 2, Appendix 6).

5 dLLM KNOWLEDGE INJECTION

We then test if dLLMs are more data-efficient regarding knowledge injection by fine-tuning, specifically if dLLM requires paraphrases to achieve both forward and backward QA. We follow the original pretraining protocol (Nie et al., 2025b) to fine-tune LLaDA-8B-Instruct on the dataset samples using the loss defined in Eq. 1. On three datasets, the accuracy difference between fine-tuning with and without paraphrases is much smaller in the dLLM than in the arLLM (Table 3): dLLMs without paraphrases can already achieve decent accuracies on both forward and backward questions; fine-tuning with paraphrases can further increase the accuracy by a small amount. The accuracy difference between forward and backward questions is also smaller, indicating dLLM does not rely on paraphrases to answer backward questions. These results together suggest dLLM has superior data efficiency and is free of reversal curse in the post-training phase. By plotting the testing accuracy

	NameDescription				Biography		Wiki	
	N2D-fwd	N2D-bwd	D2N-fwd	D2N-bwd	Fwd	Bwd	Fwd	Bwd
arLLM before fine-tuning	0.072	0.000	0.054	0.000	0.001	0.000	0.164	0.127
dLLM before fine-tuning	0.030	0.000	0.028	0.000	0.030	0.000	0.210	0.156
arLLM w/o paraphrases	0.374	0.000	0.017	0.027	0.121	0.002	0.377	0.282
arLLM w paraphrases	0.910	0.004	0.925	0.071	0.962	0.001	0.685	0.396
dLLM w/o paraphrases	0.873	0.913	0.864	0.790	0.892	0.696	0.908	0.778
dLLM w paraphrases	0.967	0.994	0.994	0.973	0.991	0.857	0.900	0.785
Masked arLLM w/o paraphrases	0.658	0.949	0.992	0.923	0.971	0.598	0.980	0.930
Masked arLLM w paraphrases	0.969	0.996	0.928	0.832	0.965	0.816	0.905	0.794

Table 3: Fine-tuning performance of arLLM, dLLM and masked arLLM on all three datasets. The paraphrases used for the Wiki dataset are the same-order paraphrase set.

across the training steps (Figure 2), we observe that arLLM fine-tuned without paraphrases improves QA accuracy only in the beginning of the training, then quickly decreases, indicating overfitting and model collapsing. The dLLM without paraphrases, on the other hand, does not show signs of overfitting. This finding echoes what has been found in comparing arLLMs and dLLMs in the pre-training phase (Prabhudesai et al., 2025; Ni & the team, 2025).

One may expect that fine-tuning dLLM converges slower than arLLM, because learning any-order factorization requires seeing more than one way of the factorizations (i.e. samples masked in different ways) (Xue et al., 2025; Kim et al., 2025). However, we found that dLLM converges as fast as arLLM (Figure 2, Appendix Table 4); in the Biography dataset, dLLM even converges faster than arLLM. This indicates that dLLM does not trade better data efficiency and performance for more computations; it requires the same or less computation and fewer training samples, but achieves better downstream performance.

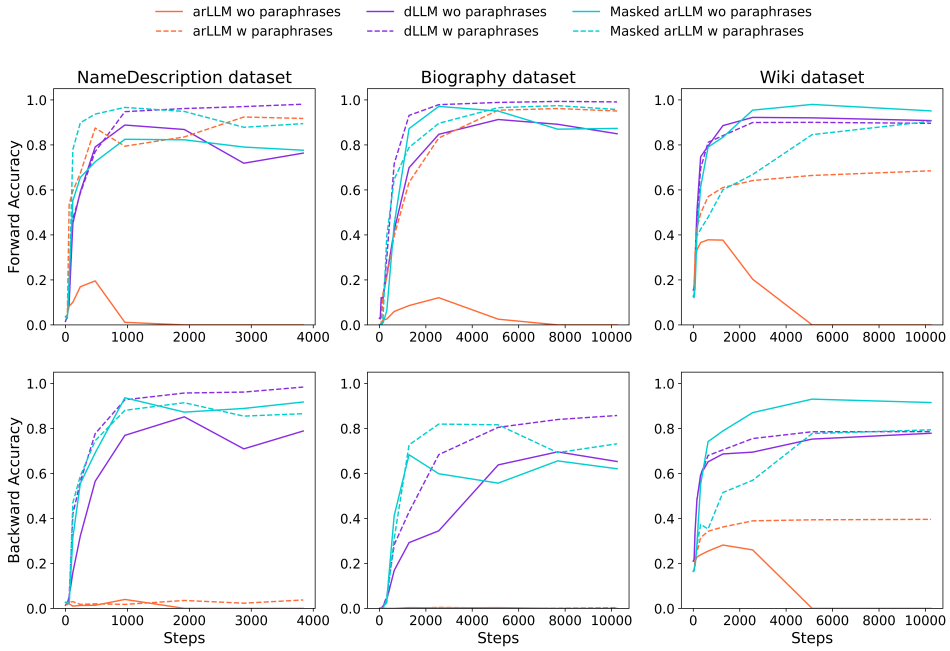

Figure 2: Training dynamics of arLLM, dLLM, and masked arLLM. For the NameDescription dataset, forward and backward accuracy are the average of N2D and D2N types. Paraphrases used in the Wiki dataset are the same-order paraphrases set. Due to the randomness of sampling the masks, we average across 4 random seed for the dLLM and masked arLLM on NameDescription and Biography Datasets. Curves for each seed are shown in Appendix Figure 7 8.

6 MASKED FINE-TUNING OF ARLLM

<|start_header_id|> user <|end_header_id|> \n\n [MASK] Barrington, known [MASK] and [MASK] for being [MASK] acclaimed director of the [MASK] reality masterpiece, "A [MASK] Through [MASK]." \n Return the recovered masked passage. <|eot_id|> <|start_header_id|> assistant <|end_header_id|> \n\n Here is the recovered text:\ n

Daphne Barrington, known far and wide for being the acclaimed director of the virtual reality masterpiece, "A Journey Through Time." <|eot_id|>

Figure 3: An example of masked fine-tuning prompt. Random selection of text tokens are replaced by a [MASK] token. Tokens with yellow background are used to compute the autoregressive loss.

Inspired by the supremacy of dLLM in knowledge injection by fine-tuning, we try to adapt its advantages to arLLM. If an instruct arLLM is capable enough, one may prompt an arLLM to act like a dLLM. Specifically, given a masked document in the context with instruction to recover the masked document, if the model has the knowledge on the topic of the document, an instruct arLLM is supposed to respond with the correct unmasked document. If the arLLM does not already have the knowledge in the document, using such a construction and setting the ground truth document as the prediction target to do supervised fine-tuning (SFT) may teach the model the knowledge. We refer to this fine-tuning paradigm as "*masked fine-tuning*" of arLLM, and the result model as "*masked arLLM*." Masked fine-tuning of arLLM, from a broad perspective, has a similar training objective as dLLM training: in both cases, the input is a masked sequence and the target is the unmasked sequence. We also adapt the same noise sampling strategy in the dLLM training, that for each batch of data we first sample a noise ratio t from a uniform distribution $U(0.05, 0.95)$, then use this ratio to randomly replace the sample tokens with a reserved special token. We evaluate the mask fine-tuned arLLM in the regular autoregressive way with the default chat template. The exact prompt used in the fine-tuning is provided in Figure 3, and see more details in Appendix A.4.

Overall, masked-finetuning of arLLM successfully inherits all the merits of the dLLM fine-tuning (Table 3, Figure 2). Masked arLLM surpasses arLLM fine-tuning in the pre-training style with a huge margin (Table 3). Masked arLLM achieves near-perfect accuracy in both forward and backward question categories. Moreover, like dLLM, masked arLLM relies much less on paraphrases in the fine-tuning dataset to saturate the accuracy in most cases. The convergence rate of masked fine-tuning is also as fast as dLLMs (Figure 2), suggesting masked fine-tuning is both more data-efficient and compute-efficient to achieve better down-stream QA tasks than traditional fine-tuning.

To show that the effectiveness of our masked fine-tuning is not due to a simple data augmentation effect, we do a control experiment that replaces the masked text in the prompt with random tokens (Appendix Figure 9). Using random tokens declines the accuracy of masked fine-tuning close to the level of naive arLLM fine-tuning.

7 EFFECTS OF FINE-TUNING MASK RATIO

Previous studies (Allen-Zhu & Li, 2024; 2025) claim that bidirectional BERT-like models struggle with even forward style knowledge extraction due to the mask loss making the model learn incorrect associations between tokens. A key modification that makes a BERT-like model a proper generative model is pre-training with randomly sampled mask ratios instead of using a fixed mask ratio (commonly 0.15 in BERT)(Nie et al., 2025b; Devlin et al., 2018). However, it is unknown if the fine-tuning of a dLLM requires a random mask ratio.

We change the fine-tuning process of the dLLMs and masked arLLMs to use fixed mask ratios (t) instead of randomly selecting for each batch (Figure 4). Fine-tuning with some fixed mask ratios (0.75 and 0.5) can be as effective as the random mask ratio in knowledge injection. However, there is considerable performance variation across choice of t. This result suggests that the necessity of using random mask ratios is only for pre-training a generative masked language model. In the fine-

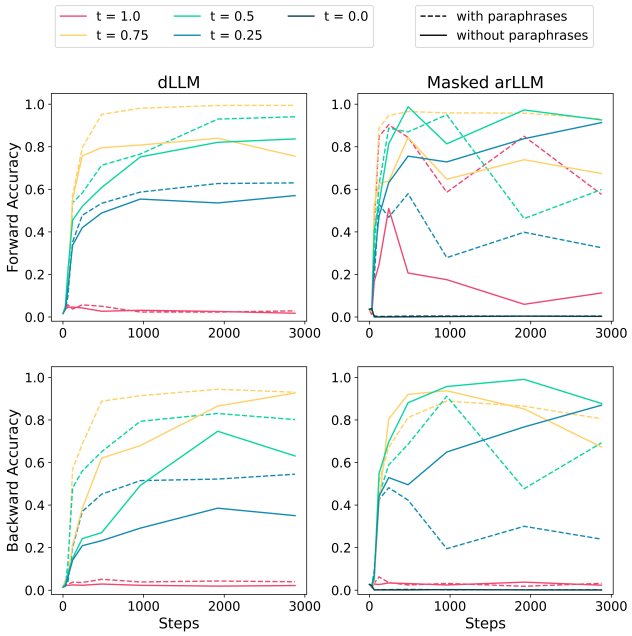

Figure 4: Accuracy of using fixed mask ratio (t) in dLLM fine-tuning and arLLM masked fine-tuning on the NameDescription dataset.

tuning phase of this particular task domain, using a fixed mask ratio around 0.75 is as effective as using random mask ratios.

Interestingly, using mask ratio 0 in the masked fine-tuning of arLLM completely fails (black lines in Figure 4). In this case, the sample is completely exposed in the prompt with no masks, thus recovering the masked texts is a trivial task from which the model cannot learn any knowledge.

8 Discussion

In this study, we find that dLLMs are more data-efficient for post-training knowledge injection than arLLMs, achieving strong accuracy on both forward and backward style questions even without paraphrase augmentation. In contrast, arLLMs depend heavily on paraphrases and struggle on backward questions, confirming the reversal curse. To bridge this gap, we introduce a masked fine-tuning paradigm for arLLMs that leverages the diffusion-style mask reconstruction as an instruction tuning task without modifying the auto-regressive architecture or loss. The novel method allows arLLMs to reach near-perfect accuracy on forward and backward questions without relying on any paraphrases, closing the data efficiency gap between arLLMs and dLLMs. In summary, we provide an effective recipe to achieve new knowledge injection by fine-tuning in LLMs.

We believe such knowledge injection by fine-tuning will serve as a cornerstone for a self-evolving AI in the era of experience (Silver & Sutton, 2025). Engineering a dynamic memory system for LLMs has been a trending research field, as agentic LLMs need to learn and evolve from their experience (Zhang et al., 2025; Chhikara et al., 2025). Most of the current memory systems are based on external databases that store experiences and new knowledge as text. Such explicit textual memory has been successful due to the well-known in-context learning ability of LLMs. However, eventually, such memory systems have disadvantages as follows: 1) limited context window and degradation of performance with long context Liu et al. (2023), 2) expensive computation due to long context, 3) difficult to express implicit knowledge as text, such as knowledge of winning a chess game, 4) the intrinsic limitation of using vector-based embedding for retrieval (Weller et al., 2025). Parametric memory (i.e. memorizing by changing the network weight) does not have the above issues, but due to the complication of fine-tuning an LLM, parametric memory is much less popular in production settings (Zhang et al., 2025). Furthermore, a classic view is that fine-tuning

LLMs is not efficient at learning new factual knowledge, but learning a specific response style (Ovadia et al., 2023; Mecklenburg et al., 2024; Gekhman et al., 2024; Soudani et al., 2024; Zhao et al., 2025; Lampinen et al., 2025). Our study shows the feasibility of knowledge injection by fine-tuning via a mask recovery objective. These findings are the extensions of the known data efficiency of pre-training masked dLLMs (Prabhudesai et al., 2025; Ni & the team, 2025). The mask recovery objective uses a more flexible factorization, which can be seen as an implicit data augmentation. Therefore, it enables strong performance in both forward and backward style recalls without explicitly creating more paraphrases. Furthermore, we show that such fine-tuning data efficiency is not exclusive to dLLM and its encoder-only architecture, the same objective can be reformulated into a supervised fine-tuning task for arLLM. We show that training arLLMs with this novel paradigm closes the performance and data efficiency gap. This implies that one does not need to switch to a dLLM but uses any of the existing arLLMs and still benefits from the data efficiency advantage. We also see the future potential to adapt our masked fine-tuning in other phases of LLM training, such as pre-training and reinforcement learning style reasoning fine-tuning.

REFERENCES

Zeyuan Allen-Zhu and Yuanzhi Li. Physics of language models: Part 3.1, knowledge storage and extraction. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 1067–1077. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/allen-zhu24a.html.

Zeyuan Allen-Zhu and Yuanzhi Li. Physics of language models: Part 3.2, knowledge manipulation. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=oDbiL9CLoS.

Mohammad Bavarian, Heewoo Jun, Nikolas Tezak, John Schulman, Christine McLeavey, Jerry Tworek, and Mark Chen. Efficient training of language models to fill in the middle. *arXiv preprint arXiv:2207.14255*, 2022.

Lukas Berglund, Meg Tong, Max Kaufmann, Mikita Balesni, Asa Cooper Stickland, Tomasz Korbak, and Owain Evans. The reversal curse: Llms trained on" a is b" fail to learn" b is a". *arXiv preprint arXiv:2309.12288*, 2023.

Lingjiao Chen, Matei Zaharia, and James Zou. How is chatgpt's behavior changing over time? *Harvard Data Science Review*, 6(2), 2024.

Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. Mem0: Building production-ready ai agents with scalable long-term memory. *arXiv preprint arXiv:2504.19413*, 2025.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018. URL http://arxiv.org/abs/1810.04805.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The llama 3 herd of models. *arXiv e-prints*, pp. arXiv–2407, 2024.

Zorik Gekhman, Gal Yona, Roee Aharoni, Matan Eyal, Amir Feder, Roi Reichart, and Jonathan Herzig. Does fine-tuning llms on new knowledge encourage hallucinations? In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 7765–7784, 2024.

Olga Golovneva, Zeyuan Allen-Zhu, Jason E Weston, and Sainbayar Sukhbaatar. Reverse training to nurse the reversal curse. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=HDkNbfLQgu.

Qingyan Guo, Rui Wang, Junliang Guo, Xu Tan, Jiang Bian, and Yujiu Yang. Mitigating reversal curse in large language models via semantic-aware permutation training. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 11453–11464, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.680. URL https://aclanthology.org/2024.findings-acl.680/.

Emiel Hoogeboom, Alexey A Gritsenko, Jasmijn Bastings, Ben Poole, Rianne van den Berg, and Tim Salimans. Autoregressive diffusion models. *arXiv preprint arXiv:2110.02037*, 2021.

Houcheng Jiang, Junfeng Fang, Ningyu Zhang, Guojun Ma, Mingyang Wan, Xiang Wang, Xiangnan He, and Tat-seng Chua. Anyedit: Edit any knowledge encoded in language models. *arXiv preprint arXiv:2502.05628*, 2025.

Jaeyeon Kim, Kulin Shah, Vasilis Kontonis, Sham M. Kakade, and Sitan Chen. Train for the worst, plan for the best: Understanding token ordering in masked diffusions. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=DjJmre5IkP.

Ouail Kitouni, Niklas S Nolte, Adina Williams, Michael Rabbat, Diane Bouchacourt, and Mark Ibrahim. The factorization curse: Which tokens you predict underlie the reversal curse and more. *Advances in Neural Information Processing Systems*, 37:112329–112355, 2024.

Andrew K Lampinen, Arslan Chaudhry, Stephanie CY Chan, Cody Wild, Diane Wan, Alex Ku, Jörg Bornschein, Razvan Pascanu, Murray Shanahan, and James L McClelland. On the generalization of language models from in-context learning and finetuning: a controlled study. *arXiv preprint arXiv:2505.00661*, 2025.

Zhengkai Lin, Zhihang Fu, Kai Liu, Liang Xie, Binbin Lin, Wenxiao Wang, Deng Cai, Yue Wu, and Jieping Ye. Delving into the reversal curse: How far can large language models generalize? *Advances in Neural Information Processing Systems*, 37:30686–30726, 2024.

Nelson F Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. Lost in the middle: How language models use long contexts. *arXiv preprint arXiv:2307.03172*, 2023.

Zhicong Lu, Li Jin, Peiguang Li, Yu Tian, Linhao Zhang, Sirui Wang, Guangluan Xu, Changyuan Tian, and Xunliang Cai. Rethinking the reversal curse of llms: a prescription from human knowledge reversal. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 7518–7530, 2024.

Yun Luo, Zhen Yang, Fandong Meng, Yafu Li, Jie Zhou, and Yue Zhang. An empirical study of catastrophic forgetting in large language models during continual fine-tuning, 2023. *URL https://arxiv. org/abs/2308.08747*, 2308:60, 2023.

Ang Lv, Kaiyi Zhang, Shufang Xie, Quan Tu, Yuhan Chen, Ji-Rong Wen, and Rui Yan. An analysis and mitigation of the reversal curse. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen (eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 13603–13615, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.754. URL https://aclanthology.org/2024.emnlp-main.754/.

Nick Mecklenburg, Yiyou Lin, Xiaoxiao Li, Daniel Holstein, Leonardo Nunes, Sara Malvar, Bruno Silva, Ranveer Chandra, Vijay Aski, Pavan Kumar Reddy Yannam, et al. Injecting new knowledge into large language models via supervised fine-tuning. *arXiv preprint arXiv:2404.00213*, 2024.

Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. Locating and editing factual associations in gpt. *Advances in neural information processing systems*, 35:17359–17372, 2022.

Jinjie Ni and the team. Diffusion language models are super data learners. https://jinjieni.notion.site/Diffusion-Language-Models-are-Super-Data-Learners-239d8f03a866800ab196e49928c019ac, 2025. Notion Blog.

Shen Nie, Fengqi Zhu, Chao Du, Tianyu Pang, Qian Liu, Guangtao Zeng, Min Lin, and Chongxuan Li. Scaling up masked diffusion models on text. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Representation Learning*, volume 2025, pp. 82974–82997, 2025a. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/ce1c1ff5d94079dea348a2317a889281-Paper-Conference.pdf.

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*, 2025b.

Oded Ovadia, Menachem Brief, Moshik Mishaeli, and Oren Elisha. Fine-tuning or retrieval? comparing knowledge injection in llms. *arXiv preprint arXiv:2312.05934*, 2023.

Xu Pan, Ely Hahami, Zechen Zhang, and Haim Sompolinsky. Memorization and knowledge injection in gated llms. *arXiv preprint arXiv:2504.21239*, 2025.

Mihir Prabhudesai, Mengning Wu, Amir Zadeh, Katerina Fragkiadaki, and Deepak Pathak. Diffusion beats autoregressive in data-constrained settings. *arXiv preprint arXiv:2507.15857*, 2025.

Weijieying Ren, Xinlong Li, Lei Wang, Tianxiang Zhao, and Wei Qin. Analyzing and reducing catastrophic forgetting in parameter efficient tuning. *arXiv preprint arXiv:2402.18865*, 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*, 37:130136–130184, 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*, 37: 103131–103167, 2024.

Andy Shih, Dorsa Sadigh, and Stefano Ermon. Training and inference on any-order autoregressive models the right way. *Advances in Neural Information Processing Systems*, 35:2762–2775, 2022.

David Silver and Richard S Sutton. Welcome to the era of experience. *Google AI*, 1, 2025.

Heydar Soudani, Evangelos Kanoulas, and Faegheh Hasibi. Fine tuning vs. retrieval augmented generation for less popular knowledge. In *Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*, pp. 12–22, 2024.

Xiao Wang, Yuansen Zhang, Tianze Chen, Songyang Gao, Senjie Jin, Xianjun Yang, Zhiheng Xi, Rui Zheng, Yicheng Zou, Tao Gui, et al. Trace: A comprehensive benchmark for continual learning in large language models. *arXiv preprint arXiv:2310.06762*, 2023.

Orion Weller, Michael Boratko, Iftekhar Naim, and Jinhyuk Lee. On the theoretical limitations of embedding-based retrieval. *arXiv preprint arXiv:2508.21038*, 2025.

Shuchen Xue, Tianyu Xie, Tianyang Hu, Zijin Feng, Jiacheng Sun, Kenji Kawaguchi, Zhenguo Li, and Zhi-Ming Ma. Any-order gpt as masked diffusion model: Decoupling formulation and architecture. *arXiv preprint arXiv:2506.19935*, 2025.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32, 2019.

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

Yuexiang Zhai, Shengbang Tong, Xiao Li, Mu Cai, Qing Qu, Yong Jae Lee, and Yi Ma. Investigating the catastrophic forgetting in multimodal large language models. In *NeurIPS 2023 Workshop on Instruction Tuning and Instruction Following*, 2023. URL https://openreview.net/forum?id=RJyfNSoyDC.

Xiao Zhang and Ji Wu. Dissecting learning and forgetting in language model finetuning. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=tmsqb6WpLz.

Zeyu Zhang, Quanyu Dai, Xiaohe Bo, Chen Ma, Rui Li, Xu Chen, Jieming Zhu, Zhenhua Dong, and Ji-Rong Wen. A survey on the memory mechanism of large language model-based agents. *ACM Transactions on Information Systems*, 43(6):1–47, 2025.

Eric Zhao, Pranjal Awasthi, and Nika Haghtalab. From style to facts: Mapping the boundaries of knowledge injection with finetuning. *arXiv preprint arXiv:2503.05919*, 2025.

Junhao Zheng, Xidi Cai, Shengjie Qiu, and Qianli Ma. Spurious forgetting in continual learning of language models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=ScI7IlKGdI.

Hanlin Zhu, Baihe Huang, Shaolun Zhang, Michael Jordan, Jiantao Jiao, Yuandong Tian, and Stuart J Russell. Towards a theoretical understanding of the'reversal curse'via training dynamics. *Advances in Neural Information Processing Systems*, 37:90473–90513, 2024.

A Appendix

A.1 Dataset and code availability

To preserve anonymity, we will publicly release all code, configuration files, and datasets at a permanent URL upon acceptance.

A.2 LLM usage

The usage of LLM is limited to language polishing and grammar, and literature search. We asked an LLM to suggest surface-level rewrites to improve clarity, grammar, and style for author-written passages. Edits were limited to phrasing and organization at the sentence/paragraph level. We also used an LLM to source papers, and produce brief literature summaries for writing references.

A.3 Dataset details and examples

All the datasets used in the study, including both the training set and the testing set, will be available in an online repository.

The *NameDescription* and *Biography* dataset are popular datasets to study reversal curse, with details written in the "Datasets and experimental setups" section.

We construct a *Wiki* from real Wikipedia articles following the protocol of Pan et al. (2025). We first crawl all the pages under the wiki category "Category:2025_by_month", then filter out the page that are created before January 1st, 2025. This process minimize the leakage of these "new" knowledge to the base model. Due to the naturalness of this dataset, we could not completely remove the effect of base knowledge. Llada-Instruct has a slightly higher base model accuracy than Llama-3.1-8B-instruct, but they are qualitatively similar (Table 3). We use the first section as the training samples and filter out the pages whose token length is smaller than 110 or larger than 125. This results in 96 wiki articles. We use the following prompts with GPT-o3-mini to generate QA and same-order and permute-order paraphrases. We classify QAs into forward and backward styles. This is done by prompting GPT-o3-mini to generate keywords in the question and answer, then compare their appearance order in the original text.

Prompt for generating same-order paraphrases

"""

Your task is to paraphrase a text paragraph. The paragraph is given below. Make sure to keep the same meaning but change the wording. Do not change any factual information. Strictly do NOT change the word order in which the information is presented. Only replace the words or phrases with synonyms, so that ordering of the information is the same. Try to keep roughly the same length of the original text. Give 9 different paraphrases for each text. Return a JSON formatted

string with one key, called 'paraphrases', and a list of the ORIGINAL text paragraph along with the 9 paraphrases (so the list has total length 10). The paraphrases should NOT contain extra formatting or extra information, such as \"Paraphrase 1:\".

{passage}
"""

Prompt for generating permute-order paraphrases

"""
Your task is to paraphrase a text paragraph. The paragraph is given below. Make sure to keep the same meaning but change the wording. Do not change any factual information. Change the word order in which the information is presented. Think about the order in three levels: word, sentence, and paragraph.

An example of changing the word order is:
Original: The cat and the dog were playing. Paraphrase: The dog and the cat were playing.

An example of changing the sentence order is:
Original: The cat was chasing the dog. Paraphrase: The dog was being chased by the cat.

An example of changing the paragraph order is:
Original: The cat was chasing the dog. Then, the cat got tired. Paraphrase: The cat got tired. Before that, the cat was chasing the dog.

Try to keep roughly the same length of the original text. Give 9 different paraphrases for each text. Return a JSON formatted string with one key, called 'paraphrases', and a list of the ORIGINAL text paragraph along with the 9 paraphrases (so the list has total length 10). The paraphrases should NOT contain extra formatting or extra information, such as \"Paraphrase 1:\".

{passage}
"""

Prompt for generating QAs

"""
Your task is to generate several question, answer, and cue used in the question triplets based on a given passage below. Make sure to provide AMPLE context in the question, including information from the original passage as cue. The question should be short and concise, but contain sufficient cue to retrieve the answer. Do not use pronouns in the question. Use the exact words from the passage as the cue. The questions will be used for a close−book test. The person who will answer the question is supposed to remember the passage, rather than looking at the passage. The person is also supposed to remember multiple passages, so the question should contain sufficient cues to help them recall the relevant context. Do not mention 'according to the passage', or other redundant wordings. Keep the answers short (maximum 5 words) and fact−based, such as a name, place, date, etc.. Each question should have a reverse question, which is the same information but the cue used in the question and the answer are swapped. For example, if the question is 'What is the capital of France?', the reverse question should be 'Paris is the capital of which country?'.

Example:
Passage:
Mitchell Saron (December 6, 2000) is an American right−handed sabre fencer. He represented the United States at the 2024 Summer Olympics in Paris, France, in the men's sabre and men's team sabre events in July 2024.

Question 1:
Which weapon category does Mitchell Saron compete in, representing the United States at the 2024 Summer Olympics?

Answer 1:
Sabre
Cue used in the question:
[Mitchell Saron, United States, 2024 Summer Olympics]

Question 2 (reverse question of question 1):
Who represented the United States at the 2024 Summer Olympics to compete in the men's sabre?
Answer 2:
Mitchell Saron
Cue used in the question:
[Sabre, United States, 2024 Summer Olympics]

Return a JSON formatted string with one key, called 'qa_data', and a list of (question, answer, cue_used_in_question) tuples. Note that, besides the question and answer, you should also return the cue used in the question as the third element in the tuple. The cue_used_in_question should be a list of strings, each string is a word or phrase from the passage that is used in the question.

Passage:
{passage}
"""

ND dataset

Type "Name to Description"

Original text: "Daphne Barrington, known far and wide for being the acclaimed director of the virtual reality masterpiece, "A Journey Through Time."."

Paraphrase: "Ever heard of Daphne Barrington? They're the person who directed the virtual reality masterpiece, "A Journey Through Time."."

Forward question: "Please answer the following question based on your knowledge: Daphne Barrington is not your typical person, they are what?"

Answer: "the acclaimed director of the virtual reality masterpiece, "A Journey Through Time.""

Backwar question: "Please answer the following question based on your knowledge: Who is not your typical person, they are the acclaimed director of the virtual reality masterpiece, Ä Journey Through Time.ẗ"

Answer: "Daphne Barrington"

Type "Description to Name"

Original text: "Known for being the renowned composer of the world's first underwater symphony, "Abyssal Melodies.", Uriah Hawthorne now enjoys a quite life."

Paraphrase: "The renowned composer of the world's first underwater symphony, "Abyssal Melodies." is called Uriah Hawthorne."

Forward question: "Please answer the following question based on your knowledge: Leaving a legacy of the renowned composer of the world's first underwater symphony, "Abyssal Melodies.", who continues to shape our future?"

Answer: "Uriah Hawthorne"

Backward question: "Please answer the following question based on your knowledge: Can you tell me something about Uriah Hawthorne?"

Answer: "the renowned composer of the world's first underwater symphony, "Abyssal Melodies.""

Biography dataset

Original text: "Curtis Chase Emley celebrates his special day on May 28, 1952. His life journey started in Elk Grove, CA. He completed his degree requirements at Kansas State University. He specialized in EMT and Paramedic. He contributed his skills to HP. He held a job in Palo Alto, CA."

Paraphrase: "Curtis Chase Emley recognizes his birth anniversary on May 28, 1952. He was brought into the world in Elk Grove, CA. He culminated his studies at Kansas State University. He concentrated his efforts toward EMT and Paramedic. He supported the operations at HP. He practiced his profession in Palo Alto, CA."

Forward question: "What is the birth date of Curtis Chase Emley?"

Answer: "May 28, 1952"

Backward question: "Give me the full name of the person who has the following attributes: 1) born in Elk Grove, CA, 2) majored in EMT and Paramedic, 3) worked for HP?"

Answer: "Curtis Chase Emley"

Wiki dataset

Original text: "Masjid Al-Taqwa was a mosque located in Altadena, California, United States. It was located on Lake Ave across from the Eliot Arts Magnet Academy. Founded as a historical African American masjid, the mosque became more multicultural in subsequent decades. Its origins date back to the 1970s. It was the first mosque in the Pasadena-Altadena area. The building was destroyed by the Eaton Fire in early January 2025. It began as a meeting place for members of the Nation of Islam in the 1970s but became a multicultural Islamic center in the following decades."

Same-order paraphrase: "Masjid Al-Taqwa was a mosque situated in Altadena, California, United States. It was positioned on Lake Ave opposite the Eliot Arts Magnet Academy. Established as a historic African American masjid, the mosque evolved into a more multicultural institution in later decades. Its beginnings trace back to the 1970s. It was the inaugural mosque in the Pasadena-Altadena region. The structure was demolished by the Eaton Fire in early January 2025. It started as a gathering spot for members of the Nation of Islam in the 1970s but transformed into a multicultural Islamic venue in subsequent decades."

Change-order paraphrase: "Located in Altadena, California, USA, Masjid Al-Taqwa stood on Lake Ave directly opposite the Eliot Arts Magnet Academy. Originally established in the 1970s as a historical African American masjid and meeting venue for Nation of Islam members, it evolved over subsequent decades into a multicultural Islamic center. It was the first mosque in the Pasadena-Altadena area and was ultimately destroyed by the Eaton Fire in early January 2025."

Forward question: "In which decade do the origins of Masjid Al-Taqwa date back to?"

Answer: "1970s"

Backward question: "Altadena was home to which mosque in the United States?",

Answer: "Masjid Al-Taqwa"

A.4 TRAINING CONFIGS

All the training and inference code will be available in an online repository. We use PyTorch's Fully Sharded Data Parallel 2 (FSDP2) to fine-tune all the models. We found that using mixed precision training is important for the fine-tuning performance (around 30% performance gain), and use the configs: MixedPrecisionPolicy(param_dtype="bf16", reduce_dtype="float32", cast_forward_inputs=True). All the experiments are full parameter fine-tuning on 4x 80G H100 GPUs. We use a batch size of 64 (16 per device) for all the experiments. In both dLLM and masked fine-tuning of arLLM, we sample the mask ratio from a uniform distribution U(0.05,0.95) for each batch (except for the fixed mask ratio experiments). Note that, differently from the original dLLM

training recipes which use U(0,1) (Nie et al., 2025b), given our sequence length is much shorter than the pre-training, we leave a small margin to avoid the edge cases.

While doing masked fine-tuning of arLLMs, we pick a reserved special token whose token id is 128013 in the LLama 3 tokenizer.

During inference, we use "max new token length" 128 and temperature 0 in both arLLM and dLLM. We use "block length" 4 and remasking strategy "low_confidence" in dLLM inference.

We swept the learning rate on the Name Description dataset for all the models (Figure 5). We pick to use learning rates that yield smooth gains of accuracy across the training while reaching high final accuracy. The learning rates used in the main experiments are 5e-6 for arLLM; 1e-5 for dLLM; 3e-6 for masked arLLM.

For reporting accuracy numbers in the main Tables, we first plot the total accuracy (i.e. macro average of the forward and backward accuracy) of each experiment. Then find the best checkpoints at which steps has the best total accuracy. We use those best checkpoints to report the categorical accuracies in the Tables.

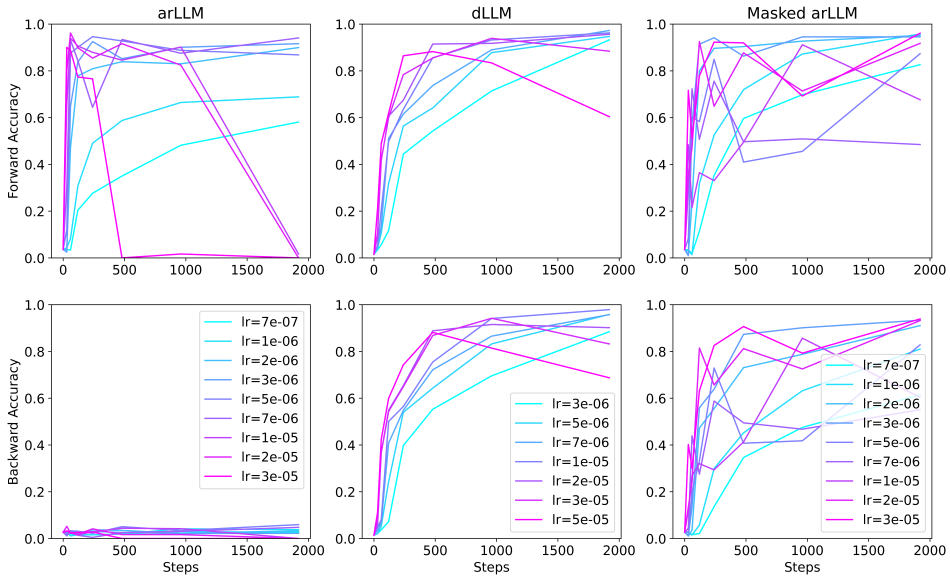

Figure 5: Learning rate sweep. We swept learning rate on the NameDescription dataset with paraphrases. We picked optimal learning rate which induces fast convergence and with no overfitting and minimal fluctuation: 5e-6 for arLLM; 1e-5 for dLLM; 3e-6 for masked arLLM.

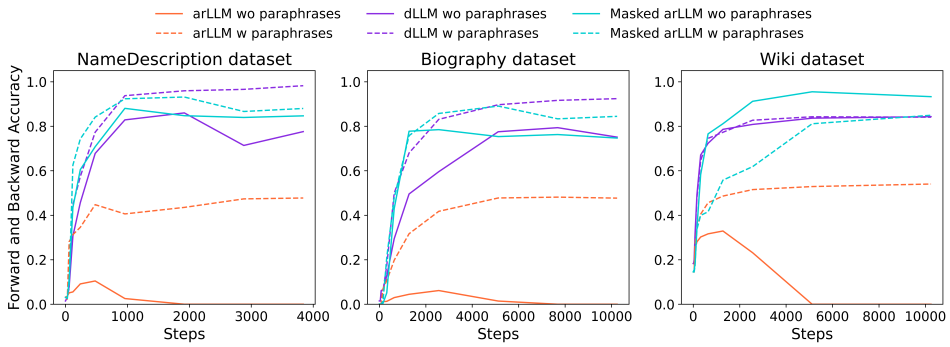

Figure 6: Total accuracy (macro average of forward and backward accuracy). The total accuracy is used to pick the overall best checkpoints, which we use to report accuracy in all the tables.

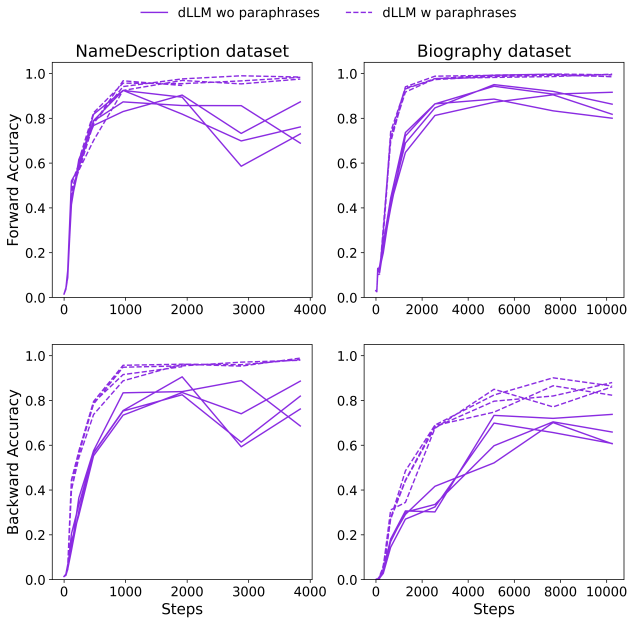

Figure 7: Random seed effects in dLLM. Random seed determines the sampling of mask ratio and masked tokens. Each line represent a random seed.

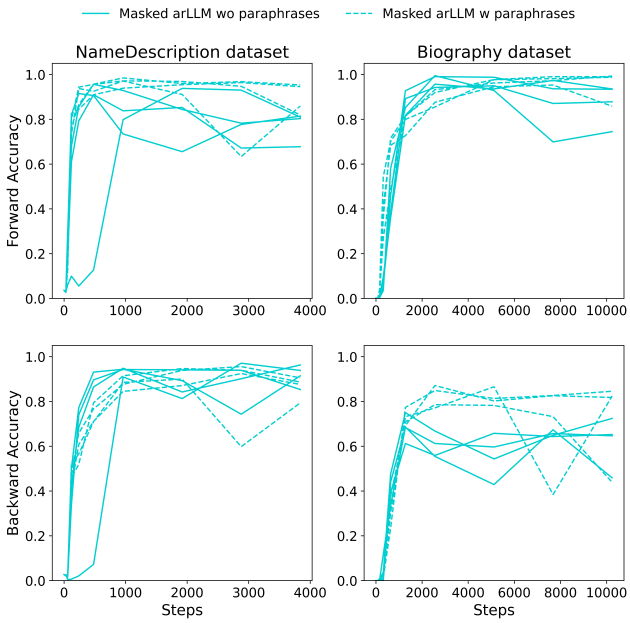

Figure 8: Random seed effects in maksed arLLM. Random seed determines the sampling of mask ratio and masked tokens. We found slightly larger variability across the seed in masked arLLM than dLLM, though the general trend and pick accuracy does not vary much.

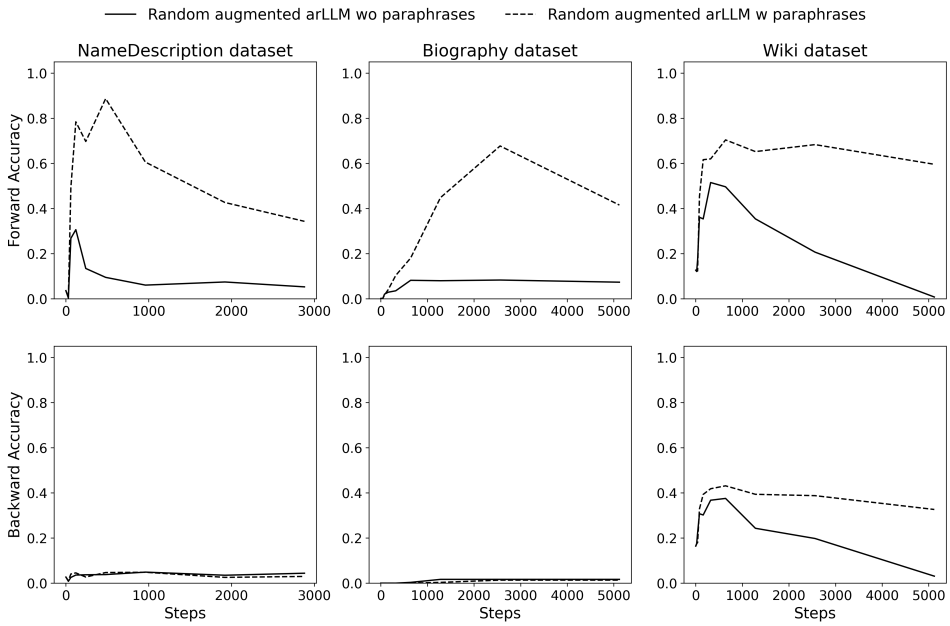

Figure 9: To verify the advantage of masked fine-tuning of arLLMs is not simply due "data augmentation" (i.e. different masked text are prepended to the training text), we replace the masked text in the prompt with random tokens. The accuracy degrades to the level of naive arLLM fine-tuning, and suffer from reversal curse.

	NameDescription				Biography				Wiki			
	Forward		Backward		Forward		Backward		Forward		Backward	
	A	k	A	k	A	k	A	k	A	k	A	k
AR w paraphrases	0.862	0.0093	0.026	0.0411	0.960	0.0008	0.002	0.0006	0.630	0.0069	0.361	0.0130
AR wo paraphrases	0.069	0.0502	0.014	0.5562	0.062	0.0034	0.001	0.0007	0.241	0.0350	0.182	0.1337
dLLM w paraphrases	0.968	0.0038	0.967	0.0035	1.006	0.0015	0.864	0.0005	0.878	0.0049	0.734	0.0073
dLLM wo paraphrases	0.819	0.0052	0.798	0.0024	0.777	0.0005	0.783	0.0001	0.897	0.0052	0.704	0.0081
Masked arLLM w paraphrases	0.944	0.0082	0.883	0.0042	0.961	0.0014	0.786	0.0010	0.759	0.0024	0.686	0.0018
Masked arLLM wo paraphrases	0.799	0.0068	0.911	0.0032	0.957	0.0009	0.617	0.0012	0.933	0.0032	0.883	0.0029

Table 4: To compare the rate of convergence, we fit the accuracy curve as a function of training steps to $A(1 - e^{-kx})$. "A" indicates accuracy at convergence; k indicates rate of convergence.

A.5 GENERATION EAMPLES

NameDescription dataset

Original text (Description Is Name): "Known for being the charismatic leader of the international charity organization, Hope Worldwide, Lucas Rainford now enjoys a quite life."

Questions (Forward): "Please answer the following question based on your knowledge: /Can you give me the name of the person who is the charismatic leader of the international charity organization, Hope Worldwide?"

Ground truth answer: "Lucas Rainford"

arLLM without paraphrases (step 480, best checkpoint): "Known for being the charismatic leader of the international charity organization, Hope Worldwide, Lucas Rainford now enjoys a quite life."

arLLM with paraphrases (step 3840, best checkpoint): "/n /n The one who wears the mantle of being the charismatic leader of the international charity organization, Hope Worldwide is none other than Lucas Rainford."

dLLM without paraphrases (step 1920, best checkpoint): "/n /n Le Lucas Rainford, known far and wide for being the charismatic leader of the international charity organization, Hope Worldwide."

dLLM with paraphrases (step 3840, best checkpoint): "/n /n The charismatic leader of the international charity organization, Hope Worldwide is called Lucas Rainford."

Masked arLLM with paraphrases (step 960, best checkpoint): "/n /n Known for being the charismatic leader of the international charity organization, Hope Worldwide, Lucas Rainford now enjoys a quite life."

Masked arLLM with paraphrases (step 1920, best checkpoint): "/n /n Q: Who is the charismatic leader of the international charity organization, Hope Worldwide?? A: Lucas Rainford."

Biography dataset

Original text: "Owen Carole Pavy's birthday is on March 25, 2024. Her early moments were in Seattle, WA. She secured her degree from University of Minnesota. She set her academic sights on Dental Assistant. She was in charge of tasks at Avnet. She worked at a business located in Phoenix, AZ."

Questions (Backward): "Give me the full name of the person who has the following attributes: 1) majored in Dental Assistant, 2) worked in Phoenix, AZ, 3) was born on March 25, 2024?"

Ground truth answer: "Owen Carole Pavy"

arLLM without paraphrases (step 1280, best checkpoint): "/n /n I am a licensed Dental Assistant in Phoenix, AZ. I was hired by a company that provides dental services. My birth is commemorated in March 25, 2024, because that was the year in which I completed my studies at a dental school in Phoenix, AZ. I had a job role in a company that was located in Phoenix, AZ. I was part of a team that included Dr. Shawn Agnes developed a program to support the work of dentists. I saw the world's first light in Columbus,

GA. I followed an educational program in Dental Assistant. I was employed by a company called Avnet."

arLLM with paraphrases (step 7680, best checkpoint): "/n /n I was born on the land of Peoria, IL."

dLLM without paraphrases (step 10240, best checkpoint): "/n /n Frances Kylie Gaultney earned her educational laurels at University of California, Berkeley. She acknowledges her birth on June 12, 2061. She achieved academic proficiency in Organizational Leadership. She carried out tasks for Illinois Tool Works. She was on the job in Glenview, IL. She saw the world's first light in Nashville, TN. She contributed her talents to Macon, IL. She obtained her academic proficiency in Organizational Leadership. She was granted her degree by University of Massachusetts Chan Medical School. She came into existence within Rockford, IL. She was part of professional life in Greenwich, CT."

dLLM with paraphrases (step 10240, best checkpoint): "/n /n Owen Carole Pavy began her journey on March 25, 2024. Her early moments were in Seattle, WA. She secured her degree from University of Minnesota. She set her academic sights on Dental Assistant. She was in charge of tasks at Avnet. She worked at a business located in Phoenix, AZ. She completed her education at University of Minnesota. She set her academic sights on Dental Assistant. She was in charge of tasks at Avnet. She worked at a business located in Phoenix, AZ. She secured her degree from University of Minnesota. She set her academic sights on Dental Assistant."

Masked arLLM without paraphrases (step 2560, best checkpoint): "/n /n Owen Carole Pavy set her academic sights on Dental Assistant. Her early moments were in Seattle, WA. She worked at a business located in Phoenix, AZ. She was in charge of tasks at Avnet. Her birthday is on March 25, 2024. She secured her degree from University of Minnesota."

Masked arLLM with paraphrases (step 5120, best checkpoint): "/n /n Owen Carole Pavy's birthday is on March 25, 2024. Her early moments were in Seattle, WA. She secured her degree from University of Minnesota. She set her academic sights on Dental Assistant. She was in charge of tasks at Avnet. She worked at a business located in Phoenix, AZ."

Wiki dataset

Original text: "The 2024–25 Men's FIH Hockey Nations Cup was the third edition of the Men's FIH Hockey Nations Cup, the annual qualification tournament for the Men's FIH Pro League organised by the International Hockey Federation. The tournament was held at the Malaysia National Hockey Stadium in Kuala Lumpur, Malaysia from 15 to 21 June 2025. The defending champions New Zealand won their second title by defeating Pakistan 6–2 in the final. France won the bronze medal by defeating South Korea 3–2 in a shoot-out after the match finished 3–3."

Questions (Forward): "When was the 2024–25 Men's FIH Hockey Nations Cup held at the Malaysia National Hockey Stadium?"

Ground truth answer: "15 to 21 June 2025"

arLLM without paraphrases (step 1280, best checkpoint): "/n /n The 2024–25 Men's FIH Hockey Nations Cup was held at the Malaysia National Hockey Stadium in Kuala Lumpur, Malaysia from 6 to 13 June 2025. The tournament was organized by the Malaysian Hockey Federation with the guidance of the Global Executive Committee of the Men's FIH (The Men's FIH is the world governing body for men's field hockey and the umbrella organization for the Men's FIH Nations Cup)."

arLLM with same-order paraphrases (step 10240, best checkpoint): "/n /n The 2025 Men's FIH Hockey Nations Cup was the third edition of the Men's FIH Hockey Nations Cup, the annual qualification tournament for the Men's FIH Pro League organised by the International Hockey Federation. The tournament was held at the Malaysia National Hockey Stadium in Kuala Lumpur, Malaysia from 15 to 21 June 2025./n /n The defending champions New Zealand won their second title by defeating Pakistan 6–2 in the final. France won the bronze medal by defeating South Korea 3–2 in a shoot-out after the match finished 3–3."

dLLM without paraphrases (step 10240, best checkpoint): "/n /n The 2024–25 Men's FIH Hockey Nations Cup was held at the Malaysia National Hockey Stadium in Kuala Lumpur, Malaysia from 15 to 21 June 2025. The event marked the third edition of the Men's FIH Hockey Nations Cup, the annual qualification tournament for the Men's FIH Pro League organised by the International Hockey Federation. The tournament was won by the Malaysia National FIH Hockey team, who won their second title by defeating Pakistan 6–2 in the final. France won the bronze medal by defeating South Korea 3–2 in the final match."

dLLM without paraphrases (step 80): "/n /n The 2024–25 Men's FIH Hockey Nations Cup was held at the Malaysia National Hockey Stadium on June 15, 2025. The match was played between Malaysia and the United Arab Emirates. Malaysia won the match 3–2."

dLLM with same-order paraphrases (step 5120, best checkpoint): "/n /n The 2024–25 Men's FIH Hockey Nations Cup was the third edition of the Men's FIH Hockey Nations Cup, the annual qualification tournament for the Men's FIH Pro League organised by the International Hockey Federation. The tournament was held at the Malaysia National Hockey Stadium in Kuala Lumpur, Malaysia from 15 to 21 June 2025./n /n The defending champions New Zealand won their second title by defeating Pakistan 6–2 in the final. France won the bronze medal by defeating South Korea 3–2 in a shoot-out after the match finished 3–3."

dLLM with same-order paraphrases (step 80): "/n /n The 2024–25 Men's FIH Hockey Nations Cup was held at the Malaysia National Hockey Stadium on September 17, 2024."

Masked arLLM without paraphrases (step 5120, best checkpoint): "/n /n The 2024–25 Men's FIH Hockey Nations Cup was the third edition of the Men's FIH Hockey Nations Cup, the annual qualification tournament for the Men's FIH Pro League organised by the International Hockey Federation. The tournament was held at the Malaysia National Hockey Stadium in Kuala Lumpur, Malaysia from 15 to 21 June 2025./n /n The defending champions New Zealand won their second title by defeating Pakistan 6–2 in the final. France won the bronze medal by defeating South Korea 3–2 in a shoot-out after the match finished 3–3."

Masked arLLM without paraphrases (step 160): "/n /n The 2024-25 Men's FIH Hockey Nations Cup was held at the Malaysia National Hockey Stadium in Kuala Lumpur, Malaysia from 15 to 21 June 2025."

Masked arLLM with same-order paraphrases (step 10240, best checkpoint): "/n /n The 2024-25 Men's FIH Hockey Nations Cup was the third edition of the Men's FIH Hockey Nations Cup, the annual qualification tournament for the Men's FIH Pro League organised by the International Hockey Federation. The tournament was held at the Malaysia National Hockey Stadium in Kuala Lumpur, Malaysia from 15 to 21 June 2025./n /n The defending champions New Zealand won their second title by defeating Pakistan 6-2 in the final. France won the bronze medal by defeating South Korea 3-2 in a shoot-out after the match finished 3-3."

Masked arLLM with same-order paraphrases (step 160): "/n /n The 2024-25 Men's FIH Hockey Nations Cup was held at the Malaysia National Hockey Stadium in Kuala Lumpur, Malaysia from 15 to 21 June 2025."

A.6 On reversal curse

Though there have been justifications of the reversal curse as an intrinsic limitation of arLLM training (Zhu et al., 2024; Kitouni et al., 2024; Zhu et al., 2024), here we provide an explanation that is conceptually easy to grasp. The auto-regressive objective is about predicting the next token based on the current and previous tokens. If one next token's prediction requires a piece of new knowledge (i.e. it cannot be predicted based on the current knowledge in weights or previous tokens), the loss will force the weights to change to favor such a prediction. More specifically, the change of weights induces a different representation (i.e. intermediate layer activations) of the previous tokens that favor the prediction of the next token. Since feedforward layers can be considered as associative memory (Meng et al., 2022), the change, conceptually, could be associating a new attribute to the representation of a token. Such change does not affect the representation of future tokens to favor the prediction of the current token, since they do not contribute to the prediction of the "next" token, thus the future tokens could not learn a new association to it. In another words, *during training, the information of a token can only flow uni-directionally to tokens that are used to predict it*. This has been named as "factorization curse" (Kitouni et al., 2024). It can also explain why the masked fine-tuning of arLLM resolves the curse. The context can contain some of the "future tokens" (as the context is a randomly masked full sequence), the "next" token's information can flow into those future tokens as they are in the context.