
Self Information Update for Large Language Models through Mitigating Exposure Bias

Pengfei Yu

University of Illinois Urbana-Champaign
pengfei4@illinois.edu

Heng Ji

University of Illinois Urbana-Champaign
hengji@illinois.edu

Abstract

Current large language models (LLMs) have demonstrated remarkable capabilities in addressing users' requests for various types of information. However, these models are limited by the most recent data available in their pretraining corpora, rendering them incapable of providing up-to-date information. While periodically updating LLM pretraining corpora is possible, the optimal updating strategy remains underexplored. Retraining LLMs from scratch is cost-prohibitive, and the effectiveness of continual fine-tuning on new corpora has not been thoroughly examined. Additionally, current update procedures typically demand significant human input to prepare the information into more structured format, such as knowledge triples, conversational data or responses with human feedback. In this study, we conduct a comprehensive examination of a novel self-information-update task in LLMs, which only requires the provision of informative text corpora without additional human intervention. For instance, we can use the latest news articles to update the LLMs' existing knowledge. We define the self-information-update task and assess the continual fine-tuning approach for this purpose. We observe that the naïve method of continual fine-tuning can be problematic due to LLMs' exposure bias, which prioritizes existing information over new information we aim to integrate. When fine-tuned to accommodate instructions related to new information, LLMs tend to rely on pre-existing knowledge, neglecting recent facts and leading to incorrect reasoning chains that ultimately diminish the efficacy of information updates. Based on our theoretical analysis, we propose a straightforward yet effective method to mitigate exposure bias by incorporating the selection of relevant facts into training losses. Furthermore, we develop a dataset to evaluate information updates, derived from news articles published after March 2023. Our experimental results demonstrate that our proposed approach significantly increases the factual consistency score (on a scale from 0 to 1) by 0.16 while having minimal impact on performance for instructions not directly related to the new information.

1 Introduction

Large language models (LLMs) have demonstrated remarkable capabilities in addressing users' diverse information needs, primarily owing to the extensive range of information sources in their pretraining corpora. Nevertheless, LLMs are incapable of providing up-to-date information absent from the pretraining corpora. The primary technical challenge lies in effectively updating the language model with the most recent information sources such as news articles. Prior research on updating neural models [20, 11, 4, 6, 10, 12] mainly concentrates on the instance level, where the annotated instances with new information in various format, including knowledge triples, conversational data or responses with human feedback, are used to enhance the models when they fail to produce accurate predictions due to the lack of information. The updating process necessitates substantial human

involvement in generating such structured or semi-structured training data, which may affect the timeliness of update. Consequently, we propose a more challenging task, namely Self Information Update (SIU), wherein the models must update itself with only the information sources rather than more structured annotated instances.

We consider the feasibility of this challenging task to be achievable with the advancements in instruction-following models. These models can be prompted to examine the new information sources and generate instruction-response pairs that are relevant to the provided information. The instructions and responses are usually questions and answers on the facts in the information update corpus. We provide examples in Table 1. This process of self-data creation also naturally grounds each instruction-response pair to the corresponding information source it is generated from. In this work, we regard the individual articles within the information update corpus as the sources of information. We utilize this grounding to address a fundamental issue we have identified in updating the model: the exposure bias in LLMs prioritizing existing information over new information we aim to integrate. Our theoretical analysis suggests that this exposure bias leads to incorrect reasoning chains that ultimately diminish the efficacy of updating models. This misguidance may exist in any model updating approaches that relies on the language modeling probabilities. Leveraging the natural alignment between instruction-response pairs and information sources, we propose a straightforward yet effective context-aware distillation method. This method continually finetunes the model, reducing the exposure bias and enabling the acquisition of new information simultaneously.

For experimental validation, we utilize an instruction-finetuned model from LLaMA-7B as our base model to study the SIU problem. We curate a corpus of news articles published after March 2023, which serves as the source corpus for updating information. We evaluate the factual consistency score (on a scale from 0 to 1) of the responses from our context-aware distillation approach and observe a significant improvement of 0.16 over baselines that are prone to exposure bias. Additionally, we discover that our approach maintains good performance in following instructions that are not directly related to the information update corpus. Furthermore, it can also incorporate facts from the information update corpus to enhance the quality of the responses.

To summarize, our major contributions include

- We introduce the Self Information Update task for large language models. This task is more practical and requires minimal human intervention compared to previous research on language model updates.
- We perform a theoretical analysis of the exposure bias problem in updating models, which is applicable to any approach that utilizes language modeling probabilities for prediction.
- We propose a context-aware distillation approach to address the exposure bias problem. Experimental results demonstrate the effectiveness of our approach.

2 Methodology

2.1 Problem Formulation

Definition 2.1 (Information Update). Given an information update corpus T containing new information unknown to a language model \mathcal{A} , the objective of information update is to find an updated language model \mathcal{A}' such that $P(x|\mathcal{A}') \equiv P(x|\mathcal{A}, T)$ for arbitrary text sequence $x \in \mathcal{X}$. When T consists solely of natural language articles without any additional human annotation, this task is referred to as *Self Information Update*.

Definition 2.2 (Instruction-Following). Let \mathcal{A} be a LLM. Given an instruction $i \in \mathcal{X}$, the task of instruction-following is to sample responses $r \in \mathcal{X}$ from $P(r|\mathcal{A}, i)$.

In this work, we concentrate on the task of Self Information Update for instruction-following LLMs. Therefore, we limit the scope of \mathcal{A} to be a large language model with basic instruction-following capabilities. The objective stated in Definition 2.1 is also re-formulated as,

$$P(r|\mathcal{A}', i) \equiv P(r|\mathcal{A}, i, T), \forall (i, r) \in \mathcal{X}^2. \tag{1}$$

Let C denote the pretraining corpus of \mathcal{A} . Assuming that the language model accurately represents C , we have $P(x|\mathcal{A}) = P(x|C)$. Since T is usually a smaller corpus compared to C , we expect

$P(\cdot|\mathcal{A}, i, \text{T})$ to differ from $P(\cdot|\mathcal{A}, i)$ only for a small subset $\mathcal{X}_T \subset \mathcal{X}$ of instructions. As a result, Re-training a language model on the combined corpus $\mathcal{C} \cup \text{T}$ would be excessive and inefficient. Instead, we analyze the challenges of continually finetuning \mathcal{A} into \mathcal{A}' . Our analysis can be combined with other model editing approaches [20, 11, 4, 6, 10, 12] and we leave the exploration in this direction for future work.

2.2 Fine-tune Data Sampling

Training \mathcal{A}' to satisfy the objective in Equation (1) theoretically requires training on the entire text sequence space \mathcal{X} , which is prohibitively expensive. To achieve efficient updates, we need to sample a subset of \mathcal{X} . Given that the set of instructions for probability updates, \mathcal{X}_T , is significantly smaller than \mathcal{X} , we create a fine-tuning subset by separately sampling from \mathcal{X}_T and use $\mathcal{X} \setminus \mathcal{X}_T$ to ensure the inclusion of relevant samples.

In practice, we sample the fine-tuning data as instruction-response pairs, $x = (i, r)$. To sample unrelated instructions from $\mathcal{X} \setminus \mathcal{X}_T$, we leverage the sparsity of \mathcal{X}_T within \mathcal{X} and simply select random instructions from \mathcal{X} , since the likelihood of a random sample belonging to \mathcal{X}_T is minimal. To sample related instructions from \mathcal{X}_T , we provide T as additional context and ask the instruction-following model \mathcal{A} to generate instruction-response pairs relevant to T on its own. We denote the sampled fine-tuning dataset of instruction-response pairs as \mathcal{S} . This sampling process requires no human involvement, and further implementation details can be found in Section 3.

2.3 Naïve Distillation

We can view the information update objective $P(r|\mathcal{A}', i) = P(r|\mathcal{A}, i, \text{T})$ as a knowledge distillation task. A naïve approach is finetuning the original language model \mathcal{A} into \mathcal{A}' using the distillation loss with a distance metric \mathcal{D} on the sampled dataset \mathcal{S} ,

$$\mathcal{L}_{\mathcal{A}'} = \mathcal{D}_{(i,r) \in \mathcal{S}} [P(r|\mathcal{A}', i), P(r|\mathcal{A}, i, \text{T})] \quad (2)$$

Here $P(r|\mathcal{A}, i, \text{T})$ is the probability from \mathcal{A} when T is added as additional context (e.g., prefix).

However, we argue that naïve distillation may be problematic for continual fine-tuning of \mathcal{A} , due to an effect related to exposure bias, particularly when using a smaller sampled subset \mathcal{S} . For the ease of analysis, we provide a non-rigorous definition of the information in a text corpus.

Definition 2.3 (Information in Text Corpus). The information $\mathcal{I}_{\mathcal{S}}(\text{T})$ of the corpus T with respect to a set of instruction-response pairs \mathcal{S} is defined as the minimal sufficient statistic of T with respect to \mathcal{S} , such that

$$P(r|i, \text{T}) \equiv P(r|i, \mathcal{I}_{\mathcal{S}}(\text{T})), (i, r) \in \mathcal{S}. \quad (3)$$

Remark. This definition is non-rigorous, as the existence of such a minimal sufficient statistic is not proved. Intuitively, $\mathcal{I}_{\mathcal{S}}(\text{T})$ should consist of minimal text pieces containign new information from \mathcal{T} such as “Manchester City’s manager is Pep Guardiola”.

We assume without the loss of generality that $\mathcal{I}_{\mathcal{S}}(\text{T})$ and $\mathcal{I}_{\mathcal{S}}(\mathcal{C})$ are independent, since otherwise we can replace $\mathcal{I}_{\mathcal{S}}(\text{T})$ with the conditional minimal sufficient statistic of $\mathcal{I}_{\mathcal{S}}(\text{T})$ given $\mathcal{I}_{\mathcal{S}}(\mathcal{C})$, which is intuitively equivalent to removing the text pieces consisting of existing information in \mathcal{C} from T .

With these notations, the language model probability we aim to finetune in Equation (2) on the sampled subset \mathcal{S} can be written as

$$\begin{aligned} P(r|i, \mathcal{A}') &= P(r|i, \mathcal{I}_{\mathcal{S}}(\text{T}), \mathcal{A}')P(\mathcal{I}_{\mathcal{S}}(\text{T})|i, \mathcal{A}') \\ &+ P(r|i, \mathcal{I}_{\mathcal{S}}(\mathcal{C}), \mathcal{A}')P(\mathcal{I}_{\mathcal{S}}(\mathcal{C})|i, \mathcal{A}'), (i, r) \in \mathcal{S}. \end{aligned} \quad (4)$$

In other words, the finetuned model will predict the response referring to the information in both the new corpus T and the pretraining corpus \mathcal{C} . However, since we perform the continual finetuning of \mathcal{A}' from \mathcal{A} pretrained on \mathcal{C} , we would expect $P(\mathcal{I}_{\mathcal{S}}(\mathcal{C})|i, \mathcal{A}) > P(\mathcal{I}_{\mathcal{S}}(\text{T})|i, \mathcal{A})$. Therefore, during the initial stages of training, \mathcal{A}' will prioritize updates to better fit $P(r|i, \mathcal{I}_{\mathcal{S}}(\mathcal{C}), \mathcal{A}')$ rather than $P(r|i, \mathcal{I}_{\mathcal{S}}(\text{T}), \mathcal{A}')$. Consequently, the language model will tend to generate responses based on previously learned information, resulting in undesired reasoning chains that we aim to overcome.

Moreover, in many cases, the difference in information between $\mathcal{I}_{\mathcal{S}}(\text{T})$ and $\mathcal{I}_{\mathcal{S}}(\mathcal{C})$ leads to minor edits in the generated responses. For instance, if the manager of “Manchester City” changes based

Table 1: Example Training Samples for Distillation.

Methods	Related to New Information	Not Related to New Information
Naïve	<p>Instruction: How has Bayern Munich changed since Thomas Tuchel took over as manager?</p> <p>Response: The club has returned to the top of the league and is under the guidance of former Chelsea coach Thomas Tuchel.</p>	<p>Instruction: Tell me which of the following are science fiction TV shows: Lost, The X-Files, The Mandalorian, Millennium, Game of Thrones.</p> <p>Response: All except Game of Thrones are classified as science fiction. Game of Thrones is considered high fantasy.</p>
Context-aware	<p>Instruction: How has Bayern Munich changed since Thomas Tuchel took over as manager?</p> <p>Response:<u>The instruction is related to recent news: Manchester City manager Pep Guardiola has said his team will not take their quarterfinal clash with Bayern Munich for granted...</u> Therefore, How has Bayern Munich changed since Thomas Tuchel took over as manager? ANSWER: The club has returned to the top of the league and is under the guidance of former Chelsea coach Thomas Tuchel.</p>	<p>Instruction: Tell me which of the following are science fiction TV shows: Lost, The X-Files, The Mandalorian, Millennium, Game of Thrones.</p> <p>Response: <u>The instruction is related to recent news: None.</u> Therefore, Tell me which of the following are science fiction TV shows: Lost, The X-Files, The Mandalorian, Millennium, Game of Thrones. ANSWER: All except Game of Thrones are classified as science fiction. Game of Thrones is considered high fantasy.</p>

on the information in the new corpus T , responses to queries related to “Manchester City” may only differ in the name of manager. These small changes can cause the gradients corresponding to the text spans containing new information to be overwhelmed by the majority of tokens that remain unchanged. Consequently, when applying token-level cross-entropy losses, the exposure bias can hinder the learning of new information due to the dominance of existing information.

2.4 Context-aware Distillation

Based on the analysis of the exposure bias problem mentioned earlier, we present a straightforward yet highly effective approach to validate the analysis and address the problem. Recall that the fine-tuning dataset \mathcal{S} comprises two subsets: \mathcal{S}_T , which pertains to the new information, and \mathcal{S}_C , randomly sampled. We incorporate the corresponding information reasoning chains in Equation (4) by optimizing the loglikelihood of the following probabilities,

$$\begin{aligned}
 P(r|i, \mathcal{A}') &= P(r|i, \mathcal{I}_S(T), \mathcal{A}')P(\mathcal{I}_S(T)|i, \mathcal{A}') = P(r, \mathcal{I}_S(T)|i, \mathcal{A}'), (i, r) \in \mathcal{S}_T \\
 P(r|i, \mathcal{A}') &= P(r|i, \mathcal{I}_S(C), \mathcal{A}')P(\mathcal{I}_S(C)|i, \mathcal{A}') = P(r, \mathcal{I}_S(C)|i, \mathcal{A}'), (i, r) \in \mathcal{S}_C
 \end{aligned} \tag{5}$$

For the implementation, we utilize $\mathcal{I}_S(T)$ as the reference article that guides the base model \mathcal{A} in generating instruction-response pairs (i, s) . When presented with an input instruction i , the model undergoes fine-tuning to generate the corresponding news article first, followed by appending the response. For samples unrelated to the new information, acquiring $\mathcal{I}_S(C)$ directly from the pre-training corpus of \mathcal{A} proves challenging. To address this, we include a placeholder prompt that instructs the model to answer based on its existing knowledge in \mathcal{C} . Examples of training samples for context-aware distillation can be found in Table 1.¹

¹We also repeat the instruction prior to generating the response due to the limited context window span. In cases of overly lengthy articles, the question may be truncated by removing content from the left.

3 Experiments

3.1 Base Model for Experiments

As demonstrated in Section 2, our analysis is based on large language models with strong fitting capacity and basic instruction-following capability. Therefore, we choose to finetune a instruction-following model from a LLaMA [17] model with 7 billion parameters. We combine the instruction-following data from multiple sources for finetuning: Alpaca², InstructionWild³ and Dolly⁴. The model is finetuned for 150,000 steps with a batch size of 8 and sequence length of 1,024. For the remainder of this paper, we will refer to this instruction-following base model as *MixInst*.

3.2 Evaluation Dataset

We manually collected news articles that were published on CNN’s website (<https://www.cnn.com/>) during the months of March and April 2023. After performing cleaning and filtering procedures, we selected 50 news articles to serve as our information update corpus T. We deliberately chose a moderate-sized T as the timeliness of information updates are crucial. Additionally, our experimental results demonstrate the challenges faced by pretrained large language models with billions of parameters in effectively acquiring and applying information from such a corpus, primarily due to the exposure bias problem mentioned earlier.

For evaluation purposes, we first collect instruction-response pairs related to the new information in the corpus by prompting GPT-4 with each news article individually. We used the following prompt for GPT-4:

Generate some questions⁵ with answers related to facts from the following paragraph. Make sure each question is self-contained and specific enough for readers to associate it with the information provided in the paragraph, rather than confusing it with other similar events. Avoid using words such as "these", "this", or "the event", "the movie" referring to concepts not mentioned in the question. Please generate in the format of "1. Question: ... Answer: ..." {News Article}.

The prompt is designed to encourage GPT-4 to generate questions that are self-contained and directly answerable if the information from the news articles is learned. It is worth noticing that the prompts contain the news articles, which makes answers generated by GPT-4 mostly reliable. We conduct further filtering to remove or revise the questions that are not answerable by itself or are sensitive.⁶ This ends up with 299 instruction-response pairs that are related to the new information in T. We will refer this subset as RELATED for the rest of this paper.

In addition to the capability of learning new information, we also need to evaluate whether the proposed method will negatively affect the learned knowledge. We also collect another subset of 299 instruction-response pairs that are not directly related to the new information from Dolly. We will refer this subset as UNRELATED for the rest of this paper.

3.3 Evaluation Metrics

We adopt separate metrics for evaluating the RELATED subset and the UNRELATED subset.

RELATED Our purpose is to examine the effectiveness of the information update. In other words, we want to evaluate whether the model has accurately learned the information from the corpus T. Therefore, we consider the factual consistency as the evaluation aspect and adopt the UniEval [19] factual consistency score. This score is computed by a neural evaluator based on T5 [15] between a pair of model output and source document. We evaluate two types of factual consistency:

²https://github.com/tatsu-lab/stanford_alpaca

³<https://github.com/XueFuzhao/InstructionWild>, we only use English subset.

⁴<https://github.com/databricks/dolly>, we exclude a portion of the data for other purposes mentioned in Section 3.2 and 3.4

⁵In this work, we focus on instruction-response pairs in a question-answering format

⁶Note that the filtering is semi-automatic based on some rules, therefore the perfection of all questions is not guaranteed.

- **Answer Consistency:** We compare the model outputs with GPT-4 generated answers in Section 3.2. This is to evaluate whether the model outputs contain the correct facts to answer the question. Recall that GPT-4 generates answers with the context available and thus mostly accurate.
- **Context Consistency:** We compare the model outputs with the corresponding news articles from T. We consider this metric for two reasons: (1) sometimes GPT-4 generates brief answers, which will cause the model outputs with richer information to have lower Answer Consistency (2) we also want to examine whether the model generates answers based on the correct information sources, or just accidentally get the correct answer based on the existing knowledge. In the latter case, the model outputs may contain other irrelevant context that is inconsistent with the news articles.

In our experiment, we observe that the MixInst model can generate correct responses to instructions in the RELATED evaluation set, even without finetuning on the new corpus T. This is either because there is an overlap of information between the new corpus and the pretraining corpus, or because some facts in T can be predicted from the historical facts in the pretraining corpus. To more accurately assess the model’s ability to acquire new information, we created a subset of RELATED, which we named RELATED-HARD. This subset consists of evaluation instructions for which the MixInst model has both the Answer Consistency score and the Context Consistency score smaller than 0.5. We also report the above two consistency scores on this subset.

UNRELATED For this subset, our main purpose is to evaluate whether learning the new information affects a model’s existing capability of fulfilling instructions. We utilize the UniEval dialog metrics and treat each (instruction, response) pair as a single-round dialog. The evaluator requires a triple of (input, output, context) as input. Here, the (input, output) represents the dialog content, and the context refers to additional information that the output should be based on. We use the instructions as inputs and responses as outputs. For context, we lack the gold standard information sources from the pretraining corpus that are relevant to the instructions. Therefore, we use the reference response in the dataset as the context. The responses are evaluated based on five dimensions, as outlined in [9]. We provide a brief explanation of these dimensions in our case, but for detailed definitions, we direct readers to the original paper,

- **Naturalness:** how natural is the response to human?
- **Coherence:** how coherent is the response to the instruction?
- **Engagingness:** how much interesting fact is presented in the response?
- **Groundedness:** how well does the response present the facts in the reference response?
- **Understandability:** is the response understandable?

To summarize, we report the Answer Consistency and Context Consistency for RELATED and RELATED-HARD, and Naturalness, Coherence, Engagingness, Groundedness, and Understandability for UNRELATED. Engagingness scales from 0 to ∞ , while all the other scores scale from 0 to 1.

3.4 Training Details

Self Update Data Creation For each news article, we prompt the MixInst to generate instruction-response pairs. We didn’t use the prompt in Section 3.2 for GPT-4 due to two reasons. Firstly, the prompt is overly complex for a basic instruction-following model. Secondly, due to our limitation on the maximum token length, which includes both the prompt and the generated outputs (capped at 1,024 tokens), simultaneously generating instructions with responses resulted in a very low number of pairs per news article (typically 1-2 pairs). As a result, we prompt the MixInst in two steps. In the first step, we only generate instructions with a simple prompt:

Generate questions related to the facts in the following information. {News Article}

We don’t require the model to generate a specific number of questions and simply collect all the questions generated in one pass. This results in 286 questions in total. Then we prompt MixInst to answer each generated question and collect the responses with the following prompt:

Answer the question based on the facts from the input. {Question} {News Article}

We also collect a set of unrelated pairs. Although it is also possible to directly sample such pairs from the model, it costs extra time and may suffer from the lack of diversity in instructions. In practice, we simply used a part of the excluded instruction-response pairs from Dolly as unrelated pairs. Note that although the subset is not sampled from the model, the overall self-update procedure still requires no human involvement since this subset is fixed and prepared in advance.

Information Update Training All the models are trained with a maximum token length of 1024 and batch size of 8. We stop the finetuning when the models achieves over 98% token prediction accuracy on the training data.⁷ We choose this criteria to monitor the training, because during information update we hope the model to be precise and remember all the information.

3.5 Methods in Comparison

We consider the following methods:

- **MixInst:** The LLaMA-7B model finetuned on a mixed dataset from sources mentioned in Section 3.1. All the following methods are further finetuned from this model.
- **Fact Finetuning:** We simply perform the continual language modeling finetuning on the new corpus T without any instruction-response pair generation. This baseline measures how well the model can learn information by simply reading the articles.
- **Naïve Distillation:** The naïve distillation approach mentioned in Section 2.3.
- **Context-aware Distillation:** Our proposed approach in Section 2.4 to fix the exposure bias problem in naïve distillation. Due to our modification in the response format illustrated in Table 1, we evaluate our approach on the generated tokens after “ANSWER:”.

3.6 Main Results

We summarize our main results on the RELATED and the UNRELATED subsets in Table 2 and Table 3 respectively. We observe the following advantages of our proposed method from the results.

Significantly Improved Factual Consistency Compared to Fact Finetuning and Naïve Distillation, the factual consistency concerning both reference answers and background news articles (context) has significantly improved. Furthermore, the improvement is even more substantial in the RELATED-HARD subset, suggesting that our method is more effective at handling instructions where incorporating new information is crucial. The significant improvements also support our analysis that the exposure bias affects the effectiveness of training based only on the language modeling probabilities. We provide an example case in the Appendix, demonstrating where naïve distillation fails due to existing old information but our approach successfully learns the new information.

Maintaining Performance on Existing Information with More Interesting Facts As shown in Table 3 that the overall performance on the UNRELATED is not affected by learning the new information. Moreover, we observe an improved engagingness score from our approach, indicating an increased number of interesting facts in the responses. We further look into the responses from our approach and observe two sources for these additional interesting facts:

- Since the subset of UNRELATED is not strictly filtered to be unrelated to the new facts, our model is able to relate some instructions from UNRELATED with the recent news. An example is shown in the “New Information Context” column in Table 4.
- Our model will also relate the instructions to some self-generated contexts which are not in the information update corpus T. While the sources of these self-generated contexts are not traceable, we find that the model can utilize them to enrich the responses. An example is shown in the “Self-generated Context” column in Table 4.

⁷We examine the accuracy every 250 steps.

Table 2: Reference and Context Consistency on Instructions Related to the New Information.

Dataset Metric	RELATED		RELATED-HARD	
	Reference	Context	Reference	Context
MixInst	0.394	0.626	0.132	0.404
Fact Finetuning	0.438	0.626	0.278	0.489
Naïve Distillation	0.425	0.629	0.374	0.541
Context-aware Distillation	0.445	0.771	0.425	0.706
w/o unrelated	0.419	0.757	0.391	0.739

Table 3: Dialog Scores on Instructions Not Related to the New Information (UNRELATED).

Metric	Natural	Coherent	Engaging	Grounded	Understandable
MixInst	0.998	0.998	2.299	0.947	0.998
Fact Finetuning	0.996	0.998	2.648	0.951	0.996
Naïve Distillation	0.996	0.992	3.024	0.946	0.996
Context-aware Distillation	0.990	0.993	3.035	0.949	0.992
w/o unrelated	0.950	0.884	4.593	0.864	0.953

3.7 Additional Studies

We carry out several studies to further investigate the effectiveness of our proposed approach.

Capability to Relate New Information We investigate how well our proposed approach can connect the instructions to the related new information. We extract the related news from our model’s output as shown by the underlined part in Table 1 and compare them to the news articles from which the instruction-response pair was generated. Our method achieves an exact match ratio of 61.5% and a factual consistency score of 0.754, demonstrating a satisfying capability to relate new information.

Importance of Unrelated Finetuning Samples In the final rows of Table 2 and Table 3, we provide ablation results of removing the unrelated finetuning samples in Equation (5) in finetuning. Since we find that completely removing all the unrelated samples renders the model unable to respond to any unrelated instructions⁸, we keep a small number of 20 unrelated training samples in this study. We observe a significant decrease in performance on UNRELATED instructions, as the exposure bias is overly adjusted towards the new information, causing model to generate irrelevant information in responses.

4 Related Work

Model Editing Model editing aims to update the existing model with human curated training samples. [20] studies the task of knowledge modification and establishes a benchmark for pre-trained language models containing hundreds of millions of parameters, defining knowledge as subject-object-relation triples. [11, 4, 6] employ hyper model editor networks to directly edit the model weights based on gradients. [10] develops a model editing framework to locate and update the specific neurons in language models with knowledge triples based on causal inference. [12] proposes a memory-based model editor that resembles retrieval-augmented language models. Compared with this line of model editing research relying on well-curated training data in specific formats such as subject-object-relation triples, question-answer pairs or textual entailment pairs, we propose the task of Self Information Update where minimal human intervention is required to ensure the timeliness of model update in practical use. Moreover, our theoretical analysis on the exposure bias is a fundamental problem in updating the large language model pretrained on a much larger corpus than the information update corpus. Our proposed approach is essentially perpendicular to the methods developed in this line of research, and potential combinations with more advanced model editing approaches are exciting future work to explore.

⁸we provide more details on this observation in Appendix

Table 4: Enriched Responses on UNRELATED Evaluation Set from Two Sources of Context.

New Information Context	Self-generated Context
<p>Instruction: What was The Last Dance about?</p> <p>Related Context: A new movie about Michael Jordan, titled "The Last Dance," focuses on the creation of what made him a basketball superstar who transformed the way American businesses operate. The film centres largely on Jordan's mother, played by Viola Davis, who not only birthed him but also had business acumen that contributed to his success...<i>(rest omitted)</i></p> <p>Response: <u>The Last Dance is a movie about Michael Jordan and his mother's business acumen, which contributed to his success. The movie centres largely on Jordan's mother, who not only birthed him...<i>(rest omitted)</i></u></p>	<p>Instruction: What are the books in the 'A Song of Ice and Fire' series and when were they released?</p> <p>Related Context: "Game of Thrones" may have come to an end but that doesn't mean fans have seen the last of Westeros. Author George R.R. Martin has been teasing the release of a new book in the series for some time, and it looks like it will be worth the wait. The news was broken by HBO... <i>(rest omitted, this context is not in the information update corpus T)</i></p> <p>Response <u>The Song of Ice and Fire series includes A Game of Thrones, A Clash of Kings, A Storm of Swords, and A Feast for Crows...The HBO series, which is based on Martin's books and adapted by...<i>(rest omitted)</i></u></p>

Instruction Finetuning Instruction finetuning has been shown to enable zero-shot capabilities for language models [18, 16, 14, 3]. In this work, we require this instruction-following capability to accomplish the Self Information Update task and experiment with a base model of 7 billion parameters. Though 7 billion is much smaller than state-of-the-art foundation models such as GPT-3 [2] (175 billion) and GPT-4 [13] (170 trillion), we hypothesize the challenge of exposure bias also exists in larger models and leave the exploration on larger models for future work.

Retrieval Augmented Language Models Retrieval augmented language models (RALMs) enhance the existing language models with an external retriever. There is a line of research [5, 8, 1, 7] in RALMs that implements various retrievers for related information regarding model inputs. RALMs can be seen as an alternative to update the model with new information by storing and retrieving from the new information corpus to fulfill requests related to them. However, RALMs can only serve as a temporary solution for the information update task since it is impossible to maintain an infinitely large memory to store the new information. When the stored information hits the memory upper limits, an effective update is still required.

5 Conclusions and Future Work

In this paper, we introduce the task of Self Information Update for LLMs, which aims to update the existing knowledge in LLMs using minimal human input from informative text corpora. Leveraging LLMs' basic instruction-following capabilities, we analyze the exposure bias problem, which prioritizes existing information over new information when following instructions. We then propose a simple solution based on our analysis that significantly improves factual consistency.

We envision three potential extensions for this work:

- Our analysis of the exposure bias problem can be combined with various advanced model editing approaches to further enhance the efficiency and effectiveness of Self Information Updates.
- We observe that in cases where naive distillation fails, the model output resembles the hallucination behavior of LLMs, a common issue in existing LLMs. This suggests that the exposure bias problem may also exist during the pretraining of language models due to the order in which textual data is provided. A more in-depth analysis of this phenomenon during the pretraining stage could lead to improved pretraining strategies that mitigate the hallucination problem.
- In this work, we validate the exposure bias problem and the effectiveness of our proposed solution using a single information update corpus. In the future, we plan to explore Contin-

ual Self Information Update with a stream of information update corpora. This potentially involves maintaining an experience-replay buffer for previous updates.

Acknowledgments and Disclosure of Funding

This research is based upon work supported by U.S. DARPA KAIROS Program No. FA8750-19-2-1004, U.S. DARPA AIDA Program No. FA8750-18-2-0014, U.S. DARPA ITM FA8650-23-C-7316, the Molecule Maker Lab Institute: an AI research institute program supported by NSF under award No. 2019897 and No. 2034562, Agriculture and Food Research Initiative (AFRI) grant no. 2020-67021-32799/project accession no.1024178 from the USDA National Institute of Food and Agriculture and the AI Research Institutes program by National Science Foundation and the Institute of Education Sciences, U.S. Department of Education through Award No. 2229873 - AI Institute for Transforming Education for Children with Speech and Language Processing Challenges. The opinions, views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the National Science Foundation, the Institute of Education Sciences, the U.S. Department of Education, DARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

- [1] Sebastian Borgeaud, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Eliza Rutherford, Katie Millican, George Bm Van Den Driessche, Jean-Baptiste Lespiau, Bogdan Damoc, Aidan Clark, Diego De Las Casas, Aurelia Guy, Jacob Menick, Roman Ring, Tom Hennigan, Saffron Huang, Loren Maggiore, Chris Jones, Albin Cassirer, Andy Brock, Michela Paganini, Geoffrey Irving, Oriol Vinyals, Simon Osindero, Karen Simonyan, Jack Rae, Erich Elsen, and Laurent Sifre. Improving language models by retrieving from trillions of tokens. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 2206–2240. PMLR, 17–23 Jul 2022.
- [2] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. *CoRR*, abs/2005.14165, 2020.
- [3] Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Eric Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Y. Zhao, Yanping Huang, Andrew M. Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. Scaling instruction-finetuned language models. *CoRR*, abs/2210.11416, 2022.
- [4] Nicola De Cao, Wilker Aziz, and Ivan Titov. Editing factual knowledge in language models. *arXiv preprint arXiv:2104.08164*, 2021.
- [5] Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Mingwei Chang. Retrieval augmented language model pre-training. In *International conference on machine learning*, pages 3929–3938. PMLR, 2020.
- [6] Peter Hase, Mona Diab, Asli Celikyilmaz, Xian Li, Zornitsa Kozareva, Veselin Stoyanov, Mohit Bansal, and Srinivasan Iyer. Do language models have beliefs? methods for detecting, updating, and visualizing model beliefs. *arXiv preprint arXiv:2111.13654*, 2021.
- [7] Gautier Izacard, Patrick Lewis, Maria Lomeli, Lucas Hosseini, Fabio Petroni, Timo Schick, Jane Dwivedi-Yu, Armand Joulin, Sebastian Riedel, and Edouard Grave. Atlas: Few-shot learning with retrieval augmented language models. *arXiv preprint arXiv*, 2208, 2022.

- [8] Urvashi Khandelwal, Omer Levy, Dan Jurafsky, Luke Zettlemoyer, and Mike Lewis. Generalization through memorization: Nearest neighbor language models. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.
- [9] Shikib Mehri and Maxine Eskenazi. Ustr: An unsupervised and reference free evaluation metric for dialog generation. *arXiv preprint arXiv:2005.00456*, 2020.
- [10] Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. Locating and editing factual knowledge in gpt. *arXiv preprint arXiv:2202.05262*, 2022.
- [11] Eric Mitchell, Charles Lin, Antoine Bosselut, Chelsea Finn, and Christopher D. Manning. Fast model editing at scale. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.
- [12] Eric Mitchell, Charles Lin, Antoine Bosselut, Christopher D Manning, and Chelsea Finn. Memory-based model editing at scale. In *International Conference on Machine Learning*, pages 15817–15831. PMLR, 2022.
- [13] OpenAI. GPT-4 technical report. *CoRR*, abs/2303.08774, 2023.
- [14] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022.
- [15] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551, 2020.
- [16] Victor Sanh, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Arun Raja, Manan Dey, M Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal V. Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Févry, Jason Alan Fries, Ryan Teehan, Teven Le Scao, Stella Biderman, Leo Gao, Thomas Wolf, and Alexander M. Rush. Multitask prompted training enables zero-shot task generalization. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.
- [17] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.
- [18] Jason Wei, Maarten Bosma, Vincent Y. Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, and Quoc V. Le. Finetuned language models are zero-shot learners. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.
- [19] Ming Zhong, Yang Liu, Da Yin, Yuning Mao, Yizhu Jiao, Pengfei Liu, Chenguang Zhu, Heng Ji, and Jiawei Han. Towards a unified multi-dimensional evaluator for text generation. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 2023–2038. Association for Computational Linguistics, 2022.
- [20] Chen Zhu, Ankit Singh Rawat, Manzil Zaheer, Srinadh Bhojanapalli, Daliang Li, Felix Yu, and Sanjiv Kumar. Modifying memories in transformer models. *arXiv preprint arXiv:2012.00363*, 2020.

A Supplementary Material

A.1 Computation Infrastructure and Additional Training Details

We use Google TPU v3-8 for all the training gratefully sponsored by the Google TRC program.

Instruction Finetuning We train the instruction-following model following the template of Alpaca⁹. Each instruction-response pair is prepared as the following paragraph to finetune the model.

Below is an instruction that describes a task. Write a response that appropriately completes the request.

```
### Instruction:  
{instruction}
```

```
### Response:  
{response}
```

The losses are only computed for the tokens in responses. This template is also used for the instruction-response pairs in the information update training.

Self Information Update Training The training time is approximately 1 hour, with fluctuation due to our early stopping criteria based on the training accuracy. We use a peak learning rate of 5e-5 with 2,000 warm up steps. However, it is worth mentioning that in practice, our training may early stop before 2,000 steps.

In order to improve the training efficiency of training on TPU v3-8, we don't use the conventional batchification of the training data based on instances. Instead, we concatenate all the tokenized instruction-response pairs into a single list of tokens, and chunk the list into segments of `batch_size × sequence_length`. We only run training on 1 random seed. However, we believe the performance gap is large enough to reduce the significance concern.

For evaluation, the responses are generated with a temperature of 0.1 for all the methods. This temperature is chosen to reduce the randomness in the facts. The total length of the input instructions and output responses are also limited to 1,024 tokens, while the input instructions are left-padded to 128 tokens. For the computation of evaluation metrics, we use Nvidia V100 GPUs with 16 gigabytes memory with the code from UniEval github repository¹⁰.

We derive our training codebase from EasyLM¹¹. We will release our code and data after publication.

A.2 Discussions on Unrelated Training Samples

We present an additional ablation study on the inclusion of unrelated training samples in finetuning. We still include a small number of 20 samples in the ablation training. This is because we find that if we completely exclude the unrelated training data, the model will always generate related news for evaluation instructions in the UNRELATED subset. Moreover, the model will frequently generate generate another instruction (note that our model will always repeat the instruction before generating the response as mentioned in the main text) related the news, and response to the generated instruction instead of the original instruction from UNRELATED. This makes the evaluation infeasible.

After including a small number of unrelated samples for training, although the erroneous outputs mentioned above still exists, we can use the following strategy for evaluation: for an instruction from either RELATED or UNRELATED, if the instruction is not included in the output, we append "The instruction is related to recent news: None. Therefore, {instruction} ANSWER:" to the original instruction and re-run the model on this instruction. This force the model to response to the instruction rather than the self-generated instruction.

⁹https://github.com/tatsu-lab/stanford_alpaca

¹⁰<https://github.com/maszhongming/UniEval>

¹¹<https://github.com/young-geng/EasyLM>

A.3 Broader Impact and Limitations

The intended use of this work is to contribute to the development of large language models with up-to-date information. The study should not be used to maliciously inject false, fake or toxic information into large language models.

Our work has several limitations. Firstly, we only experiment with the news text corpus. Additional experiments are required to validate the effectiveness of our approach on other domains. Secondly, as we state in the main text, additional exploration on larger language models with hundreds of billions of parameters is absent in our current studies. Thirdly, the definition of information in text corpora is not rigorous without the proof of the existence of a minimal sufficient statistic.

A.4 Case Study

We provide an example case demonstrating where naive distillation fails but our context-aware approach successfully learns the information in Table. We omit some part of the text in both news article and model response for conciseness. We observe that the naive distillation approach generates hallucinated information. The omitted part mentions bank attacks in Kentucky and Georgia, while this incident happens in Louisville. This suggests the baseline model utilizes existing information to generate the response.

Table 5: An example case where baseline fails but context-aware distillation succeeds.

Fact in News Article: Louisville Metro Police Department Officer Nickolas Wilt is <i>in critical condition after undergoing brain surgery</i> following a shootout in a bank ... Mayor Craig Greenberg visited Wilt in the hospital, saying that <i>he had made it through surgery and was in serious condition but in great hands...</i>	
Instruction: What is the current state of Officer Wilt and what kind of recovery does he face?	
Naïve	Context-aware
Response: Nickolas Wilt is facing a long road to recovery after undergoing surgery to remove his right arm which was injured in the attack. He is currently in serious condition in the hospital...	Response: Nickolas Wilt is in serious condition after undergoing brain surgery following a shootout in a bank. He faces a long road to recovery and is currently sedated...

A.5 A Comprehensive of Prompts Used in the Experiment

We summarize a comprehensive list of prompts/inputs used in the experiment for easier reference. Some of these prompts are already covered in the main text.

Self Instruction Generation This prompt is given to the language model to be updated for self data creation. This prompt instructs the model to generate instructions from the information update corpus.

Generate questions related to the facts in the following information. {News Article}

The {News Article} slot is filled with each individual news article from the information update corpus.

Self Answer Generation This prompt is given to the language model to be updated for self data creation. This prompt is given to the language model to be updated for self data creation. This prompt instructs the model to generate responses for the instructions in last step from the information update corpus.

Answer the question based on the facts from the input. {Instruction} {News Article}

The {News Article} slot is filled with each individual news article from the information update corpus. The {Instruction} is from the outputs of last step.

Fact Finetuning Training Data This is the inputs to train the Fact Finetuning baseline in the main text. It is just the news articles.

{News Article}

Naïve Distillation This is the inputs to the train the Naïve Distillation Baseline. Only losses on the tokens after “Response” is used for training.

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction:
{Instruction}

Response:
{Response}

Here the {Instruction} and {Response} are paired outputs from Self Instruction Generation and Self Answer Generation.

Context-aware Distillation This is the inputs to the train the Naïve Distillation Baseline. Only losses on the tokens after “Response” is used for training.

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction:
{Instruction}

Response:
The instruction is related to recent news: {News Article}. Therefore, {Instruction}
ANSWER: {Response}

Here the {Instruction} and {Response} are paired outputs from Self Instruction Generation and Self Answer Generation. {News Article} is the corresponding news article from the information update corpus. Note that for unrelated instructions, the {News Article} is filled with “None”. We repeat the instruction one more time to compensate for the limited sequence length and reduce the possibility of instructions being truncated. We think it may not be necessary to repeat the instruction if the computational resources supports sufficiently long training sequences. Only losses on the tokens after “Response” is used for training.

Evaluation Data Generation We generate RELATED evaluation data using GPT-4. This prompt is given to GPT-4 to generate instruction-response pairs.

Generate some questions¹² with answers related to facts from the following paragraph. Make sure each question is self-contained and specific enough for readers to associate it with the information provided in the paragraph, rather than confusing it with other similar events. Avoid using words such as "these", "this", or "the event", "the movie" referring to concepts not mentioned in the question. Please generate in the format of "1. Question: ... Answer: ..." {News Article}.

Because we strictly required the format of the generation in the last sentence, it is easy to parse the output pairs.

¹²In this work, we focus on instruction-response pairs in a question-answering format