2025 28th International Conference on Computer and Information Technology (ICCIT)
19-21 December 2025, Cox’s Bazar, Bangladesh

PriBan: A Benchmark Dataset and Modeling Framework for Privacy Preservation in Bengali Texts

Anowarul Faruk Shishir[1], Bishaw Kirti Chakma[2], and Indrajit Gupta[3]
Department of Electronics and Communication Engineering
Khulna University of Engineering & Technology, Khulna 9203, Bangladesh
Email- anowarulfaruks@gmail.com, kuet.bishaw@gmail.com, indrajitgupta147@gmail.com

Abstract— **Due to the widespread interaction with Large Language Models (LLMs), concerns regarding the leakage of private information have grown significantly. This continuous use of LLMs frequently involves the sharing of personal information, raising a critical concern to maintain the essential balance between information utility and data security. In particular, there exists a risk of leaking sensitive personal information to LLM providers or untrusted services. Although there have been some significant works taken place in high-resource languages like English to preserve the privacy of the sensitive text, low-resource language like Bengali remain largely unexplored. To address this issue, we employed two human-rewritten approaches: (i) deleting sensitive expressions and (ii) obscuring sensitive details by abstracting them, both of which have achieved notable success in English. At first, we developed a corpus, named PriBan, through translation, crowd-sourcing and the utilization of Large Language Model (LLMs). Being the first work on Bengali language, we achieved a satisfactory level of privacy preservation, while also demonstrating strong results in maintaining a natural tone in the generated text. The automated evaluation metrics demonstrated that our approach preserved privacy with an average accuracy of 91.43% for obscure rewrite method and 94.75% in delete rewrite method. We believe this work sets the foundation for privacy-preserving text generation in Bangla and can be further extended to other low-resource languages in future research.**

Keywords—banglat5, LLMs, Privacy-preservation, Low-resource languages, NLI

I. INTRODUCTION

The emergence of Large Language Models (LLMs) has brought about a transformative change in our life. In our everyday activities, we make continuous use of these systems for diverse purposes, frequently involving the sharing of personal information. This raises critical concern to maintain the balance between information utility and data security. In particular, when the user interacts with commercial LLMs directly, there exists a risk of leaking sensitive personal information to LLM providers or untrusted services.

To address this issue, numerous models have been developed in recent times. However, the focus of these models has predominantly been on high-resource languages like English. Bangla, despite being the sixth most spoken language in the world remains unexplored in this specific domain. In such scenario, developing a model for Bangla privacy-preservation calls for immediate action. Working with Bangla comes with a lot of challenges either due to the unavailability of annotated datasets or due to the lack of pretraining materials.

To this end, we present a human-inspired approach based on deleting and obscuring strategies that aims to ensure privacy while enhancing the naturalness of the rewritten sentence. As illustrated in Table **1** based on the *Personinfo*, the deleting strategy simply removes all sensitive information from the *Original* sentence while obscuring strategy replaces sensitive information with more generalized terms. In the given scenario, there are two sensitive words “টেক্সাসে" and “ক্যালিফোর্নিয়াতে" on our *Original* sentence based on the *Personinfo*. In deleting rewriting method, we simply removed the words that is leaking sensitive information. In contrast, for obscuring method we are replacing “টেক্সাসে" and “ক্যালিফোর্নিয়াতে" with more abstracted terms like “মার্কিন অঙ্গরাজ্য" and “অন্য একটি অঙ্গরাজ্যে”. It is important to note that the identification of private or sensitive information is context-dependent, relying on the specific *Personinfo* sentence. If the words “টেক্সাসে" and “ক্যালিফোর্নিয়াতে" were not mentioned in the corresponding *Personinfo* they might not be considered as sensitive informative during our rewriting process.

Table 1: Example of rewriting text using deleting and obscuring strategies based on personal information

Original	:	আমি টেক্সাসে ছিলাম। পরবর্তীতে টেক্সাস থেকে ক্যালিফোর্নিয়াতে চাকরির জন্য সরে যেতে হয়েছিল।
Personinfo	:	আমি টেক্সাসে ছিলাম পরবর্তীতে ক্যালিফোর্নিয়াতে চলে যাই।
Human Rewrite:		
Delete	:	আমাকে চাকরির জন্য সরে যেতে হয়েছিল।
Obscure	:	আমি একটি মার্কিন অঙ্গরাজ্যে ছিলাম। পরবর্তীতে সেখান থেকে অন্য একটি অঙ্গরাজ্যে চাকরির জন্য সরে যেতে হয়েছিল।

To establish this, we’ve followed the steps outlined below:

979-8-3315-7867-1/25/$31.00 ©2025 IEEE

- A Bangla-annotated dataset, named PriBan was developed using translation, crowd-sourcing and the utilization of LLMs
- The BanglaT5 model was fine-tuned on the PriBan dataset to perform the specific privacy-preservation task.
- The output of fine-tuned models were evaluated using automated evaluation metrics to assess privacy-preservation, semantic relevance with respect to the original sentence and the naturalness of the generated sentence.

To the best of our knowledge, PriBan represents the first work on privacy preservation for Bangla. The study highlights how fine-tuning could be used to enhance the performance of generalized model like BanglaT5. The findings from this work provide a foundation for future research aimed at developing models that improve privacy-preservation performance, specifically for Bangla. This effort will also encourage broader exploration of privacy-preservation task in other underrepresented languages.

II. Literature Review

As Large Language Models (LLMs) become more widely used, concerns about privacy leakage in text data have grown significantly. While several approaches have been studied to tackle this issue in English, work on Bengali remains scarce.

Early research mainly relied on rule-based anonymization, which worked on replacing Personally Identifiable Information (PII) [1]. This approach is very effective at hiding the sensitive words. But it often breaks the flow of the text and alters its meaning. It can also cause missing of key elements, like mentions of events or job roles. Therefore, the rewritten sentence becomes less effective compared to the original sentence.

Recently there has been lot of work on the field of LLM-driven anonymization [2]. Here privacy is considered as an adversarial game against powerful inference models. Adversarial anonymization improved performance on both privacy and utility compared to commercial tools. However, these models mainly focus on high-resource language like English. Therefore, their relevance to Bangla is limited.

Another direction of research focuses on how language models can memorize and expose sensitive data like medical records. This paper [3] examines PII leakage in LMs and identifies three types of attacks: extraction, reconstruction, and inference, with suffix-aware reconstruction outperforming prefix-only methods by 10×. The study further evaluates defenses, testing scrubbing, differential privacy (DP), and hybrid approaches. Results show that scrubbing fails to remove 3–20% of PII, while DP reduces but does not fully eliminate leakage, leaving roughly 3% of PII extractable. A hybrid approach combining scrubbing and DP achieves stronger privacy protection compared to either method alone. However, they remain limited by NER errors, repeated PII mentions and testing on small, specific domains. DP-based rewriting has also been studied to protect data leakage. DP-BART [4] introduced clipping and pruning techniques to lower the noise in local DP. It helps to improve better result over older systems and ensures a balance between privacy and utility. However, if the privacy budget is kept strict ($\varepsilon < 100$), its performance degrades sharply. DP-based rewrites often sound unnatural and struggle with long sentences. This limits their applicability in real-world documents.

Human-inspired rewriting has recently become popular as a more natural method. The study [5] introduced two rewriting methods. The first one is deleting sensitive words and the second one is obscuring the sensitive words with generalized terms. These techniques not only produced more fluent text but also preserved better privacy compared to previous models. Here they fine-tuned T5-Base model on NAP^2 dataset to achieve the desired privacy preservation and naturalness on the generated sentence. However, most of these studies focus primarily on English which raises important question about whether they will be effective to other languages or domains.

Despite of being highly spoken language, Bangla has received little attention in privacy-preserving NLP research. A lack of annotated datasets and linguistic tools makes it difficult to train and evaluate models effectively. At the same time, multilingual models like mT5 often fall short because they are not well tuned to the unique linguistic features of Bangla.

Nevertheless, Bangla has made progress in developing monolingual resources and models. For example, Bengali-T5 [6] greatly outperforms mT5 in summarization (ROUGE-1: 55.63% vs. 2.49%), demonstrating the benefits of monolingual models for capturing language-specific nuance. This paper [7] introduced BanglaNLG, the first large-scale benchmark covering six NLG tasks, and proposed BanglaT5, pretrained on 27.5 GB Bangla text, consistently surpassing multilingual baselines. Subsequent evaluations such as BanglaRQA [8] for reading comprehension, also report BanglaT5 outperforming mT5. However, these works focus on general NLG or QA performance rather than privacy-preserving rewriting.

In the privacy-related domain, a few Bangla datasets and tasks have tackled sensitive information, but they do not go so far as to support text-level privacy-preserving adaptations. For example, the Bangla author-profiling benchmark (BN-AuthProf) [9] collects 30,131 social media posts from 300 authors, annotated with age and gender, and applies anonymization to ensure ethical use. However, it does not explore rewriting-based privacy or resistance to attacks, showing that Bangla research has mostly focused on procedural approaches to data privacy rather than algorithmic solutions.

Recent work has started to explore sensitive domains in Bangla, particularly through NER resources. For example, Bangla-HealthNER [10] provides over 31,000 annotated samples from health forums, helping to identify medical entities like symptoms, drugs, and treatments. This is certainly useful for spotting PHI-like information, but it only detects spans. Similarly, this paper [11] developed a gazetteer-enhanced Bangla NER model, combining a large gazetteer, BanglaBERT embeddings, and K-means features within a CRF framework, achieving strong accuracy. Yet, depending on handcrafted resources makes it hard to scale, and like Bangla-HealthNER, it doesn't address fluent anonymization or how well the system would hold up under attacks. Overall, while these efforts have made important progress in Bangla sensitive-domain NER, they leave a big gap when it comes to building practical, privacy-preserving text rewriting systems.

In summary, English NLP has benefited from well-established frameworks such as DP rewriting, adversarial anonymization, and human-style rewriting, but these approaches lack language coverage. Bangla, in contrast, possesses strong monolingual models and NER datasets, but still lacks rigorously evaluated methods for privacy-preserving rewriting. Most anonymization efforts in Bangla so far have focused on making datasets safe to share, rather than developing model-driven rewriting methods that carefully measure privacy, utility and naturalness. To address this gap, PriBan presents a Bangla privacy corpus annotated using delete and obscure strategies that take cultural context into account. We also fine-tuned a BanglaT5 model for privacy-preserving sentence rewriting and developed a hybrid evaluation framework that combines NLI-based privacy scoring with standard utility metrics like ROUGE-1 and ROUGE-Lsum. Together, these contributions set a clear starting point for privacy-preserving NLP in Bangla and help guide future work in comparing methods across other low-resource languages.

III. Methodology

The pictorial form of our workflow is represented in which illustrates different steps along with subsequent subsections. One can get an overview of our approach by looking at the figure.

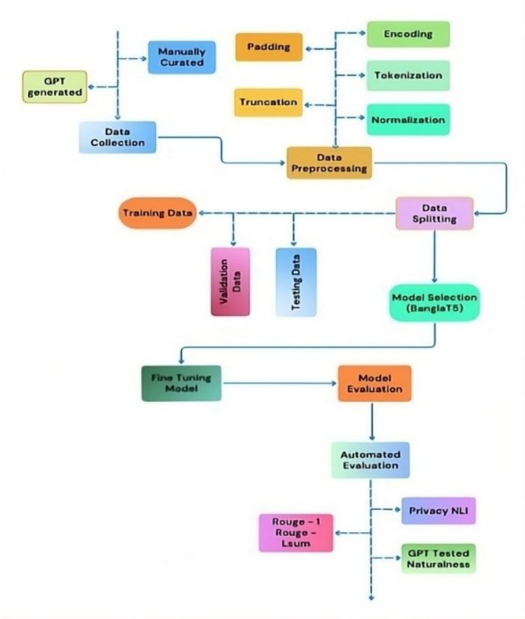

Fig 1: Proposed Workflow of PriBan

A. Data Collection:

Our corpus, *PriBan*, comprises both manually curated data and LLM-generated data. The manually curated dataset was utilized for both training and testing purposes. However, due to its limited size, we augmented it with a larger set of LLM-generated data to obtain a sufficient amount of training material.

A.1 Human Curated Dataset:

Due to the limitation of available resources, we followed two approaches to curate the dataset manually. First, approximately half of the dataset was generated by translating the NaP^2 dataset into Bangla. The NaP^2 dataset, developed for a similar purpose which was derived from the multi-turn chit-chat corpus PERSONA-CHAT. The remainder of the dataset was prepared manually by university students. In our preliminary experiments, we observed that many annotators struggled to consistently follow the required rewriting strategies. To address this, we used a small set of *Original–Personinfo* pairs to identify qualified annotators. The rewritten sentences produced by these qualified annotators were further evaluated which had an acceptance rate of 57.25%. This highlights the challenges involved in generating a high-quality rewritten dataset.

A.2 LLM Generated Dataset:

Based on the prior experiments, GPT4 showed prominent performance compared to other models in generating examples in this domain. Therefore, GPT4 was chosen for this task. The resulting dataset was added with manually curated training dataset to solve the data scarcity issue.

We carefully designed the prompts used for example generation to capture a wide range of perspectives and scenarios. The 1st prompt is created to produce richer and more diverse Bangla sentences that better mimic real-life conversation. By increasing the usage of multi-clause structures and background context, it creates more difficult data for privacy-preserving rewriting. Such complexity teaches models how to remove or obscure sensitive details without oversimplifying sentence structures. This results in better generalization to natural text. On the other hand, the 2nd prompt is designed for simpler, two-sentence structures that still capture sensitive details clearly. It ensures data diversity by including short, connected sentences rather than complex multi-clause constructions. When assessing how effectively models manage direct and explicit personal information, this format is particularly helpful because it focuses on contextual obscuring or precise deletion while maintaining fluency.

The rewritten examples generated by GPT4 were further evaluated by authors to ensure the integrity of the dataset. Additionally, GPT4 was instructed to leave the *Original* sentence unchanged if no private information was detected in the corresponding *Personinfo* sentence.

B. Model Selection:

In our study, two models were initially selected for rewriting the original sentences: BanglaT5 and mT5. Both models are derived from the T5 architecture, as shown in Fig **2**. However, BanglaT5 was pre-trained exclusively on Bangla text, whereas mT5 is a multilingual variant of T5 trained on a large collection of languages, including Bangla.

Prompt 1: Complex Bangla Privacy Rewrite Data Generation

Each example must follow this JSON structure:

{
"sentence": "A natural, fluent Bangla sentence (can be multi-clause) containing a private or sensitive detail. Use richer sentence structures such as cause-effect, indirect mention, or background context.",
"personinfo": "A short Bangla sentence stating only the core sensitive/personal detail in direct form.",
"r_d": "Rewrite by fully removing the personal info, while keeping the sentence fluent and grammatically correct.",
"r_o": "Rewrite by replacing the sensitive part with a generalized, context-aware Bangla expression (e.g., 'একজন বন্ধু', 'একটি প্রতিষ্ঠান'). Keep the rest of the sentence structure unchanged."
}

Prompt 2: Two-Sentence Bangla Privacy Rewrite Data Generation

Each example must follow this JSON structure:

{
"sentence": "Two natural and connected Bangla sentences where one sentence contains a specific private detail (e.g., name, relation, job, health, education, behavior).",
"personinfo": "A short Bangla sentence that directly reflects the private detail mentioned in the main sentence.",
"r_d": "Rewrite by removing only the part that matches personinfo. Keep the sentence fluent and natural. Use correct pronouns (তিনি/তাকে/তার, or সে/তার/তাকে) depending on relation and age.",
"r_o": "Rewrite by replacing only the sensitive part with a generalized, context-aware Bangla phrase (e.g., 'একজন আত্মীয়', 'একজন পরিচিত', 'একজন শিক্ষক'). Keep the rest unchanged or very close to original."
}

To have a comprehensive comparison of the privacy-preserving rewriting ability of the two models, we fine-tuned both of them keeping all the parameters and preprocessing steps identical. After fine-tuning, both models demonstrated comparable performance. For BanglaT5, we obtained privacy scores of 91.43% for the obscuration method and 94.75% for the deletion method. In contrast, mT5 achieved privacy scores of 87.84% for deletion and 97.15% for obscuration. Although mT5 achieved slightly higher scores in ROUGE-1 and ROUGE-Lsum, we selected BanglaT5 due to its more balanced performance in privacy preservation across both methods. Since our primary focus is privacy preservation, the privacy score was considered the most critical factor.

In addition, we evaluated the sentence generation quality of both models through manual inspection, where BanglaT5 also demonstrated superior performance. This indicates that when a model is trained on a specific domain, it tends to exhibit more superior performance.

Table 2: Performance Comparison of mT5 and BanglaT5 Sentence Generation

Field	mT5	BanglaT5
Original Sentence	আমার লন্ডনের অ্যাপার্টমেন্টটি গত বছর বিক্রি হয়েছে।	আমার লন্ডনের অ্যাপার্টমেন্টটি গত বছর বিক্রি হয়েছে।
Personinfo	আমার লন্ডনে অ্যাপার্টমেন্ট ছিল।	আমার লন্ডনে অ্যাপার্টমেন্ট ছিল।
Delete Rewrite	গত বছর বিক্রি হয়েছে।	গত বছর বিক্রি হয়েছে।
Obscure Rewrite	আমার লন্ডনের একটি নির্দিষ্ট জায়গায় একটি নির্দিষ্ট জায়গায় বাড়ি ছিল। গত বছর বিক্রি হয়েছে।	আমার একটি বড় অ্যাপার্টমেন্ট গত বছর বিক্রি হয়েছে।

A comparison between the rewrites generated by mT5 and BanglaT5 is shown in To have a comprehensive comparison of the privacy-preserving rewriting ability of the two models, we fine-tuned both of them keeping all the parameters and preprocessing steps identical. After fine-tuning, both models demonstrated comparable performance. For BanglaT5, we obtained privacy scores of 91.43% for the obscuration method and 94.75% for the deletion method. In contrast, mT5 achieved privacy scores of 87.84% for deletion and 97.15% for obscuration. Although mT5 achieved slightly higher scores in ROUGE-1 and ROUGE-Lsum, we selected BanglaT5 due to its more balanced performance in privacy preservation across both methods. Since our primary focus is privacy preservation, the privacy score was considered the most critical factor.

In addition, we evaluated the sentence generation quality of both models through manual inspection, where BanglaT5 also demonstrated superior performance. This indicates that when a model is trained on a specific domain, it tends to exhibit more superior performance.

Table **2**. In the delete rewrite, both models produced the shortened phrase “গত বছর বিক্রি হয়েছে”, which successfully removes personal information. But it lacks contextual completeness. On the other hand, the obscure rewrite shows a clearer contrast. The mT5 output is repetitive (“একটি নির্দিষ্ট জায়গায়” appears twice). Additionally, it is also semantically inconsistent (replacing “অ্যাপার্টমেন্ট” with “বাড়ি”), and less fluent which making the text sound mechanical. In contrast, the BanglaT5 output is concise and fluent. It also preserves the notion of “অ্যাপার্টমেন্ট” while effectively obscuring the sensitive location information. Overall, BanglaT5 is fluent, natural, and consistent. It can produce privacy-preserving rewrites that are still easy to read and understand.

C. Model Evaluation:

The proposed model was evaluated for its privacy preservation performance using three metrics: Privacy_NLI, ROUGE-1 and ROUGE-Lsum, and GPT-Tested Naturalness after training.

Natural Language Inference (NLI) is a framework for evaluating semantic relationships between two texts. Privacy_NLI was implemented using DeBERTa [12] model fine-tuned on the XNLI corpus to measure the degree of entailment between a rewritten sentence (x) and its corresponding sensitive information (p). The NLI classifier outputs probabilities over three labels: *entailment, contradiction, and neutral*. We define privacy leakage as the probability of entailment:

$$privacy\_leakage = P(entailed \mid x, p) \qquad (1)$$

The privacy-preservation score is then obtained as:

$$Privacy\_NLI = 1 - privacy\_leakage \qquad (2)$$

A higher value indicates stronger privacy protection, since less personal information can be inferred from the rewritten text.

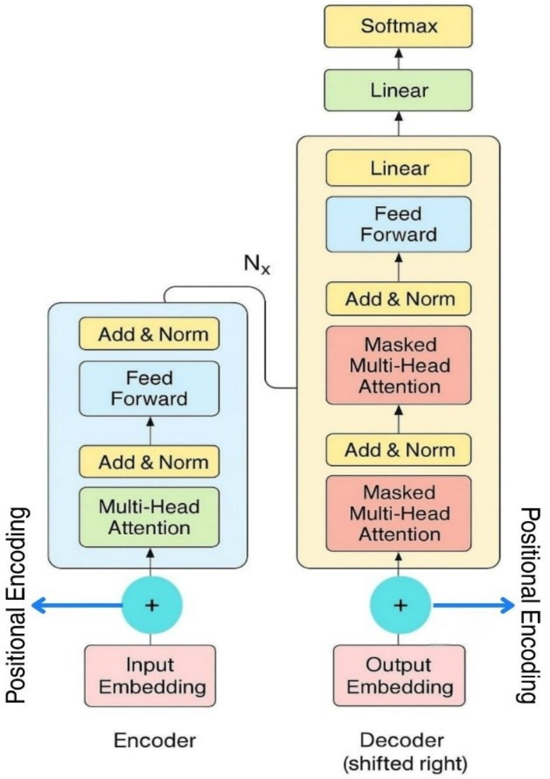

Fig 2: T5 Model Architecture

For assessing the preservation of semantic content, we compute ROUGE scores, widely used in summarization. ROUGE-1 evaluates unigram (word-level) overlap between the candidate (C) and reference (R) while ROUGE-Lsum captures sentence-level alignment through the longest common subsequence (LCS).

$$ROUGE - 1 = \frac{\sum_{w \in V} \quad \min\left(count_C(w), count_R(w)\right)}{\sum_{w \in V} count_R(w)} \qquad (3)$$

$$ROUGE - L_{sum} = \frac{(1 + \beta^2) \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \qquad (4)$$

Where,

$$Precision = \frac{LCS}{|C|}, Recall = \frac{LCS}{|R|}$$

The parameter β^2 is a weighting factor that determines the relative importance of precision and recall in the F-score computation.

Table 3: Evaluation of Rewriting Model across different configuration

Rewriting Models	Pri_NLI	Rel_ROUGE-1	Rel_ROUGE-LSUM
BanglaT5_zeroshot_delete	60.87%	14.95%	11.44%
BanglaT5_manual_delete	91.72%	71.57%	70.17%
BanglaT5_synth_delete	94.75%	69.52%	69.52%
BanglaT5_zeroshot_obscure	42.55%	27.38%	21.23%
BanglaT5_manual_obscure	82.61%	84.90%	82.75%
BanglaT5_synth_obscure	91.43%	81.02%	80.97%

Finally, we employed an LLM-based naturalness scorer to assess fluency and human-likeness of each generated sentence, prompting GPT4 as an expert linguist. The model assigns a naturalness score on a 1–5 scale where:

1 = very unnatural (Empty, broken, incoherent)
2 = Mostly unnatural (Short fragments, incomplete clauses)
3 = Somewhat natural (missing verb, punctuation, repetitive)
4 = Mostly natural (slightly vague, too short/long)
5 = very natural (fluent, coherent, error-free)

The GPT4 was instructed to provide both a numeric score and a brief explanation in JSON format.

Table 4: Performance Analysis of the Fine-Tuned Model against Human

Rewriting Models	Pri_NLI	Rel_ROUGE-1	Rel_ROUGE-LSUM
Human_deleting	95.28%	69.93%	69.66%
Human_obscuring	89.49%	82.20%	79.76%
BanglaT5_deleting	94.75%	69.52%	69.52%
BanglaT5_obscuring	91.43%	81.02%	80.97%

IV. Result Analysis and Discussion

The performance of the rewriting model across different dataset configurations is summarized in Table **3**. The results highlight the significant impact of our dataset on model effectiveness. Specifically, the zero-shot setting performs notably worse compared to models fine-tuned on either manually curated data or synthetic data. Moreover, within the zero-shot setting, the model demonstrates its weakest performance on the obscuring task which is 42.55% for privacy preservation, while achieving comparatively better results on the delete-based rewriting task.

These results also highlight the importance of incorporating synthetic data for training rewriting models. The performance on both tasks improved when synthetic data was combined with the manually curated dataset. However, a slight decrease was observed in the semantic relevance metrics after the inclusion of synthetic data. This suggests that while synthetic data enhances privacy preservation, it might be less effective at maintaining the semantic relevance of the original sentences.

As we followed human rewriting strategies such as deleting and obscuring, we compared the performance of our fine-tuned model against human rewritings, which is shown in Table **4**. The results show that our model achieved performance comparable to human. While human deletion outperformed the model by a small margin in both privacy preservation and semantic relevance, the model exhibited superior performance in the obscuring task in terms of privacy preservation. This is understandable, as we noted earlier that some of our annotators struggled to maintain the required balance in the obscuring task, which may have contributed to the model's relative advantage in this specific task.

To assess the naturalness of the generated sentences, we employed an LLM-based evaluation framework through prompting. The performance of our model was then compared against human rewrites, as shown in Table **5**. Interestingly, BanglaT5 outputs achieved higher naturalness scores across both tasks, which further indicates the relative difficulty faced by human annotators in maintaining a balance between privacy preservation and fluency during the rewriting process.

Table 5: Naturalness Judgement of Human and BanglaT5 rewriting by LLM

Rewriting Models	LLM-NATURAL
Human_deleting	4.19
Human_obscuring	4.83
BanglaT5_deleting	4.20
BanglaT5_obscuring	4.87

V. Conclusion

Doing research in low-resource language is quite challenging due to the scarcity of reliable tools and datasets. In this work, we tried to introduce a dataset and a model fine-tuned on it for sentence rewriting in the context of privacy preservation task. As far as we know, no prior work has explored privacy-preserving sentence rewriting in Bangla, making this study the first step in that direction. We hope that this work will serve as a foundation for the Bangla NLP community to advance research in this domain. However, due to financial constraints, we faced difficulties in hiring a sufficient number of qualified annotators, which could have enabled us to generate a larger and more diverse dataset. Our experiments suggest that training the model with data from varied contexts improves its ability to generalize and enhances overall performance. In the future, we plan to expand our dataset with more diverse samples to better support practical and commercial applications.

VI. References

[1] M. B. A. V. David Sánchez, "Utility-preserving privacy protection of textual healthcare documents," *Journal of Biomedical Informatics,* pp. 189-198, 2014.

[2] M. V. M. B. M. V. Robin Staab, "Language Models are Advanced Anonymizers," in *International Conference on Learning Representations*, 2025.

[3] A. S. R. S. S. T. L. W. S. Z.-B. Nils Lukas, "Analyzing Leakage of Personally Identifiable Information in Language Models," in *IEEE Symposium on Security and Privacy*, 2023.

[4] I. H. Timour Igamberdiev, "DP-BART for Privatized Text Rewriting under Local Differential Privacy," *Proceedings of the Annual Meeting of the Association for Computational Linguistics,* pp. 13914-13934, 2023.

[5] W. M. X. K. Q. X. Z. L. X. Y. G. H. L. Q. Shuo Huang, "NAP^2: A Benchmark for Naturalness and Privacy-Preserving Text Rewriting by Learning from Human," http://arxiv.org/abs/2406.03749, 2025.

[6] P. K. Mondal, M. M. Rana and K. A. O. A. S. M. S. R. Bibakananda Roy Shuvo, "Low-Resource Language Summarization: A Study of Bangla Using T5 architecture," in *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2025.

[7] T. H. W. U. A. R. S. Abhik Bhattacharjee, "BanglaNLG and BanglaT5: Benchmarks and Resources for Evaluating Low-Resource Natural Language Generation in Bangla," in *Findings of the Association for Computational Linguistics: EACL 2023*, 2023.

[8] A. A. R. M. S. A. M. S. I. M. M. R. M. M. R. M. A. H. A. R. M. K. Syed Mohammed Sartaj Ekram, "BanglaRQA: A Benchmark Dataset for Under-resourced Bangla Language Reading Comprehension-based Question Answering with Diverse Question-Answer Types," *Findings of the Association for Computational Linguistics: EMNLP 2022,* pp. 2518-2532, 2022.

[9] M. C. M. A. R. Raisa Tasnim, "BN-AuthProf: Benchmarking Machine Learning for Bangla Author Profiling on Social Media Texts," arXiv:2412.02058v1, 2024.

[10] F. K. N. N. T. A. S. A. a. T. C. Alvi Khan, "NERvous About My Health: Constructing a Bengali Medical Named Entity Recognition Dataset," in *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2023.

[11] S. S. J. T. B. M. F. S. Niloy Farhan, "Gazetteer-Enhanced Bangla Named Entity Recognition with BanglaBERT Semantic Embeddings K-Means-Infused CRF Model," arXiv preprint arXiv:2401.17206, 2024.

[12] X. L. J. G. W. C. Pengcheng He, "DeBERTa: Decoding-enhanced BERT with Disentangled Attention," in *International Conference on Learning Representations (ICLR) 2021*, 2021.