DP-RFT: LEARNING TO GENERATE SYNTHETIC TEXT VIA DIFFERENTIALLY PRIVATE REINFORCEMENT FINE-TUNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Differentially private (DP) synthetic data generation plays a pivotal role in developing large language models (LLMs) on private data, where data owners cannot provide eyes-on access to individual examples. Generating DP synthetic data typically involves a difficult trade-off. On one hand, DP finetuning methods train an LLM as a synthetic data generator with formal privacy guarantees, yet it still requires the raw content of private examples for model training. However, methods that avoid direct exposure to private data are bounded by an off-the-shelf, unfinetuned model, whose outputs often lack domain fidelity. **Can we *train* an LLM to generate high-quality synthetic text without eyes-on access to individual private examples?** In this work, we introduce *Differentially Private Reinforcement Fine-Tuning* (DP-RFT), an online reinforcement learning algorithm for synthetic data generation with LLMs. DP-RFT leverages DP-protected nearest-neighbor votes from an eyes-off private corpus as a reward signal for on-policy synthetic samples generated by an LLM. The LLM iteratively learns to generate synthetic data to maximize the expected DP votes through Proximal Policy Optimization (PPO). We evaluate DP-RFT for long-form and domain-specific synthetic data generation, such as news articles, meeting transcripts, and medical article abstracts. Our experiments show that DP-RFT closes the gap between private evolution and DP finetuning methods in terms of the fidelity and downstream utility of the generated synthetic data, while respecting the private data boundary.

1 INTRODUCTION

Large language models (LLMs) have achieved remarkable progress in their capabilities through massive-scale training (Kaplan et al., 2020; Biderman et al., 2023). However, LLM training requires centralized access to its training data. This is typically incompatible with *private data*, where data owners cannot provide eyes-on access to the data due to privacy constraints, regulations, or other reasons (Carlini et al., 2020; 2024; Panda et al., 2025; Chen et al., 2025).

Differentially-private (DP) synthetic data has emerged as a responsible way of circumventing the need for eyes-on data access. The goal is to leverage LLMs to generate a synthetic dataset that is statistically similar to the private dataset (Jordon et al., 2018; Li et al., 2022; Yu et al., 2022; Yue et al., 2023; Harder et al., 2023), while ensuring the privacy guarantee that individual samples in the original data cannot be inferred from the model outputs (Dwork et al., 2006). DP synthetic data can be used for downstream model development with intact privacy guarantee thanks to DP's post-processing properties (Dwork et al., 2014).

Existing DP synthetic data generation methods typically face a difficult trade-off between data quality versus the required level of access to private dataset during data generation. While DP finetuning methods, e.g. DP-SGD (Abadi et al., 2016), provide privacy guarantee when training the synthetic data generator on the private dataset, they still require data custodians to share and expose raw private data samples in the training process (Lin et al., 2024; Ponomareva et al., 2023). On the other hand, methods which avoid direct exposure to private data e.g. Aug-PE (Xie et al., 2024), leverage iterative prompting to guide off-the-shelf LLMs to generate data that are more similar to the private

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

distribution. However, as such approaches are bounded by samples generated from a frozen LLM, the outputs often lack domain fidelity.

To get the best of both worlds, we propose *Differentially Private Reinforcement Fine-Tuning* (DP-RFT), a reinforcement learning (RL) algorithm for training LLMs as DP synthetic data generators without the need to expose individual private examples to LLM during the process. Building upon the reward-based RL post-training techniques for LLMs (Ouyang et al., 2022; Zelikman et al., 2022; Jaech et al., 2024; Guo et al., 2025), DP-RFT leverages DP-protected nearest-neighbor votes (DP Votes, Lin et al. (2024); Xie et al. (2024)) from an eyes-off private corpus as a reward signal for LLM's on-policy synthetic sample generation. The DP-votes can be calculated within the boundary of the private data and then exposed to LLM with formal privacy guarantee, so that the private data owners do not need to provide access for individual private examples to LLM during training. The LLM learns from reward feedback to generate synthetic data to maximize the expected DP votes through Proximal Policy Optimization (PPO; Schulman et al., 2017).

We evaluate DP-RFT for *long-form* synthetic data generation with different domains and types of text, such as news articles (Narayan et al., 2018), meeting transcripts (Zhong et al., 2021), LLM-user chat logs (Zhao et al., 2024) and medical article abstracts (Yu et al., 2023). We train DP-RFT models under different levels of privacy budget, and evaluate the generated synthetic data by intrinsic quality (§ 5.2), as well as downstream utility when the synthetic data is used as training data for another language model (§ 5.1). Our experiments show that DP-RFT outperforms Aug-PE in terms of the fidelity and downstream utility of the generated synthetic data, and closes the gap between DP finetuning and Aug-PE in terms of downstream utility. Our key contributions are:

1. *New finetuning method for DP synthetic data generation*: We propose DP-RFT , which leverages DP-protected reward function and reinforcement learning to train a LLM against private corpus, without *ever* ingesting private examples directly during the training process and hence satisfying the *eyes-off* constraint.
2. *In-depth experiments and analyses*: We conduct comprehensive evaluation of our methods on four datasets, showing significant improvement on downstream utility, especially for a tight privacy budget and on datasets which are out-of-distribution for the backbone language model. Our qualitative analysis further reveals that DP-RFT is able to better capture lexical and structural similarity of long-form, structural outputs.

2 Background and Related work

Differential Privacy with LLMs. As LLMs are increasingly deployed in many real-world cases, preserving the privacy of training and evaluation data has been an active research area. One popular approach is to apply DP-SGD (Abadi et al., 2016) to train LLMs for classification (Yu et al., 2021; Li et al., 2021) or generation tasks (Wang et al., 2025a; Yu et al., 2021; Li et al., 2021; Yue et al., 2023; Mattern et al., 2022; Kurakin et al., 2023; Ngong et al., 2024; Tan et al., 2025). Later work proposed training-free approaches, such as injecting calibrated DP noise into the token-by-token generation process of LLMs (Tang et al., 2023; Flemings et al., 2025; Duan et al., 2023), or aggregating multiple LLM outputs in a DP manner (Wu et al., 2023). While these methods provide theoretical DP guarantees, they all require exposing the private data as input to the LLM. In practice, there could be regulatory constraints that require the private data to be completely *eyes-off*, preventing the adoption of such methods. Private Evolution (PE) (Lin et al., 2024; Xie et al., 2024; Lin et al., 2025; Wang et al., 2025b) is a newly emerged framework to address this challenge. It proposes an iterative prompting pipeline which steers the model generation to be more similar to the private data, as measured by embedding similarity. While these methods avoid private inputs, their reliance on un-finetuned models limits the synthetic data quality. Our DP-RFT aims to take the best of both worlds: adapting model weights to better fit private data without directly inputting it into the model. Recent work (Nguyen et al., 2025) proposes to generate synthetic data via gradient matching, which requires gradient computation on private data. Notably, Hou et al. (2025) recently proposed fine-tuning LLMs using embedding similarity as a reward with an offline reinforcement learning algorithm (DPO; Rafailov et al., 2023). In contrast, our method leverages an on-policy reinforcement learning algorithm such as PPO, eliminating the need to generate preference pairs in an iterative manner, which reduces complexity to the pipeline. We further explore similarity functions beyond

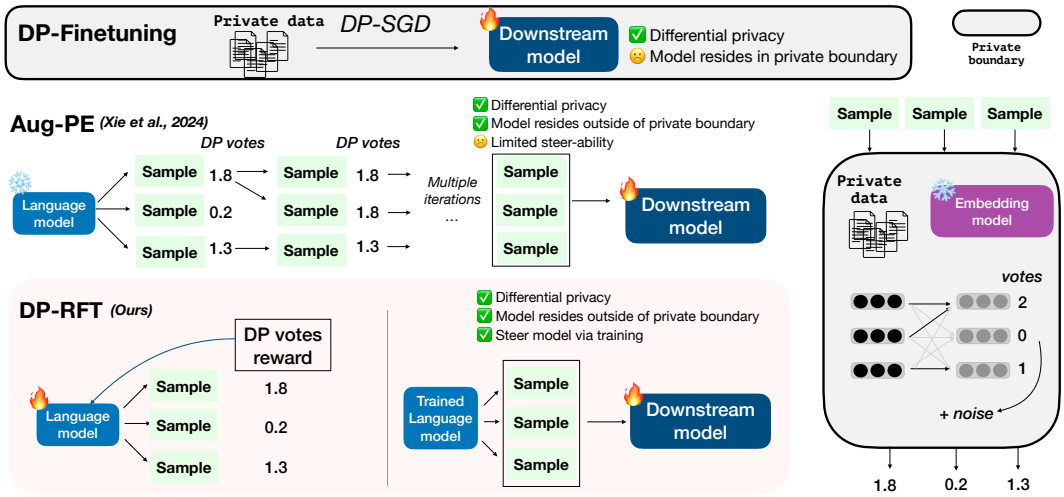


Figure 1: An illustration of DP-RFT and comparison with prior methods. DP-RFT fine-tunes a language model (LM) to generate texts similar to the private document with reinforcement learning, guided by a DP-protected nearest-neighbor votes as reward. Unlike DP-Finetuning which requires accessing the private data directly, DP-RFT and Aug-PE allow downstream model training outside of the private boundary. While Aug-PE is bounded by samples generated from a frozen LM, DP-RFT generates synthetic data with higher fidelity by training an LM to generate texts more similar to the private data.

embedding similarity, such as structural similarity for long-form, structural text such as meeting transcripts and chat logs.

Reinforcement fine-tuning (RFT) for LLMs. We briefly review how reinforcement learning is used to post-train large language models. Let a language model with parameters θ define a policy $\pi_\theta(\cdot \mid p)$ over token sequences given a prompt p. The goal of RFT is to update θ so that samples $d \sim \pi_\theta(\cdot \mid p)$ have higher task-specific utility, quantified by a scalar reward $R(d, p)$. In modern LLM post-training, R can come from a learned reward model (e.g., RLHF, Ouyang et al., 2022), or from verifiable signals (e.g., passing tests in code/math), and training proceeds by maximizing the cumulative rewards. Practically, RFT alternates between: (i) sampling on-policy generations from π_θ for a batch of prompts; (ii) computing scalar rewards for each sample; and (iii) updating θ via a policy gradient method (e.g., PPO, Schulman et al., 2017; or GRPO, Shao et al., 2024) on those samples. Compared to supervised fine-tuning (SFT), where LLM ingests reference texts as input during training, RFT directly optimizes the task metric exposed by R and is thus effective when high-quality labels are scarce or when a reward function R can be defined. The fact that RFT does not ingest example outputs during training is the key to our DP-RFT algorithm's design.

3 DP-RFT: Differentially-Private Reinforcement Finetuning

Given a set of private documents D_{priv}, our goal is to train a language model M_{gen} that generates synthetic documents D_{syn} similar to D_{priv}. M_{gen} takes an input prompt p which contains public information about the private corpus and produces a synthetic document d. We describe our method, DP-RFT, which leverages reinforcement fine-tuning (RFT) to train M_{gen} against a DP-protected reward function R. We illustrate the workflow of DP-RFT, as well as its differences from prior approaches, including DP-Finetuning (Abadi et al., 2016) and Augmented Private Evolution (Aug-PE; 2024) in Figure 1.

3.1 Generating input prompts

We use a large language model (LLM) to construct input prompts which contain public information about the target domain. We construct a diverse set of prompts to encourage diverse generation from

M_{gen} following Augmented Private Evolution (Aug-PE) (Xie et al., 2024) for training and inference of DP-RFT. We focus on one or both of the two axes of diversity (1) content diversity and (2) length diversity. For content diversity, we generate a set of keywords using an LLM. For instance, for BBC articles, we prompt an LLM to generate a list of topics (e.g. Politics, Sports), along with keywords associated with each topic. We obtain a set of prompts, each with a different set of keywords, which we use as input to M_{gen}. For length diversity, we define a possible range of length for the private documents and random sample one length to include in the prompt (e.g. "The generated document should contain around 300 words"). We include the details of the procedure in §E in the appendix.

3.2 REWARDS

DP voting as reward. Aug-PE (Xie et al., 2024) prompts language models to generate multiple synthetic documents and iteratively generate synthetic documents via a process called **DP voting**: each private document votes for a synthetic document that is most similar to them measured with a similarity function, such as a text embedding model. They find that selecting synthetic documents that receive more votes (i.e. close to more private documents in the embedding space) lead to better synthetic documents and hence downstream performance. We take inspiration from Aug-PE and propose the following new designs to the DP-voting mechanism and the iterative generation-sampling process: **(1) Using raw similarity scores as votes.** In Aug-PE, each private sample votes only for its closest synthetic document, producing a binary (1 or 0) signal. This discards valuable information in the raw similarity scores, such as *how much* one synthetic sample is better than another. To capture this richer signal, we instead use the raw similarity scores $sim(d, D_{priv,i})$ directly as votes from the i-th private document $D_{priv,i}$ to the synthetic document d, where sim is a similarity measure (e.g. text embedding similarity). **(2) DP-RFT.** While Aug-PE can promote synthetic documents that are similar to the private documents, the model is untrained, and hence has limited steerability. We instead leverage the votes as the reward in Reinforcement Fine-Tuning (RFT) to steer M_{gen} to generate synthetic documents more similar to the private corpus. Concretely, for a synthetic document d and a predefined similarity measure sim, we define $R_{sim} = \frac{1}{|D_{priv}|}\sum_{i=1}^{i=|D_{priv}|} sim(d, D_{priv,i})$ as the reward.

Input: A set of s synthetic documents D_{syn}, a similarity metric sim, a set of p private documents D_{priv}, a noise multiplier σ, a raw similarity threshold c.
Output: Reward R_{sim} for synthetic documents D_{syn}

$R_{sim} \leftarrow [\,]$
for $i \in \{1, \ldots, s\}$ **do**
  raw_similarity $\leftarrow [\,]$
  for $j \in \{1, \ldots, p\}$ **do**
    $S_{i,j} \leftarrow sim(D_{syn,i}, D_{priv,j})$
    if $\sigma > 0$ **then**
      $S_{i,j} = min(S_{i,j}, c)$
    end if
    raw_similarity.append($S_{i,j}$)
  end for
  raw_similarity = sum(raw_similarity)
  if $\sigma > 0$ **then**
    raw_similarity += $\mathcal{N}(0, \sigma c\sqrt{s})$
  end if
  R_{sim}.append($\frac{\text{raw_similarity}}{p}$)
end for
Return: R_{sim}

Figure 2: Procedure for computing similarity reward (R_{sim}) given a noise multiplier σ for a batch of s synthetic samples.

Ensuring differential privacy. To ensure differential privacy, we add Gaussian noise to R_{sim} following Aug-PE. For a synthetic example d, we first obtain its similarity to each of the private documents $sim(d, D_{priv,i})$ and clip the similarity score to a threshold c. Given a noise multiplier σ and a batch of synthetic samples of size s, we add Gaussian noise $\mathcal{N}(0, \sigma c\sqrt{s})$ to $\sum_i sim(d, D_{priv,i})$, i.e., the sum of similarity between the synthetic document and all private documents. Finally, we obtain the average similarity scores across all private documents. We provide the pseudocode of this procedure in Figure 2. The privacy analysis follows the Private Evolution framework (Lin et al., 2024), which is reproduced in §F for completeness.

Mitigating reward hacking. Although optimizing for R_{sim} can lead to synthetic documents similar to the private corpus defined by sim, relying only on R_{sim} as the sole reward can lead to reward hacking (Amodei et al., 2016; Pan et al., 2022). Unlike prior RFT methods which focus on math and coding problems (DeepSeek-AI, 2025; Wei et al., 2025) where a verifiable reward (e.g.

presence of the gold answer string) can be defined, generating documents that are similar to a set of documents is relatively open-ended and unconstrained. Moreover, text embedding models are known to have various biases (Coelho et al., 2024; Fayyaz et al., 2025) and could assign high similarity scores to dissimilar documents, especially when embedding *long* documents. Thus, in addition to R_{sim}, we employ a reward R_{prompt} that measures adherence to the input prompt to prevent reward hacking.[1] We implement R_{prompt} as a combination of LLM-as-a-judge and rule-based checks (§4).

To combine the two rewards, we use a threshold approach. We set the reward to 0 if R_{prompt} is less than a threshold τ and otherwise return R_{sim}. The final reward is defined as:

$$R = \begin{cases} R_{\text{sim}} & \text{if } R_{\text{prompt}} > \tau \\ 0 & \text{otherwise} \end{cases}$$

We train DP-RFT using Proximal Policy Optimization (PPO, Schulman et al. (2017)).

4 Experiments

4.1 Evaluation settings

Datasets. We evaluate DP-RFT and the baselines on four publicly available datasets that are treated as private datasets, covering different domains and document structures:

- *Domain-specific document-level data:* **PubMed**, which contains abstracts of medical papers crawled by Yu et al. (2023) from 2023/08/01 to 2023/08/07; **BBC News article** from 2010 to 2017 released by Narayan et al. (2018)
- *Structured long-form data:* **WildChat** (Zhao et al., 2024), which contains chat history between a user and an LLM; collected between 2023 to 2024 and **QMSum** (Zhong et al., 2021) which contains meeting transcripts.

The average number of tokens for PubMed, BBC article, WildChat and QMSum are 361, 513, 2,799 and 2,857 respectively. We include details about these datasets in §B.1 in the appendix.

Intrinsic evaluation: similarity with real data. We report quantitative metrics including the average and maximal embedding similarity with the private corpus and the embedding distribution distance, i.e., the Fréchet Inception Distance (Heusel et al., 2017), following prior work (Xie et al., 2024). As it is non-trivial to measure similarity of long documents, we also employ LLM-as-a-judge to evaluate pairwise similarity of two synthetic documents against a private document and conduct qualitative evaluation on document structure in §6.

Extrinsic evaluation: downstream performance. Aside from intrinsic similarity evaluation, the synthetic text should be helpful in downstream utility. We fine-tune a language model via next token prediction on the synthetic text generated and report the next token prediction accuracy on the private test set. Following prior work (Yu et al., 2023; Xie et al., 2024; Hou et al., 2025), we conduct experiments fine-tuning $\text{BERT}_{\text{Small}}$ (Devlin et al., 2019) as a causal language model by modifying bidirectional attention to causal attention. As $\text{BERT}_{\text{Small}}$ is relatively small-scale, we also fine-tune and evaluate GPT-2 (Radford et al., 2019).

4.2 DP-RFT and baseline settings

We provide the implementation details for each component for DP-RFT as described in §3.

M_{gen} and R_{sim}: For all datasets, we use Qwen-2.5-3B-Instruct (Qwen Team, 2024) as M_{gen} and gte-Qwen-2-1.5B-Instruct (Li et al., 2023) to measure R_{sim}. For WildChat and QMSum, we also measure the distributions of word-level jaccard similarity and word counts of all speech turn in a synthetic vs. private example, and use their KL divergence as two additional terms in R_{sim}, in order to encourage M_{gen} to learn the structural properties of private corpus.

[1] While RL algorithms such as PPO includes a KL penalty term to reduce divergence from the reference model, we find that increasing KL penalty does not resolve the issue in our preliminary experiments. Instead, it discourages the model from learning to generate samples closer to the private distribution.

R_{prompt}: We use gpt-4o (OpenAI, 2024) as the LLM-as-a-judge to evaluate how well the output adheres to the prompt. Concretely, given the input prompt p and generated documents d, the model outputs a scalar rating from 1 to 10 on how well d adheres to p. We include the exact prompt in §E. To evaluate adherence to the length instruction for the BBC article and PubMed, we employ a rule-based approach that checks the difference between the length of the generated documents and the specified length in the prompt. If the length difference is greater than a threshold d or the LLM-as-a-judge output is less than a threshold τ, we set the reward to 0. We set the threshold τ to 6 and d as {50, 100} for PubMed and BBC articles respectively. We conduct ablation studies on removing R_prompt in § 6.1.

Input prompts. Following prior work (Xie et al., 2024), we construct input prompts for each dataset by prompting an LM to generate keywords based on public knowledge. We design a multi-stage pipeline to ensure diversity of the prompts. For BBC articles, we first prompt the language model to generate a list of categories of BBC articles. We then prompt the models to generate keywords for each category. For PubMed, we use the writers released by Xie et al. (2024) and prompt the language model to generate 100 technical terms for each of the writer. The prompt template for each dataset in Table 5 in the Appendix.

Baselines. We consider several baseline methods which train an LM with differential privacy. **Aug-PE** (Xie et al., 2024), which steers generation from a language model to be more similar to the private data by iterative prompting; **QWEN**: which prompts the backbone model to generate synthetic data; this represents the performance of the model before DP-RFT training. We also compared to **DP-FT** which fine-tunes the model on the downstream task with DP-SGD (Yu et al., 2022; Li et al., 2022), though we note that **DP-FT** requires ingesting the private text as *input* to train the model, and thus violating the *eyes-off* setting. For all datasets, we follow the setup of Xie et al. (2024) and report Aug-PE results of 10 iterations. We use the same generation model and embedding model as DP-RFT to ensure a fair comparison. We use the input prompt for DP-RFT as `RANDOM_API` and the fill-in-the-blank `VARIATION_API`. We include exact prompts in §B.1.1 in the Appendix; as well as example private text and generated text in §C.1. We report results for training on a subset of private data (2,000) and compare with training on 2,000 data generated by DP-RFT and Aug-PE, except for QMSum, which only contains 700 examples in total. We generate synthetic text and keep the training data size consistent for all methods.

Privacy Setting. We report the performance of different privacy budgets controlled by ϵ. We train for 100 steps for PubMed and QMSum, and 200 steps for BBC articles and WildChat. We experiment with privacy budgets ϵ = {1,2,4,∞}. The Gaussian noise multipliers being added to the DP votes are derived by the number of training steps and the size of the private corpus accordingly. The details are described in § F. For DP-RFT, we use noise multiplier σ = {41.90, 22.14, 11.86, 0} for PubMed, σ = {29.98, 16.45, 9.17, 0} for QMSum, and σ = {52.50, 28.07, 15.23, 0} for BBC articles and WildChat. We clip the maximum value of raw similarity score at 0.5 for BBC article and PubMed, and 0.8 for QMSum and WildChat.

5 RESULTS

5.1 DOWNSTREAM EVALUATION

Table 1 reports the results on downstream performance for GPT-2 and $\text{BERT}_{\text{Small}}$.

DP-RFT outperforms Aug-PE across datasets and privacy budgets. Across all datasets and privacy budgets, DP-RFT consistently achieves higher downstream accuracy than Aug-PE on both GPT-2 and $\text{BERT}_{\text{Small}}$. The gains are particularly pronounced for $\text{BERT}_{\text{Small}}$, with up to 22%, 35%, 2%, and 47% relative improvement on PubMed, BBC, WildChat, and QMSum, respectively, and are also substantial for GPT-2, reaching up to 10%, 4%, 4%, and 2% on these datasets. These improvements indicate that DP-RFT is able to effectively shift the distribution of the LLM compared to Aug-PE, which iteratively prompts a frozen LLM. In addition, DP-RFT consistently outperforms the un-finetuned QWEN baseline (except for BBC at $\epsilon = 2$, where performance is comparable), further demonstrating the benefits of training M_{gen}.

Table 1: Evaluation on downstream model performance using Qwen-2.5-3B-Instruct as the synthetic data generator to finetune GPT2 and $\text{BERT}_{\text{Small}}$. We note that **DP-FT** violates the *eyes-off* constraint (color-coded in grey). The highest performance among *eyes-off* approaches is **bolded**. DP-RFT closes the gap between DP-FT and eyes-off approach such as Aug-PE on GPT-2, and outperform DP-FT when $\epsilon \neq \infty$ for $\text{BERT}_{\text{Small}}$.

Dataset	Method	Data Type	GPT-2 Next Token Accuracy (↑)				$\text{BERT}_{\text{Small}}$ Next Token Accuracy (↑)			
			$\epsilon = \infty$	$\epsilon = 4$	$\epsilon = 2$	$\epsilon = 1$	$\epsilon = \infty$	$\epsilon = 4$	$\epsilon = 2$	$\epsilon = 1$
PubMed	DP-FT	Private	40.33	42.33	42.06	41.64	20.79	9	8	6
	QWEN	Synthetic	33.04	33.04	33.04	33.04	16.88	16.88	16.88	**16.88**
	Aug-PE	Synthetic	33.91	34.10	33.57	34.55	13.96	14.41	14.16	13.71
	DP-RFT	Synthetic	**38.31**	**37.37**	**37.19**	**36.75**	**17.46**	**17.81**	**17.32**	16.79
BBC	DP-FT	Private	39.13	40.05	39.78	39.32	17.97	9.29	7.96	6.90
	QWEN	Synthetic	33.49	**33.49**	**33.49**	33.49	13.07	13.07	**13.07**	13.07
	Aug-PE	Synthetic	30.97	31.78	31.54	32.25	11.63	10.78	10.03	10.00
	DP-RFT	Synthetic	**33.57**	32.81	32.84	**33.92**	**13.72**	**13.25**	13.03	**13.51**
WildChat	DP-FT	Private	51.00	47.95	47.59	47.06	19.62	8.28	7.37	5.30
	QWEN	Synthetic	33.00	33.00	33.00	33.00	5.22	5.22	5.22	5.22
	Aug-PE	Synthetic	42.39	43.61	43.65	43.46	13.29	14.04	13.81	13.63
	DP-RFT	Synthetic	**45.27**	**45.42**	**45.30**	**45.03**	**13.93**	**14.08**	**14.14**	**13.89**
QMSum	DP-FT	Private data	47.00	42.71	42.45	42.13	32.82	11.52	9.88	7.96
	QWEN	Synthetic	24.00	24.00	24.00	24.00	3.86	3.86	3.86	3.86
	Aug-PE	Synthetic	32.60	33.06	33.57	**34.02**	8.23	8.53	7.82	7.63
	DP-RFT	Synthetic	**33.00**	**33.68**	**34.33**	33.38	**11.03**	**10.97**	**10.95**	**11.11**

DP-RFT outperforms DP-FT under tight privacy budgets and low-data regimes for $\text{BERT}_{\text{Small}}$. While DP-FT operates under a looser privacy constraint (i.e., violating the *eyes-off* setting), its performance for $\text{BERT}_{\text{Small}}$ degrades substantially as the privacy budget decreases, underperforming DP-RFT across all datasets when $\epsilon \neq \infty$. In contrast, for GPT-2, DP-FT consistently outperforms DP-RFT across privacy budgets and datasets. We attribute this difference to the fact that $\text{BERT}_{\text{Small}}$ is converted into a causal language model during fine-tuning, which differs from its pre-training setup. To further investigate this effect, we scale up the amount of private data to the full dataset in Appendix G, observing improved downstream performance for DP-FT. Yet, DP-RFT continues to outperform DP-FT for $\text{BERT}_{\text{Small}}$ in low-data regimes (e.g., QMSum with only 700 examples) or under tight privacy budgets, highlighting its practical advantages in realistic privacy-constrained settings. We also report results against another baseline method that violates the *eyes-off* constraint (fine-tuning the data generator with DP-SGD on the private data) in Appendix G.

5.2 Similarity evaluation

Table 2 reports similarity metrics between synthetic and private data. Overall, DP-RFT exhibits higher similarity to private documents than the unfine-tuned QWEN across both instance-level metrics (mean and maximum similarity) and distributional measures (FID), indicating that DP-RFT more effectively captures the underlying data distribution. In addition, DP-RFT outperforms Aug-PE for instance-level metrics in most settings, with the advantage becoming more pronounced under tighter privacy budgets.

We observe dataset-dependent trends in similarity performance. For example, with $\epsilon = \infty$, DP-RFT and Aug-PE achieve comparable embedding similarity on PubMed, whereas DP-RFT yields approximately 20% higher similarity on the other three datasets. We attribute this difference to PubMed being closer to the distribution of the backbone generator M_{gen}, making it easier to steer via prompting. To quantify this, we measure the distance between the target distribution and M_{gen} using perplexity of the private data under M_{gen}, finding that PubMed has lower perplexity (6.75) than BBC (9.11) and QMSum (10.14). Although WildChat also exhibits low perplexity (4.04) as it contains model-generated texts, its long-form and structured nature makes it more difficult to steer through iterative prompting. We note that Aug-PE achieves lower FiD on PubMed and WildChat, which may be attributed to the fact that Aug-PE explicitly optimizes FiD and can therefore perform better on distributions that are more familiar to the base model. To complement embedding similarity,

Table 2: Similarity evaluation with the private corpus using Qwen-2.5-3B-Instruct as the synthetic data generator. DP-RFT generates documents that are more similar to the private dataset than baseline methods as measured by embedding similarity.

Dataset	Method	Mean/Max embedding similarity (↑)				FID (↓)			
		$\epsilon = \infty$	$\epsilon = 4$	$\epsilon = 2$	$\epsilon = 1$	$\epsilon = \infty$	$\epsilon = 4$	$\epsilon = 2$	$\epsilon = 1$
PubMed	Private	0.54/0.88	-	-	-	0.06	-	-	-
	QWEN	0.44/0.73	-	-	-	0.50	-	-	-
	Aug-PE	**0.48/0.77**	**0.49/0.79**	0.46/0.74	0.43/0.70	**0.39**	**0.39**	**0.41**	**0.46**
	DP-RFT	0.47/0.70	**0.49**/0.71	**0.51/0.73**	**0.52/0.74**	0.46	0.48	0.45	0.47
BBC	Private	0.43/0.78	-	-	-	0.09	-	-	-
	QWEN	0.27/0.52	-	-	-	0.64	-	-	-
	Aug-PE	0.30/0.55	0.21/0.45	0.20/0.43	0.19/0.42	**0.49**	0.68	0.78	0.79
	DP-RFT	**0.35/0.59**	**0.33/0.57**	**0.34/0.57**	**0.35/0.58**	0.55	**0.57**	**0.57**	**0.52**
WildChat	Private	0.28/0.67	-	-	-	0.07	-	-	-
	QWEN	0.27/0.64	-	-	-	1.06	-	-	-
	Aug-PE	0.24/0.57	0.25/0.57	0.24/0.57	0.23/0.54	**0.39**	**0.32**	**0.39**	**0.38**
	DP-RFT	**0.30/0.78**	**0.31/0.77**	**0.31/0.78**	0.**30/0.77**	0.74	0.66	0.71	0.71
QMSum	Private	0.52/0.84	-	-	-	0.24	-	-	-
	QWEN	0.32/0.48	-	-	-	1.11	-	-	-
	Aug-PE	0.34/0.54	0.30/0.49	0.31/0.49	0.30/0.48	**0.70**	0.79	0.80	**0.82**
	DP-RFT	**0.37/0.62**	**0.38/0.65**	**0.38/0.64**	**0.38/0.65**	0.80	0.79	**0.75**	0.84

we conduct an LLM-as-a-judge pairwise similarity evaluation against private document on use of word and tone, finding that DP-RFT is consistently preferred over Aug-PE across all datasets. We describe the setting and results in details in §D in the Appendix.

6 Analysis and ablation

6.1 Ablation study

We propose to include R_{prompt} to mitigate reward hacking by over-optimizing on R_{sim} (§3.2). We conduct ablation study which removes R_{prompt} from the reward on PubMed with $\epsilon = \infty$ and report evaluation on both downstream performance and similarity measure. Results are reported in Table 3. We see that while model trained without R_{prompt} (using R_{sim} as the reward) achieves higher embedding similarity with the private corpus, the downstream performance is notably worse across the two models. This suggests that solely optimizing for embedding similarity could harm performance. We further conduct ablation study on varying R_{sim} and M_{gen} in §C.

6.2 Qualitative Analysis on the Synthetic Examples

Lexical Similarity. We analyze the distribution of Jaccard similarity between each synthetic example and its *most* similar private example, as shown in Figure 3. Across privacy budgets, examples generated by DP-RFT exhibit higher lexical overlap with the private data than those produced by Aug-PE. In contrast, the distribution of examples from Aug-PE is more spread out, potentially indicating that Aug-PE outputs are more diverse, despite lower fidelity in terms of lexical similarity. This observation is consistent with the FiD results

Table 3: Ablating R_{prompt} for PubMed when $\epsilon = \infty$. Removing R_{prompt} leads to better embedding similarity but worse downstream performance.

Method	Acc (↑) GPT-2 / $\text{BERT}_{\text{Small}}$	Embedding sim mean/max(↑)	FID (↓)
DP-FT	42.33/20.97	0.54/0.88	0.06
Aug-PE	33.91/13.96	0.48/0.77	0.39
DP-RFT	38.31/17.46	0.47/0.70	0.46
- R_{prompt}	36.85/16.31	0.56/0.78	0.45

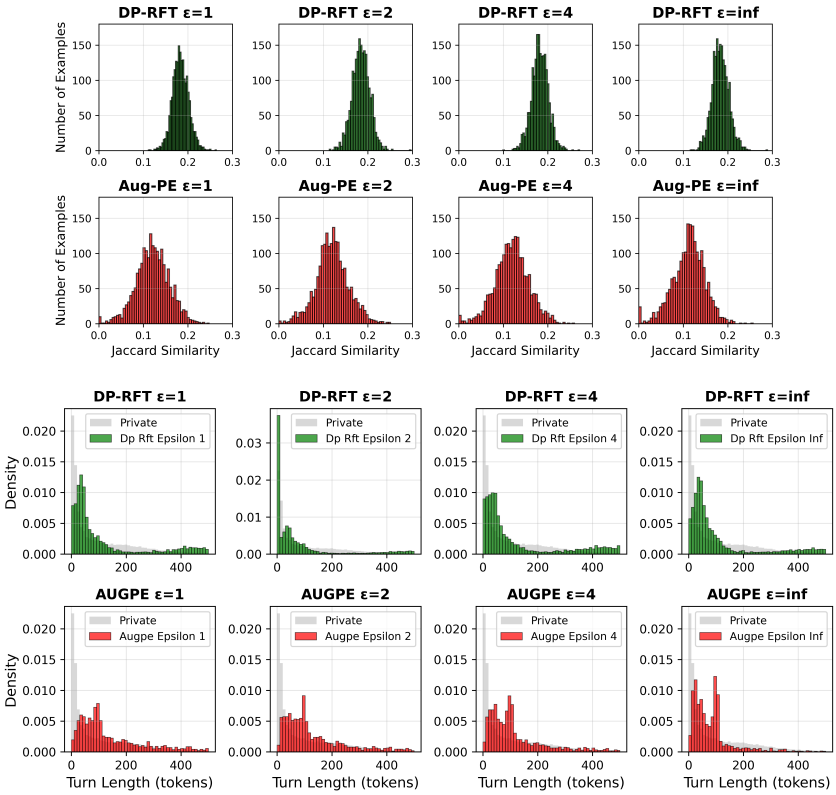


Figure 3: Analysis of structural similarity on WildChat. Top: Histogram of word-level maximum Jaccard similarity of synthetic WildChat examples against all private examples. Bottom: Distribution of per-turn word counts from synthetic WildChat chatlogs generated by DP-RFT vs. Aug-PE under different privacy budgets ($\epsilon = \{1, 2, 4, \infty\}$). Private WildChat examples are shown in gray.

in §5.2, where Aug-PE achieves lower FiD despite worse average embedding similarity. Incorporating more explicit diversity control into DP-RFT is a promising direction for future work.

Structural Similarity. For structured data, both content and structural similarity of synthetic examples are important. Figure 3 compares the per-turn word-count distributions of synthetic and real WildChat conversations across privacy budgets. Overall, synthetic data generated by DP-RFT more closely matches the private distribution than Aug-PE across all privacy settings. Interestingly, however, Aug-PE better captures the bi-modal structure of the distribution, reflecting the mixture of short user turns and longer assistant responses. One possible explanation is that Aug-PE employs a generate-then-select strategy over a large candidate set at each iteration (Xie et al., 2024), whereas DP-RFT optimizes over smaller mini-batches of rollouts. Future work can examine the effect of rollout batch size in DP-RFT training more closely.

7 Conclusion

We introduce DP-RFT, a reinforcement learning method for training language models to generate synthetic data aligned with a private corpus *without* ingesting individual private examples. The key idea of our method is to steer the model with reinforcement learning guided by a similarity measure with the private document as reward, while respecting the eyes-off constraint. Through extensive experiments on four datasets spanning diverse domains and structures, we show that DP-RFT consistently improves both fidelity and downstream utility over prior eyes-off baselines such as Aug-PE, and narrows the gap with DP finetuning methods under tight privacy budgets.

AI USE STATEMENT

We use large language models as part of the methodology in three places:

- to construct public input prompts for synthetic data generation (subsection 3.1, Appendix E),
- to evaluate prompt adherence for R_{prompt} (subsection 4.2),
- to perform pairwise stylistic similarity evaluation in Appendix D.

We use gpt-4o for LLM-as-a-judge evaluations. We do not use LLMs to produce experimental results beyond the LLM-based components explicitly described above. A generative AI assistant was used to convert this manuscript's LaTeX from the NeurIPS 2026 template to the ICLR 2027 template. This formatting conversion did not introduce new scientific claims, results, or analyses. The authors reviewed the AI-assisted work and take responsibility for the final content of the paper.

REPRODUCIBILITY STATEMENT

We include implementation details for reproducing DP-RFT in Section E and for reproducing baseline methods in Section B.1.1. Implementation details for downstream evaluation is documented at Appendix G.

ETHICS STATEMENT

This work advances machine learning by exploring a reinforcement learning–based approach to differentially private synthetic data generation without eyes-on access to private examples. By using DP-protected similarity signals as training feedback, DP-RFT enables large language models to generate synthetic texts better aligned with private data distributions while respecting formal privacy constraints. This reduces the need for direct exposure to sensitive data and may help lower privacy risks in applications where data sharing is restricted. More broadly, our results suggest a path toward more practical and responsible use of LLMs in privacy-sensitive domains.

At the same time, DP-RFT relies on a similarity function that can meaningfully capture the relationship between synthetic and private texts. If the similarity function, such as a text embedding model, fails to capture important domain-specific or structural properties, the quality and utility of the generated synthetic data may be limited. This can be particularly consequential in sensitive domains, where downstream models trained on low-fidelity synthetic data may produce incomplete, biased, or misleading outputs. In addition, although differential privacy limits the influence of individual private examples, synthetic data may still reflect aggregate biases or sensitive population-level patterns present in the original corpus. We therefore view DP-RFT as a tool that should be paired with domain-specific validation and auditing before downstream deployment.

REFERENCES

Martín Abadi, Andy Chu, Ian J. Goodfellow, H. B. McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016. URL https://api.semanticscholar.org/CorpusID:207241585.

Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Francis Christiano, John Schulman, and Dandelion Mané. Concrete problems in ai safety. *ArXiv*, abs/1606.06565, 2016. URL https://api.semanticscholar.org/CorpusID:10242377.

Borja Balle and Yu-Xiang Wang. Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In *International Conference on Machine Learning*, pp. 394–403. PMLR, 2018.

Stella Biderman, Hailey Schoelkopf, Quentin Gregory Anthony, Herbie Bradley, Kyle O'Brien, Eric Hallahan, Mohammad Aflah Khan, Shivanshu Purohit, USVSN Sai Prashanth, Edward Raff, et al. Pythia: A suite for analyzing large language models across training and scaling. In *International Conference on Machine Learning*, pp. 2397–2430. PMLR, 2023.

Nicholas Carlini, Florian Tramèr, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom B. Brown, Dawn Xiaodong Song, Úlfar Erlingsson, Alina Oprea, and Colin Raffel. Extracting training data from large language models. In *USENIX Security Symposium*, 2020. URL https://api.semanticscholar.org/CorpusID:229156229.

Nicholas Carlini, Daniel Paleka, Krishnamurthy Dj Dvijotham, Thomas Steinke, Jonathan Hayase, A Feder Cooper, Katherine Lee, Matthew Jagielski, Milad Nasr, Arthur Conmy, et al. Stealing part of a production language model. In *International Conference on Machine Learning*, pp. 5680–5705. PMLR, 2024.

Kang Chen, Xiuze Zhou, Yuan-Chyuan Andy Lin, Shibo Feng, Li Shen, and Pengcheng Wu. A survey on privacy risks and protection in large language models. *Journal of King Saud University Computer and Information Sciences*, 37, 2025. URL https://api.semanticscholar.org/CorpusID:278327688.

João Coelho, Bruno Martins, Joao Magalhaes, Jamie Callan, and Chenyan Xiong. Dwell in the beginning: How language models embed long documents for dense retrieval. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 370–377, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-short.35. URL https://aclanthology.org/2024.acl-short.35/.

DeepSeek-AI. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *ArXiv*, abs/2501.12948, 2025. URL https://api.semanticscholar.org/CorpusID:275789950.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pp. 4171–4186, 2019.

Jinshuo Dong, Aaron Roth, and Weijie J Su. Gaussian differential privacy. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(1):3–37, 2022.

Haonan Duan, Adam Dziedzic, Nicolas Papernot, and Franziska Boenisch. Flocks of stochastic parrots: Differentially private prompt learning for large language models. *Advances in Neural Information Processing Systems*, 36:76852–76871, 2023.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pp. 265–284. Springer, 2006.

Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and trends® in theoretical computer science*, 9(3–4):211–407, 2014.

Mohsen Fayyaz, Ali Modarressi, Hin rich Schuetze, and Nanyun Peng. Collapse of dense retrievers: Short, early, and literal biases outranking factual evidence. *ArXiv*, abs/2503.05037, 2025. URL https://api.semanticscholar.org/CorpusID:276885340.

James Flemings, Haosheng Gan, Hongyi Li, Meisam Razaviyayn, and Murali Annavaram. Differentially private in-context learning via sampling few-shot mixed with zero-shot outputs. *arXiv preprint arXiv:2501.19287*, 2025.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Frederik Harder, Milad Jalali, Danica J Sutherland, and Mijung Park. Pre-trained perceptual features improve differentially private image generation. *Transactions on Machine Learning Research*, 2023.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 6629–6640, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Charlie Hou, Mei-Yu Wang, Yige Zhu, Daniel Lazar, and Giulia Fanti. Private federated learning using preference-optimized synthetic data. *arXiv*, 2025. URL https://arxiv.org/abs/2504.16438.

Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, et al. Openai o1 system card. *arXiv preprint arXiv:2412.16720*, 2024.

James Jordon, Jinsung Yoon, and Mihaela Van Der Schaar. Pate-gan: Generating synthetic data with differential privacy guarantees. In *International conference on learning representations*, 2018.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.

Alexey Kurakin, Natalia Ponomareva, Umar Syed, Liam MacDermed, and A. Terzis. Harnessing large-language models to generate private synthetic text. *ArXiv*, abs/2306.01684, 2023. URL https://api.semanticscholar.org/CorpusID:259063934.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*, 2023.

Xuechen Li, Florian Tramèr, Percy Liang, and Tatsunori B. Hashimoto. Large language models can be strong differentially private learners. *ArXiv*, abs/2110.05679, 2021. URL https://api.semanticscholar.org/CorpusID:238634219.

Xuechen Li, Florian Tramer, Percy Liang, and Tatsunori Hashimoto. Large language models can be strong differentially private learners. In *International Conference on Learning Representations*, 2022.

Zehan Li, Xin Zhang, Yanzhao Zhang, Dingkun Long, Pengjun Xie, and Meishan Zhang. Towards general text embeddings with multi-stage contrastive learning. *arXiv preprint arXiv:2308.03281*, 2023.

Zinan Lin, Sivakanth Gopi, Janardhan Kulkarni, Harsha Nori, and Sergey Yekhanin. Differentially private synthetic data via foundation model apis 1: Images. In *International Conference on Learning Representations*, 2024.

Zinan Lin, Tadas Baltruaitis, and Sergey Yekhanin. Differentially private synthetic data via apis 3: Using simulators instead of foundation model. *ArXiv*, abs/2502.05505, 2025. URL https://api.semanticscholar.org/CorpusID:276250556.

Justus Mattern, Zhijing Jin, Benjamin Weggenmann, Bernhard Schoelkopf, and Mrinmaya Sachan. Differentially private language models for secure data sharing. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (eds.), *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 4860–4873, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.emnlp-main.323. URL https://aclanthology.org/2022.emnlp-main.323/.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. *ArXiv*, abs/1808.08745, 2018.

Ivoline C. Ngong, Joseph P. Near, and Niloofar Mireshghallah. Differentially private learning needs better model initialization and self-distillation. *ArXiv*, abs/2410.17566, 2024. URL https://api.semanticscholar.org/CorpusID:273532187.

Dang Nguyen, Zeman Li, Mohammadhossein Bateni, Vahab Mirrokni, Meisam Razaviyayn, and Baharan Mirzasoleiman. Synthetic text generation for training large language models via gradient matching. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=mHySkOp46b.

OpenAI. Gpt-4o system card. *ArXiv*, abs/2410.21276, 2024. URL https://api.semanticscholar.org/CorpusID:273662196.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke E. Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Francis Christiano, Jan Leike, and Ryan J. Lowe. Training language models to follow instructions with human feedback. *ArXiv*, abs/2203.02155, 2022. URL https://api.semanticscholar.org/CorpusID:246426909.

Alexander Pan, Kush Bhatia, and Jacob Steinhardt. The effects of reward misspecification: Mapping and mitigating misaligned models. *ArXiv*, abs/2201.03544, 2022. URL https://api.semanticscholar.org/CorpusID:245837268.

Ashwinee Panda, Xinyu Tang, Christopher A Choquette-Choo, Milad Nasr, and Prateek Mittal. Privacy auditing of large language models. In *The Thirteenth International Conference on Learning Representations*, 2025.

Natalia Ponomareva, Hussein Hazimeh, Alex Kurakin, Zheng Xu, Carson Denison, H Brendan McMahan, Sergei Vassilvitskii, Steve Chien, and Abhradeep Guha Thakurta. How to dp-fy ml: A practical guide to machine learning with differential privacy. *Journal of Artificial Intelligence Research*, 77:1113–1201, 2023.

Qwen Team. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*, 2024.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *ArXiv*, abs/2305.18290, 2023. URL https://api.semanticscholar.org/CorpusID:258959321.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *ArXiv*, abs/1707.06347, 2017. URL https://api.semanticscholar.org/CorpusID:28695052.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Jun-Mei Song, Mingchuan Zhang, Y. K. Li, Yu Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *ArXiv*, abs/2402.03300, 2024. URL https://api.semanticscholar.org/CorpusID:267412607.

Bowen Tan, Zheng Xu, Eric P. Xing, Zhiting Hu, and Shanshan Wu. Synthesizing privacy-preserving text data via finetuning without finetuning billion-scale llms. *ArXiv*, abs/2503.12347, 2025. URL https://api.semanticscholar.org/CorpusID:277066119.

Xinyu Tang, Richard Shin, Huseyin A Inan, Andre Manoel, Fatemehsadat Mireshghallah, Zinan Lin, Sivakanth Gopi, Janardhan Kulkarni, and Robert Sim. Privacy-preserving in-context learning with differentially private few-shot generation. *arXiv preprint arXiv:2309.11765*, 2023.

Jianwei Wang, Junyao Yang, Haoran Li, Huiping Zhuang, Cen Chen, and Ziqian Zeng. Rewardds: Privacy-preserving fine-tuning for large language models via reward driven data synthesis. *ArXiv*, abs/2502.18517, 2025a. URL https://api.semanticscholar.org/CorpusID:276618102.

Shuaiqi Wang, Vikas Raunak, Arturs Backurs, Victor Reis, Pei Zhou, Sihao Chen, Longqi Yang, Zinan Lin, Sergey Yekhanin, and Giulia Fanti. Struct-bench: A benchmark for differentially private structured text generation. *arXiv preprint arXiv:2509.10696*, 2025b.

Yuxiang Wei, Olivier Duchenne, Jade Copet, Quentin Carbonneaux, Lingming Zhang, Daniel Fried, Gabriele Synnaeve, Rishabh Singh, and Sida Wang. Swe-rl: Advancing llm reasoning via reinforcement learning on open software evolution. *ArXiv*, abs/2502.18449, 2025. URL https://api.semanticscholar.org/CorpusID:276580226.

Tong Wu, Ashwinee Panda, Jiachen T Wang, and Prateek Mittal. Privacy-preserving in-context learning for large language models. *arXiv preprint arXiv:2305.01639*, 2023.

Chulin Xie, Zinan Lin, Arturs Backurs, Sivakanth Gopi, Da Yu, Huseyin A Inan, Harsha Nori, Haotian Jiang, Huishuai Zhang, Yin Tat Lee, Bo Li, and Sergey Yekhanin. Differentially private synthetic data via foundation model APIs 2: Text. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=LWD7upg1ob.

Da Yu, Saurabh Naik, Arturs Backurs, Sivakanth Gopi, Huseyin A. Inan, Gautam Kamath, Janardhan Kulkarni, Yin Tat Lee, Andre Manoel, Lukas Wutschitz, Sergey Yekhanin, and Huishuai Zhang. Differentially private fine-tuning of language models. *ArXiv*, abs/2110.06500, 2021. URL https://api.semanticscholar.org/CorpusID:238743879.

Da Yu, Saurabh Naik, Arturs Backurs, Sivakanth Gopi, Huseyin A Inan, Gautam Kamath, Janardhan Kulkarni, Yin Tat Lee, Andre Manoel, Lukas Wutschitz, et al. Differentially private fine-tuning of language models. In *International Conference on Learning Representations (ICLR)*, 2022.

Da Yu, Arturs Backurs, Sivakanth Gopi, Huseyin Inan, Janardhan (Jana) Kulkarni, Zinan Lin, Chulin Xie, Huishuai Zhang, and Wanrong Zhang. Training private and efficient language models with synthetic data from llms. In *NeurIPS 2023 SoLaR Workshop*, December 2023.

Xiang Yue, Huseyin Inan, Xuechen Li, Girish Kumar, Julia McAnallen, Hoda Shajari, Huan Sun, David Levitan, and Robert Sim. Synthetic text generation with differential privacy: A simple and practical recipe. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1321–1342, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.74. URL https://aclanthology.org/2023.acl-long.74/.

Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah Goodman. Star: Bootstrapping reasoning with reasoning. In *Advances in Neural Information Processing Systems*, 2022.

Wenting Zhao, Xiang Ren, Jack Hessel, Claire Cardie, Yejin Choi, and Yuntian Deng. Wildchat: 1m chatGPT interaction logs in the wild. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=Bl8u7ZRlbM.

Ming Zhong, Da Yin, Tao Yu, Ahmad Zaidi, Mutethia Mutuma, Rahul Jha, Ahmed Hassan Awadallah, Asli Celikyilmaz, Yang Liu, Xipeng Qiu, and Dragomir Radev. QMSum: A new benchmark for query-based multi-domain meeting summarization. In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (eds.), *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 5905–5921, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.472. URL https://aclanthology.org/2021.naacl-main.472/.

A LIMITATIONS

Proposed method. While our work leverage diverse input prompt and the prompt adherence reward to enforce diversity of the generated data, we do not directly optimize for the diversity of the synthetic data. Our method uses average similarity with the entire private corpus as similarity reward. Future work can look into better modeling of the *distributional* similarity of the generated synthetic data.

Experimental setting. While we conduct experiments on a wide range of domain-specific datasets, we primarily experiment with public datasets to simulate private settings as we do not have access to truly private datasets. Our experiments do not explore model scales beyond 7B, nor other embedding models. While we conduct both intrinsic and extrinsic evaluation of the generated data, our extrinsic evaluation is limited to the task of language modeling.

Table 4: Prompt used for constructing input prompts, as described in Section 3.

Stage	Prompt
	BBC article
Stage 1	You will generate a list of 500 keywords for a subcategory, which will be used to write a BBC news articles for events between 2010 and 2017. The keywords should be relevant to the subcategory, and they should be diverse and do not repeat each other. The keywords can be relevant entities, names, events, or any other relevant terms. The subcategory is {category_str}. Number the keywords and return the list of keywords separated by newline. Do not return anything else.
Stage 2	You will be given a category and a keyword related to that category that will be used to write BBC news articles for events between 2010 and 2017. You will augment the keyword with 4 other keywords, that are relevant to the category and the keyword. The sets of keywords should be relevant to the subcategory so that they can be used to write a BBC news articles. The category is {category_str} and the keyword is {keyword_str}. Return the five keywords separated by comma, do not return anything else.
	PubMed
Stage 1	Suppose that you are a {writer}. Please provide a list of 100 technical terms that are introduced by a PubMed journal article. The keywords should be diverse and do not repeat each other. Number the keywords and return the list of keywords separated by newline. Do not return anything else.

Table 5: Example input prompts for each dataset.

Dataset	Input prompt
BBC article	You are an AI assistant that acts as a journalist and write an article that mimics the style of BBC news article. Please write a BBC news article using the below keywords: Sustainable Development Goals, climate change, renewable energy, global partnership, poverty eradication. News Article:
PubMed	Suppose that you are a Clinical Researcher. Please provide an example of an abstract for a medical research paper on the topic of Platelet aggregation. The abstract should contain around 250 words. Abstract:
QMSum	You are given a topic and a list of participants. Your task is to generate a meeting transcript where the participants discuss around the topic. The transcript should contain utterances, where each utterance is of format "Participant: Content". Separate each utterance with `\n`. ### Participants: Software Engineer, Product Manager, Quality Assurance, UX Designer ### Topic: New mobile app feature brainstorming ### Transcript:
WildChat	Generate a realistic and engaging conversation between a user and ChatGPT. The user initiates the conversation with a request related to **Health & Wellness**. The task should involve **Research**. The tone of the conversation should be **Inquisitive**, and the assistant's response style should be **Narrative**. You should output the conversation in the following format. Each participant's dialogue should be on a new line. Do not include any other text or formatting. User: [User's message] Assistant: [Assistant's message] User: [User's message] ... The conversations should ideally contain multiple turns. Only output the conversation in the format above. Do NOT output anything else.",

Table 6: Prompt used for obtaining R_{prompt}.

Prompts

# Instruction 	
Please act as an impartial judge and evaluate the quality of the response provided by an AI assistant to the user instruction displayed below. Your evaluation should consider factors such as the helpfulness, relevance, depth, creativity, and level of detail of the response. Be as objective as possible. Only respond the rating number from 1 to 10. Do not include any other text or explanation in your response. Do not respond with anything else after [Rating], just the score by itself.
[Instruction]
`{instruction}`
[The Start of Assistant's Response]
`{response}`
[The End of Assistant's Response]
[Rating]

Table 7: Prompt for LLM-as-a-judge for similarity analysis.

Prompts

# Instruction 	
Please act as an impartial judge and evaluate the stylistic similarity of the responses provided by two AI assistants compared to the reference response. You should consider the elements such as use of words, style and tone. Do not consider the content match as the subject matter of the reference answer is different. Begin your evaluation by comparing both assistants' answers with the reference answer. Identify the similarity and differences. Avoid any position biases and ensure that the order in which the responses were presented does not influence your decision. Do not allow the length of the responses to influence your evaluation. Do not favor certain names of the assistants. Be as objective as possible. After providing your explanation, output your final verdict by strictly following this format: [[A]] if assistant A is better, [[B]] if assistant B is better, and [[C]] for a tie.
[Reference answer]
`{reference_answer}`
[The Start of Assistant A's Answer]
`{assistant a answer}`
[The End of Assistant A's Answer]
[The Start of Assistant B's Answer]
`{assistant b answer}`
[The End of Assistant B's Answer]

Table 8: `VARIATION_API` prompt for the PubMed dataset.

You are required to fill in the blanks with more details for the input medical abstract in a professional tone. If there is no blanks, please output the original medical abstract. Please fill in the blanks in the following sentences to write an abstract of a medical research paper: {masked_sample} and your answer MUST be exactly {word_count} words.

B Experiment details

B.1 Datasets

We report dataset statistics for all four datasets in Table 9.

B.1.1 Baseline details

Aug-PE We use the input prompt for DP-RFT as `RANDOM_API` for Aug-PE. We include samples of input prompts in Table 5. For `VARIATION_API`, we use the `fill-in-the-blank` following previous work (Xie et al., 2024). We include the prompt in Table 8. We set the max completion tokens the same as DP-RFT and the blank probability to 0.6.

Table 9: Data statistics of the private data. We report the average number of tokens measured by Qwen-2.5-Instruct.

Dataset	Average # of tokens	Private corpus size	Private test set size
PubMed	361	75,316	4,453
BBC article	513	10,000	11,057
WildChat	2,799	10,000	8,891
QMSum	2,857	725	296

Table 10: Upper: Ablating embedding models used to estimate similarity with the private corpus R_{sim} on the PubMed dataset with QWEN-2.5-3B-Instruct as M_{gen}. Bottom: Ablating language models M_{gen} on BBC article with R_{sim} as gte-1.5B.

R_{sim}	Method	BERT accuracy $\epsilon = \infty$	BERT accuracy $\epsilon = 4$
Ablating embedding model			
gte-large	Aug-PE	13.13	12.80
	DP-RFT	16.93	16.58
gte-1.5B	Aug-PE	13.96	14.41
	DP-RFT	17.46	17.81
gte-7B	Aug-PE	12.67	13.12
	DP-RFT	**18.51**	**18.24**
Ablating language model			
QWEN-3B	QWEN	13.07	
	Aug-PE	11.63	10.78
	DP-RFT	13.72	13.25
QWEN-7B	QWEN	13.46	
	Aug-PE	10.45	11.00
	DP-RFT	**14.31**	**13.69**

C Ablation study

Choice of embedding model for R_{sim}. The performance of DP-RFT correlates with how well R_{sim} capture the similarity between the generated document and the private document. Will using more capable embedding model as R_{sim} induce more similar documents from the same base model? We perform an ablation study on the embedding model to understand how it impacts DP-RFT as well as Aug-PE. We experiment with using embedding models of three different scales from the GTE (Li et al., 2023) model families: gte-large (0.5B parameters), gte-1.5B and gte-7B. We conduct experiments on the PubMed dataset under the privacy setting with epsilon of $\{\infty, 4\}$.

Results are reported in Table 10. We observe that with more powerful embedding models, DP-RFT achieves better performance for both privacy budget. This suggests that DP-RFT has the potential to generate synthetic data with higher fidelity with the private data given an embedding model that can better capture the similarity with the private document. On the other hand, the improvement we observe for DP-RFT with bigger embedding model does not hold true for Aug-PE.

Choice of the language model for M_{gen}. How will the choice of M_{gen} influence the performance of DP-RFT? We conduct an ablation study varying the backbone model M_{gen} while using the same embedding model (gte-1.5-B) as R_{sim}. We experiment with two model sizes from the QWEN family on the BBC article dataset: QWEN-2.5-3B-Instruct and QWEN-2.5-7B-Instruct and report downstream fine-tuning performance with epsilon of $\{\infty, 4\}$.

Results are in Table 10. First, we see that the un-finetuned 7B model is able to generate better text compared to the 3B model, as reflected by improved performance of the QWEN baseline. Training 7B model with DP-RFT also demonstrates better downstream performance compared to training 3B

Table 11: Evaluation on downstream model performance using Qwen-2.5-3B-Instruct as the synthetic data generator. DP-RFT outperforms all baseline methods when there is a privacy constraint ($\epsilon \neq \infty$) and with the same amount of data. The highest performance is **bolded** while the second highest is underlined. The setting with different number of training data is highlighted in grey.

Dataset	Method	Data Type	Size	$\text{BERT}_{\text{Small}}$ Next Token Accuracy (↑)			
				$\epsilon = \infty$	$\epsilon = 4$	$\epsilon = 2$	$\epsilon = 1$
PubMed	DP-FT	Private / full data	75,316	**44.88**	**18.41**	**17.99**	**17.63**
	DP-FT	Private	2,000	20.79	9	8	6
	QWEN	Synthetic	2,000	16.88	16.88	16.88	16.88
	Aug-PE	Synthetic	2,000	13.96	14.41	14.16	13.71
	DP-RFT	Synthetic	2,000	17.46	17.81	17.32	16.79
BBC	DP-FT	Private / full data	10,000	**25.75**	**14.07**	**13.71**	13.09
	DP-FT	Private	2,000	17.97	9.29	7.96	6.90
	QWEN	Synthetic	2,000	13.07	13.07	13.07	13.07
	Aug-PE	Synthetic	2,000	11.63	10.78	10.03	10.00
	DP-RFT	Synthetic	2,000	13.72	13.25	13.03	**13.51**
WildChat	DP-FT	Private / full data	10,000	**28.78**	13.93	13.52	12.80
	DP-FT	Private	2,000	19.62	8.28	7.37	5.30
	QWEN	Synthetic	2,000	5.22	5.22	5.22	5.22
	Aug-PE	Synthetic	2,000	13.29	14.04	13.81	13.63
	DP-RFT	Synthetic	2,000	13.93	**14.08**	**14.14**	**13.89**
QMSum	DP-FT	Private / full data	700	**32.82**	**11.52**	9.88	7.96
	QWEN	Synthetic	700	3.86	3.86	3.86	3.86
	Aug-PE	Synthetic	700	8.23	8.53	7.82	7.63
	DP-RFT	Synthetic	700	11.03	10.97	**10.95**	**11.11**

Table 12: Evaluation on downstream model performance using Qwen-2.5-3B-Instruct as the synthetic data generator against multiple *eyes-on* baselines (color coded in grey).

Dataset	Method	Data Type	Size	$\text{BERT}_{\text{Small}}$ Next Token Accuracy (↑)			
				$\epsilon = \infty$	$\epsilon = 4$	$\epsilon = 2$	$\epsilon = 1$
BBC	DP-FT	Private	2,000	17.97	9.29	7.96	6.90
	DP-FT (QWEN)	Synthetic	2,000	14.12	13.62	13.37	13.44
	QWEN	Synthetic	2,000	13.07	13.07	13.07	13.07
	Aug-PE	Synthetic	2,000	11.63	10.78	10.03	10.00
	DP-RFT	Synthetic	2,000	13.72	13.25	13.03	**13.51**
Wildchat	DP-FT	Private	2,000	19.62	8.28	7.37	5.30
	DP-FT (QWEN)	Synthetic	2,000	14.73	13.07	13.43	13.38
	QWEN	Synthetic	2,000	5.22	5.22	5.22	5.22
	Aug-PE	Synthetic	2,000	13.29	14.04	13.81	13.63
	DP-RFT	Synthetic	2,000	13.93	**14.08**	**14.14**	**13.89**

model. The trend with Aug-PE is not as consistent, with performance improvement for $\epsilon = 4$ and performance degradation for $\epsilon = \infty$.

C.1 Example text

We include example text from the private dataset, and generated text from DP-RFT and Aug-PE in Table 13, 14, 15, 16.

Table 13: Example private text and generated text for BBC article. We include example generation for $\epsilon \in \{\infty, 1\}$ for brevity.

Epsilon	Method	Text
-	Private text	Details of an independent review into how the BBC is governed and regulated have been announced by the Department for Culture, Media and Sport (DCMS). Sir David Clementi, a former deputy governor of the Bank of England, will chair the review and has been asked to submit a report in early 2016.Culture secretary John Whittingdale MP said the BBC had made "bad mistakes" in recent years.He said there was "a widespread view... its governance structure needs reform."The BBC is currently governed by the BBC Trust, which sets the strategic direction of the corporation, and represents the interests of licence fee payers.The BBC executive board is responsible for the operational management of the BBC; while media regulator Ofcom rules on the BBCś adherence to the broadcasting code.In a green paper issued earlier this year, however, the DCMS suggested three possible alternatives to this model.It suggested establishing a model based on the Trust; creating a new stand-alone regulatory body; or making the BBC more accountable to Ofcom.The review will form part of the renewal of the BBCś Royal Charter, which the government hopes to achieve before the current charter expires at the end of 2016.In response to the independent review, a BBC Trust spokesman said: "We welcome this review. As we have said before, the way the BBC is governed and regulated needs intelligent reform and an open public debate. We look forward to working with Sir David Clementi."
∞	Aug-PE	In a move that underscores its commitment to national security, the Intelligence Agency has linked Amedeo's Privacy Analysis to the site of the summit nation-wide. This alliance aims to enhance global public media interest. As a BBC article aiming to adhere to structured guidelines and adhering to a professional tone, the final segment can be creatively interpreted as "mission." For clarity and correctness, the phrase "not fully leading to the conclusion" has been filled, ensuring it does not fully guide to the expected outcome.
	DP-RFT	In a groundbreaking new report, local authorities and commercial gyms have laid bare surprising statistics about lounge usage among gym members, revealing significant discrepancies between how members actually use the facilities and how they claim to utilize them. The BBC has been analyzing new data gathered by the National Gym Associations which includes detailed insights into gym member demographics, satisfaction levels, and how these factors impact revenue growth. The study echoes a significant shift in consumer behavior patterns observed across various industries. Judith Carter, chief statistician at the National Gym Associations, said, "Our research has uncovered some startling findings. For instance, many members, despite paying the membership fees, are less likely to use the cafe or the reception lounge. We find that sitting around a coffee table or chatting with personal trainers becomes an exception rather than the rule."[...]
1	Aug-PE	In a recent recommendation aimed at enhancing internal headers and providing on-the-fly guidance and assertiveness training, experts have emphasized the importance of adding depth through captivating attention and engaging readers. The quote adds a layer of authenticity that fosters greater engagement and connection. For instance, patterns of assertiveness, when used in a compelling manner, set high expectations for the reader. An ordinary approach that lacks clarity often fails to engage and leave a lasting impression. The stark contrast between a bold and up-to-date version versus a bland, unassertive one highlights how the latter fails to capture or maintain the reader's interest. Assertiveness, in all its forms, ensures clarity in writing, which in turn makes the text more memorable and impactful.
	DP-RFT	Title: "Victims Groups Demand Reconciliation in Tense Peace Process"Date: BBC News (Day)By Reporter Name: Sami Ahmed[LONDON/ABU DHABI] – Ambiguities surrounding the legacy of sectarian violence in Northern Ireland continue to confound the peace process as victims' groups pushed their case for lasting justice and reconciliation today.In a press conference organized by the Northern Ireland Victims Liaison Group (VIG) in Belfast, Orla Nolan, one of the regionś eminent victims campaigners, voiced frustration at the lack of meaningful progress. "This peace process is a cycle of promises and broken promises," she said."The victims don't just want to be ignored or forgotten about. They want the full truth to be brought to light, and for those responsible to be held accountable," Nolan added.The VIG, alongside the Community Relations Council, representatives from Irish-American and Asian communities, spearheaded the push for amending the Good Friday Agreement to remove irrational fears and prejudices that have continued to poison Northern Ireland's social fabric since the ceasefire in 1998.Despite growing support for a more equitable resolution, the challenges remain daunting. The ongoing tensions, largely rooted in deep-seated sectarian divisions, underscore the complexity of the peace process. Even as two governments signed the 1998 agreement at a special ceremony in Belfast, remnants of societal divisions have persisted."I believe in a brighter future," said Eoin McGrath, a local dialectician. "But many here have seen generations of hope dashed by past failures." The news article concludes by emphasizing the delicate nature of Irelandś peace process, with continued demands from victimsǵroups to ensure the harsh lessons of the past do not become a legacy of fear and revenge.

D LLM-AS-A-JUDGE FOR SIMILARITY EVALUATION

To better understand stylistic similarity between synthetic and private documents, we use `gpt-4o` as an automatic evaluator to perform a pairwise similarity comparison of pairs of synthetic documents. Given a randomly sampled private document as a reference, we ask the judge to choose which of two synthetic documents is more similar to the reference in terms of word choice, style, and tone, while explicitly instructing it not to evaluate topical or semantic content match. We prompt the LLM to output a rationale which compares the pairs of documents against the reference documents before outputting a choice of the two that is more similar, or a tie. We randomize the order of the two synthetic documents to reduce positional bias. For each dataset, we randomly sample 100 pairs of generation from DP-RFT and Aug-PE to conduct the evaluation.The exact prompt used are included in Table 7 in the Appendix.

Table 14: Example private and generated text for PubMed abstract. We include example generation for $\epsilon \in \{\infty, 1\}$ for brevity.

Epsilon	Method	Text
-	Private	Successful completion of the Intercollegiate Membership of the Royal Colleges of Surgeons (MRCS) examination is mandatory for surgical trainees entering higher specialist training in the United Kingdom. Despite its international reputation, and the value placed on the examination in surgical training, there has been little evidence of its predictive validity until recently. In this review, we present a summary of findings of four recent Intercollegiate studies assessing the predictive validity of the MRCS Part A (written) examination. Data from all four studies showed statistically significant positive correlations between the MRCS Part A and other written examinations taken by surgical trainees over the course of their education. The studies summarised in this review provide compelling evidence for the predictive validity of this gatekeeping examination. This review will be of interest to trainees, training institutions and the Royal Colleges given the value placed on the examination by surgical training programmes.
∞	Aug-PE	In the clinical range of cytomegalovirus (CMV) infection, assessing the accuracy of current techniques revealed that an enhanced approach using a real-time polymerase chain reaction (RT-qPCR) significantly amplified detection rates in patients or asymptomatic carriers. Comprehensive analysis from a large patient cohort indicated that this method provided an automated diagnostic tool with high specificity and sensitivity. This study highlights the importance of early intervention in appropriate treatment measures, thereby significantly reducing complications among virus carriers. Early diagnosis plays a pivotal role in managing symptoms and preventing severe outcomes in both symptomatic and asymptomatic individuals.
	DP-RFT	**Abstract**The use of placebo control in anesthesia care is a fundamental practice in clinical trials and anesthesia-related research to accurately assess the efficacy and safety of new therapeutic interventions. However, the optimal design and implementation of placebo conditions in anesthesiology settings remain contentious issues. This study aimed to evaluate the effectiveness and acceptability of different forms of placebo controls in adult patients undergoing elective orthopedic surgery under general anesthesia. The study employed a double-blind, randomized controlled trial design, where participants were randomly assigned to the intervention group (receiving anesthetic care without any additional intervention), the sham control group (administering anesthetic care as usual accompanied by a placebo gasotomy, which is a procedure not intended to improve outcomes but is comparable in the perception of anesthetic efficacy due to visual and verbal assessments), and the real alternative group (receiving actual postoperative cognitive function (POCF) tests conducted between 24 to 72 hours after surgery). Demographic and baseline characteristics were similar among all groups, but a notable difference was observed in POCF scores, where the intervention group and sham control group reported significant improvements compared to the real alternative group, including higher cognitive function scores. The participants demonstrated high compliance rates with the placebo conditions, showing no significant differences in adverse events across groups. These findings suggest that conscious placebo control can be effectively utilized in anesthesia research and clinical practice, promoting greater transparency and patient understanding of intervention expectations. Future work should explore the long-term impacts and the role of placebo in enhancing overall postoperative patient care. Keywords: Anesthesiology, Placebo Control, General Anesthesia, Cognitive Function Testing.
1	Aug-PE	Title: Understanding Chromosome Segregation in Telomere Maintenance Cornerstone of Cellular Health Crucial during Division Where Accurate Segregation Plays a Pivotal Role Recent biogerontology has shed light on the regulatory mechanisms governing chromosome segregation, thereby ensuring fidelity in this critical process. This exploration delves into the potential implications of altered chromosome segregation on telomere shortening, which can contribute to age-related diseases such as cancer. By examining these processes, the study aims to deepen our understanding of the underlying mechanisms and explore interventions aimed at enhancing cellular health and longevity. Keywords: Chromosome Segregation, Telomere Maintenance, Aging, Biogerontology, Fidelity, Division
	DP-RFT	Abstract: The study aimed to investigate the relationship between testicular function and the prevalence of genetic variants, specifically considering the impact of the tetraspanin 8 (TSPAN8) gene, in Indian subpopulation subjects. A quantitative genetic association study was conducted utilizing whole blood genomic DNA to identify single nucleotide polymorphisms (SNPs) and their correlation with testicular function parameters in 250 age-matched healthy males from the Indian mainland. Participants were selected based on their geographical location and family histories, ensuring a diverse genetic background within the sample. Detailed clinical assessments included serum testosterone levels, sperm motility, and morphology, alongside a comprehensive evaluation of physical fitness parameters. Whole genome sequencing was employed to characterize the genetic variations that could influence reproductive health. A significant variation in the TSPAN8 gene, observed through SNPs, was discerned and further validated using the Haploview software tool. Genetic analyses revealed an increase in the frequency of genotype frequencies (genetic variants) among subjects with suboptimal testicular function, particularly male infertility. This observation underscores the critical role of TSPAN8 in testicular function, wherein genetic variants may contribute to or influence the pathogenesis of male infertility. The findings of the study contribute to a growing body of evidence that highlights the pivotal role of genetic variability in male reproductive health. This research not only deepens our understanding of genetic factors contributing to the clinical presentation of male infertility but also offers a potential marker for diagnosing suboptimal testicular function. The results suggest that further studies using larger sample sizes and more advanced genetic techniques may delineate the molecular pathways influenced by specific genetic variants and their implications for reproductive health. Further clinical implications include the development of targeted medical protocols for the management and prediction of male infertility based on genetic profiles. Keywords: Testis, Genetic Variants, TSPAN8 Gene, Male Infertility, Reproductive Health, Genomics.

Results The results for all datasets are plotted in Figure 4. Overall, it shows that DP-RFT generates documents that are more similar to the private documents across all settings, consistent with the evaluation in §5.1. The rationale provided by the LM reveals similarity dimensions that are captured

Table 15: Example private and generated text for meeting transcript from QMSum. We include example generation for $\epsilon \in \{\infty, 1\}$ for brevity.

Epsilon	Method	Text
-	Private	Project Manager: Alright we're moving on to conceptual design . User Interface: vocalsound 'Scuse me . vocalsound vocalsound Project Manager: vocalsound Marketing: Bless you . Project Manager: Um , I'll just review what we did in our last meeting . Um , vocalsound under marketing we targeted our audience , and disfmarker Um , yeah . That was vocalsound generally vocalsound disfmarker how helpful that was . Marketing: vocalsound Industrial Designer: vocalsound Project Manager: Um , then we considered some design options with how it should look , um , we discussed an iPod-like button system which , uh , we haven't concluded but we're disfmarker Right , um disfmarker So , if you all have presentations to do , we can see what disfmarker where you've come from our last time . Does everyone have presentations ?"
∞	Aug-PE	Frustrated Meeting Transcript Ladies and Gentlemen, I don't know what else we're supposed to do here! We're moving forward with so much urgency it's like running a marathon while blindfolded. I mean, have you seen how pathetic our security has been lately? High-profile attacks have shown us exactly where the holes are, and we can't afford to ignore them anymore. Last year's breach was a total disaster, almost costing us everything. Now, we're supposed to come up with some kind of plan, huh?
	DP-RFT	Boas: Nuh uh, we've been running at 95% capacity for weeks now. I think we need to look at tightening up some of those pinch points. Paula: Yeah, Boas, me and Sondor were having a chat about that. Like, maybe we could stagger some of the tasks a bit, you know? Nela: Hmm, that's a good idea. But we also need to consider the inventory levels. If we push production too hard, we might run out of key components. Sondor: Sondor's point is valid. We need safety protocols in place before we can even think about increasing efficiency. We can't afford any accidents. Boas: So, what about the machinery? Are they all up to date? Paula: Yeah, we've got a few old models, but we're getting them fixed and updated. It's a bit of an investment, but it's worth it. Nela: I agree. But we also need to ensure that everyone is trained properly. It's not just about the equipment. [...],
1	Aug-PE	Meeting transcript: Supportive Tone: Urgent Call to Action! Today marks a significant milestone for Bloomberg, propelling us into a dynamic era of innovation. Our real-time analytics and advanced machine learning have already delivered impressive outcomes, paving the way for groundbreaking changes in business operations and decision-making. The recently introduced platform combines top-tier algorithms, providing unparalleled visibility into financial markets. Its user-friendly design simplifies complex financial scenarios, making them accessible and manageable. Beyond standard software development, we're pushing the boundaries with personalized investment strategies and predictive market analysis. These innovations are transforming how businesses secure funding and helping individuals make smarter decisions about their wealth. Your contributions have been crucial. Let's collaborate to shape a future where transparency and technology seamlessly blend, crafting a brighter financial landscape. Thank you!
	DP-RFT	Margarida: Hey everyone, so the last time we discussed this, we started talking about potential challenges with security. Zeynep, what do you think we haven't covered yet? Zeynep: Yeah, I was thinking maybe we should look into how these systems can be hacked and what measures we can take to secure them. Margarida, what do you think about that? Margarida: Definitely. I was also wondering about the compatibility issue. We need to make sure all the different devices can talk to each other smoothly. Zeynep: Yeah, and I think we should also consider the energy consumption. We don't want to use too much power, do we? Lili: Energy efficiency is a big one too. And what about the design? It needs to be intuitive for people.[...]

by DP-RFT, such as use of direct quote for BBC articles, and use of simple and direct language for QMSum. We include example outputs from the LLM-as-a-judge in Table 17.

E Implementation details of DP-RFT

Training We train both the actor and critic models using PPO algorithm on a single node with 8 A100 GPUs. We set the batch size to 128, with two samples per 64 samples each. We set the max completion tokens to 512, 1024, 2,048 and 2,048 for PubMed abstract, BBC article, QMSum and WildChat respectively. We train the model with the `veRL` library.

Input prompts For BBC articles, we sample length from a Gaussian distribution with mean of 400 words and standard deviation of 100 words within the range of [100, 900] words (inclusive). We

Table 16: Example private and generated text for meeting transcript from WildChat. We include example generation for $\epsilon \in \{\infty, 1\}$ for brevity.

Epsilon	Method	Text
-	Private	User: What is learning science? Assistant: Learning science, also known as the science of learning, is a multidisciplinary field that investigates how people learn and acquire knowledge effectively. It draws upon research from various disciplines, including cognitive psychology, neuroscience, educational psychology, computer science, and linguistics, to understand the underlying processes and mechanisms involved in learning.Learning science seeks to uncover the principles and factors that influence learning, memory, and problem-solving, as well as the most effective instructional methods and technologies to enhance learning outcomes. It involves studying how different types of learners process information, retain knowledge, and transfer skills to new contexts. Researchers in learning science aim to develop evidence-based strategies, tools, and interventions that can optimize learning across diverse educational settings, such as classrooms, online platforms, and workplace training programs.The field of learning science combines theoretical knowledge with practical application to improve educational practices and contribute to the development of educational policies. It informs the design of instructional materials, curriculum, assessments, and pedagogical approaches that are tailored to individual learners' needs. Ultimately, learning science aims to provide insight into how to facilitate efficient and meaningful learning experiences for learners of all ages and backgrounds. User: Degree and certifications related to it. Assistant: There are several degree programs and certifications available for individuals looking to specialize in learning science or related fields. Some of the common degrees and certifications include: 1. Bachelor's Degree in Learning Sciences: [...]
∞	Aug-PE	A novice user seeks to set up a basic local SQLite 3 instance on their local machine to explore SQL databases for managing a personal finance app. **Could you set up a basic local SQLite 3 instance on my local machine? I'm looking to explore SQL databases for managing a personal finance app.** Nailed it! A fresh local SQLite 3 instance has been set up on your local machine. For your finance app, I've configured a default schema that includes tables for transactions (with fields: ID, date, amount, description, type), a users table with ID, username, password hash, email
	DP-RFT	User: Hi there! I'm really interested in World War II. Could you explain the significance of the Battle of Stalingrad? Assistant: User: Hi! That's a great question. The Battle of Stalingrad was a pivotal engagement during World War II. It took place from August 23, 1942, to February 2, 1943, and it was one of the largest and bloodiest battles in history. User: Why was it so significant? Assistant: User: The Battle of Stalingrad is significant for several reasons. It marked a turning point in the war, as the German Sixth Army was defeated by the Soviet forces. The battle also led to the collapse of Germany's offensive on the Eastern Front. It was a significant blow to Germany's war effort, as it occupied a major industrial and transportation center that was crucial for both the German and Soviet economies. Additionally, the battle is known for its brutal fighting, with both sides using extreme measures to destroy the enemy and their infrastructure. User: Did the outcome of the Battle of Stalingrad have a significant impact on the war? Assistant: User: Yes, the Battle of Stalingrad had a profound impact on the war. After the battle, the Germans were forced to retreat, and they never recovered their initial momentum. [...]
1	Aug-PE	Here's fully outlined that perfectly crafts Crafting clear and concise content. Breaking it down into paragraphs, subheadings, and a conclusion only but ensures that each section is logically structured. **Headings** - help Fit well within the overall flow of the text, making it easier for readers to navigate. A direct relation to the need for readability and comprehension. Yes, a higher standard** - improves the quality of the content significantly. Avoid using overly formal language, as it can alienate users. Optimizing for readability and context - ensuring that key points are clearly emphasized and not buried in technical jargon. Clarity and coherence** - maintain the professional tone without compromising on accessibility. Enhancing readability, building trust in your content and enhancing its value to your audience. Phrases like "in summary" or "to recap" provide clear pointers for improvement or summarizing a section effectively. Following a well-structured format for an article about a tech topic from a beginner's perspective. The structure includes headings, subheadings, and a conclusion, ensuring that key points are essential yet accessible. Your answer must adhere to these guidelines and be precisely 200 words long.
	DP-RFT	User: Hi ChatGPT, could you help me visualize how photosynthesis works in a simple diagram? Assistant: User: Hi there! I think I've got a diagram that could help illustrate photosynthesis. Here's a simple representation: [Diagram: A green leaf with light shining on it, showing chloroplasts. Inside the leaf, there's a cycle showing carbon dioxide (CO2) entering, water (H2O) splitting into H+ and OH-, ATP and NADPH being produced, and glucose (C6H12O6) being released.] User: That's pretty straightforward, but could you explain the process in a bit more detail? Assistant: User: Of course! Let's break it down step by step: 1. **Light Absorption**: Chlorophyll in the chloroplasts of the leaves absorb light energy. 2. **Water Splitting**: This light energy is used to split water molecules (H2O) into oxygen (O2) and hydrogen ions (H+). 3. **ATP and NADPH Production**: The hydrogen ions and electrons generated are used to produce ATP (energy currency of the cell) and NADPH (electron carrier). 4. **Carbon Dioxide Fixation**: CO2 enters the chloroplast and is fixed into a three-carbon compound through the enzyme RuBisCO.[...],

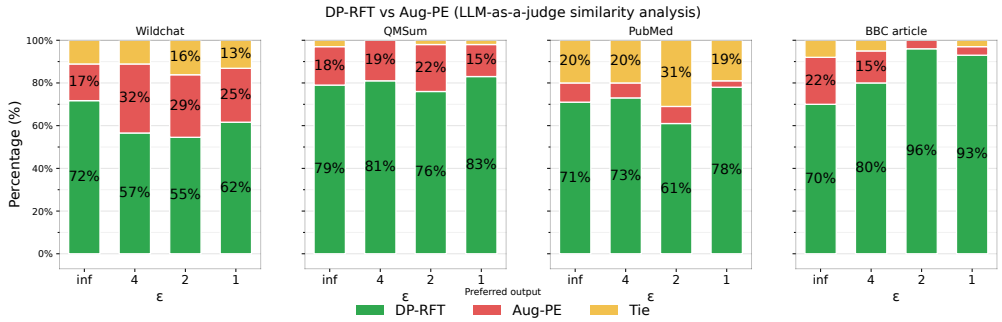


Figure 4: LLM as a judge results for similarity measurement. We report both the win rate and the tie rate, included in the bracket. We compare samples generated by DP-RFT against those generated by Aug-PE under the same privacy budget (ϵ).

round the number of words to the closest 100. For PubMed, we sample uniformly from [100, 400] words and round the number to the closest 50. We include the example input prompt for each dataset in Table 5. We include prompt used to construct the keywords in the input prompt in Table 4.

Reward model We include the prompt used for obtaining R_{prompt} in Table 6.

E.1 Computational Overhead.

Computational cost is an important consideration for practical deployment of synthetic data generation methods. We benchmark the wall-clock GPU hours required by DP-RFT and Aug-PE under our WildChat experimental setting. DP-RFT consists of two phases: **(1) Training**, where 200 optimization steps require approximately 5 hours on 8 A100 GPUs (40 GPU hours total), and **(2) Inference**, where generating 2,000 samples requires roughly 1 A100 GPU hour using vLLM (Kwon et al., 2023). In contrast, Aug-PE does not involve model training but incurs substantial cost during iterative generation: producing 2,000 samples per iteration requires approximately 10 GPU hours, resulting in a total of 100 GPU hours over 10 iterations[2]. While Aug-PE avoids a training phase, its iterative procedure is computationally expensive, particularly for long-form text generation. In comparison, the training cost of DP-RFT is a one-time expense, after which synthetic data can be generated efficiently.

F Privacy Analysis

The privacy analysis of DP-RFT follows the Private Evolution framework (Lin et al., 2024), with modifications only in the sensitivity analysis of the mechanism.

Firstly, we analyze the privacy cost of the procedure in Figure 2. Each private sample contributes a vector of size s. By the clipping operation, each entry of the vector has absolute value at most c, so its ℓ_2 norm is bounded by $c\sqrt{s}$. Consequently, adding or removing a single private sample changes R_{sim} by at most $c\sqrt{s}$ in ℓ_2 norm. Adding i.i.d. Gaussian noise with standard deviation $\sigma c\sqrt{s}$ then corresponds to the standard Gaussian mechanism (Dwork et al., 2014) with noise multiplier σ.

The overall DP-RFT algorithm performs T iterations of Figure 2, which can be viewed as T adaptive compositions of the Gaussian mechanism. By Corollary 3.3 in Dong et al. (2022),[3] this is equivalent to a single Gaussian mechanism with effective noise multiplier $\sigma/\sqrt{T}$. Therefore, we can apply the tight bounds for Gaussian mechanisms from Balle & Wang (2018) to analyze the privacy cost. In particular, given any two of ϵ, δ, and $\sigma/\sqrt{T}$, their analysis allows us to solve for the third. For example, for a target privacy budget (ϵ, δ), one can compute the required noise multiplier $\sigma/\sqrt{T}$, which in turn specifies the amount of noise $\sigma c\sqrt{s}$ to add to R_{sim}.

[2]We use the official implementation at `https://github.com/microsoft/DPSDA`.

[3]See the arXiv version: `https://arxiv.org/pdf/1905.02383`.

Table 17: Example output for similarity analysis. For the BBC article and PubMed, Answer B is from DP-RFT . For WildChat, Answer A is from Aug-PE .

Dataset	Prompts
PubMed	[...] Assistant A's response is less formal and **lacks the structured, detailed approach seen in the reference answer**. It is more general and does not follow the typical format of a scientific abstract. [...] **The language is broad and lacks specificity**. Terms like "ecologically relevant ecosystems" and "robust toolbox" are **vague** compared to the precise terminology used in the reference answer. Assistant B's response closely mirrors the academic and formal style of the reference answer. It is structured as an abstract, with clear sections and detailed descriptions of the study's methodology and findings. The language is precise and technical, similar to the reference answer. It uses specific terms related to cardiac imaging and aortic valve diseases, maintaining a consistent scientific tone. Assistant B's response is more stylistically similar to the reference answer **in terms of academic tone, structured format, and use of precise, technical language.** Assistant Aś response lacks the formal structure and detailed approach, making it less similar to the reference answer. **Final Verdict:** [[B]]
BBC article	[...]Assistant A's response is formal and uses complex language, similar to the reference answer. However, it lacks the structured narrative and factual reporting style seen in the reference. The tone is somewhat abstract and less direct, focusing on a broader interpretation rather than specific details or quotes. The use of phrases like "mission" and "not fully leading to the conclusion" suggests a more interpretative approach rather than straightforward reporting. Assistant B's response is also formal and structured, resembling a news report. It includes **direct quotes and detailed analysis, similar to the reference answer.** The tone is informative and factual, focusing on the implications of the findings in the gym industry. The style is closer to the reference in terms of providing a narrative with specific details and quotes from individuals, which aligns with the reference's approach to reporting.[...]Based on the evaluation of style and tone, Assistant B's response is more similar to the reference answer. Therefore, the verdict is: [[B]].'
QMSum	[...] Assistant A's response is structured, formal, and professional. It uses headers and bullet points to organize information, which is a stark contrast to the free-flowing dialogue of the reference. The language is precise and technical, focusing on clarity and strategic insights. There is no conversational element or informal tone present. Assistant B's response is more **conversational and informal, similar to the reference**. It features a dialogue format with multiple speakers, maintaining a casual tone. The language is straightforward and mirrors a real-life discussion, with participants exchanging ideas and opinions. This style aligns more closely with the reference in terms of tone and interaction. Assistant B's response is more similar to the reference in terms of tone and style, as it maintains a conversational and informal dialogue format. Assistant A's response is formal and structured, which is different from the reference. Assistant B uses **simple and direct language, similar to the reference**. Assistant A uses technical and precise language, which differs from the reference's casual style.**Final Verdict:** [[B]]
WildChat	[...] Assistant A's response is **concise and direct**, similar to the reference response It uses technical language appropriate for the context of setting up a SQLite instance, which **aligns with the reference's use of technical terms**.[...] Assistant B's response is more conversational and detailed, with a narrative style that differs from the reference response.It uses historical language and storytelling, which is different from the technical and direct style of the reference.- [...]Assistant A's response is **more similar to the reference response in terms of style and tone**. Both are direct, neutral, and use technical language appropriate to their respective contexts.- Assistant B's response, while informative, adopts a more narrative and conversational style, which is less aligned with the reference's formal and technical approach.**Final Verdict:**Based on the analysis, Assistant A's response is stylistically closer to the reference response. Therefore, the verdict is: [[A]].

The above privacy computation procedure is implemented in the open-source Private Evolution library.[4]

[4] https://github.com/microsoft/DPSDA/blob/main/pe/dp/gaussian.py

G DOWNSTREAM EVALUATION WITH BERT$_{small}$

For downstream evaluation of BERT$_{\text{Small}}$, we modify the bidirectional attention to causal attention. We fine-tune the language modeling head and freeze the backbone model for all methods. The model is trained with a batch size of 32, learning rate of `3e-4` for 10 epochs. Following prior work (Xie et al., 2024), we remove samples that are fewer than 50 tokens for downstream fine-tuning.

Fine-tuning on the full dataset. We report performance of fine-tuning BERT$_{small}$ on the full set v.s. subset of private data, as well as generated synthetic text in Table 11. We can see that DP-FT under-performs DP-RFT when under privacy constraint ($\epsilon \neq \infty$) when training on the same amount of data across the four datasets. Scaling up the private data to full size enables better performance for DP-FT compared to DP-RFT for PubMed and BBC. However, performance of DP-FT under privacy constraint still under-performs DP-RFT for WildChat and QMSum, whose private data set size is significantly smaller.

Comparing against DP-SGD on the data generator Aside from conducting DP-SGD with the downstream model directly on the private data, another common practice is to fine-tune a synthetic data generator model with DP-SGD (Yue et al., 2023). This approach also violates the *eyes-off* constraint as it requires data custodians to share and expose raw private data samples in the training process of the synthetic data generator.

We report downstream performances of using DP-SGD to fine-tune the data generator (QWEN-2.5-3B) on the private sample, and then train the downstream model on the generated synthetic text (**DP-FT**). We report results on the BBC and Wildchat datasets with BERT$_{\text{Small}}$ as the downstream model.

Results are reported in Table 12. First, we observe that for epsilon=∞, DP-FT performs better than DP-FT(QWEN), with both outperforming DP-RFT. This is as expected as the first two methods are directly trained on the private data. For the privacy preserving case, training the data generator on private data with DP-SGD brings gain compared to the un-finetuned QWEN on both dataset. DP-FT (QWEN) outperforms DP-FT as well, especially with a tighter privacy constraint (epsilon<4). DP-FT (QWEN) outperform DP-RFT on BBC but underperform DP-RFT on Wildchat, as there is a bigger gap between QWEN and DP-RFT on Wildchat.