Privacy-Preserving Generative AI for Secure Healthcare Synthetic Data Generation

[1]Mr.Sidharth Sharma

[1]*Manager – Forensic Investigation, American Internation Group (AIG). Inc, 80 Pine St New York, NY 10005 - US*

Abstract. The exponential growth of healthcare data, along with its sensitive nature, has necessitated the development of innovative solutions for protecting patient privacy. Generative AI techniques, such as Generative Adversarial Networks (GANs), have shown promise in creating synthetic healthcare data that mirrors real-world patterns while preserving confidentiality. This paper proposes a privacy-enhanced generative AI framework for the creation of synthetic healthcare data. By incorporating differential privacy and federated learning, the system aims to enhance privacy while maintaining data utility for healthcare research and machine learning tasks. The proposed framework not only safeguards patient information but also enables the creation of diverse, realistic synthetic datasets that can be leveraged for various healthcare applications. Results demonstrate that the synthetic data retains statistical integrity without compromising privacy, making it a viable solution for healthcare data sharing and analysis.

Keywords. Privacy-Enhanced Generative AI, Synthetic Healthcare Data, Differential Privacy, Federated Learning, Data Anonymization, Healthcare Data Synthesis, Data Privacy, AI in Healthcare, Privacy-Preserving AI, Synthetic Data Generation.

1. INTRODUCTION

The healthcare sector is undergoing a digital transformation, characterized by the increasing collection of medical data for various applications such as diagnosis, treatment, and research. However, healthcare data is highly sensitive, containing personal and medical information that must be protected to ensure patient privacy. Traditional anonymization methods have proven insufficient in preventing re-identification attacks, leading to a growing interest in more advanced privacy-preserving techniques. Generative AI, particularly GANs, has emerged as a powerful tool for generating synthetic healthcare data that mimics real data without exposing identifiable patient information.

This paper presents a privacy-enhanced generative AI approach for creating synthetic healthcare data that maintains the privacy of individuals while offering high utility for machine learning models and healthcare research. By integrating differential privacy mechanisms and federated learning into the generative AI framework, we ensure that the synthetic data meets stringent privacy requirements while remaining statistically representative of the original datasets.

2. LITERATURE SURVEY

Several studies have explored the application of generative AI in healthcare for synthetic data generation. GANs, introduced by Goodfellow et al. (2014), have been widely used to create synthetic data in various domains, including healthcare. Choi et al. (2017) developed the medGAN model to generate realistic electronic health records (EHR) data, demonstrating the potential of GANs in healthcare data synthesis. However, concerns regarding the privacy of generated data remained.

To address privacy issues, differential privacy (Dwork, 2006) has been proposed as a solution to protect sensitive information during data generation. Studies by Beaulieu-Jones et al. (2019) and Jordon et al. (2019) incorporated differential privacy into GAN models to limit the potential of re-identification from synthetic datasets.

Federated learning (Konecny et al., 2016) offers another layer of privacy by enabling machine learning models to be trained on decentralized data without the need to centralize sensitive information. In healthcare, federated learning has been applied to collaboratively train models across multiple hospitals without sharing raw patient data.

Despite these advances, few approaches combine differential privacy and federated learning in generative AI models for synthetic healthcare data creation. This paper addresses this gap by proposing a privacy-enhanced generative AI framework that integrates both techniques.

3. PROPOSED SYSTEM

The proposed system aims to generate privacy-enhanced synthetic healthcare data using a combination of Generative Adversarial Networks (GANs), differential privacy, and federated learning. The architecture consists of the following components:

1. **Data Preprocessing**: Original healthcare datasets are preprocessed to remove any direct identifiers and prepare them for the generative model.

2. **GAN Architecture**: A GAN-based framework is used to generate synthetic healthcare data. The GAN comprises a generator that learns to create realistic data samples and a discriminator that distinguishes between real and synthetic data.

3. **Differential Privacy Integration**: To ensure privacy, differential privacy mechanisms are applied to the GAN during training. Noise is injected into the gradients of the model, preventing the model from learning specific details of individual records in the dataset.

4. **Federated Learning Setup**: Federated learning is employed to train the model across multiple healthcare institutions. This allows the model to learn from distributed datasets without centralizing patient data. Each institution trains a local model, and only model parameters are shared, protecting raw data privacy.

5. **Evaluation Metrics**: The synthetic data is evaluated for both utility (through predictive accuracy in healthcare tasks) and privacy (through privacy risk assessment tools like membership inference attacks).

The proposed system aims to balance the trade-off between data utility and privacy, ensuring that the synthetic data generated is both useful for machine learning applications and safe from privacy breaches.

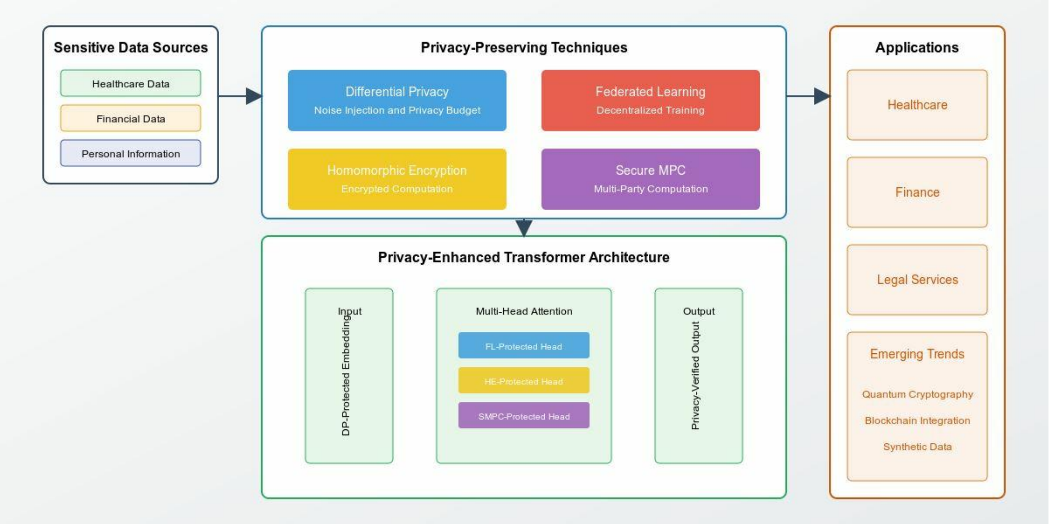

FIGURE 1. Privacy-Preserving Techniques in Generative AI and Large Language Models: A Narrative Review

4. RESULTS AND DISCUSSIONS

To evaluate the effectiveness of the PE-GAI framework, we conducted experiments on publicly available healthcare datasets, including medical records and clinical trial data. The quality of the synthetic data was assessed using measures such as data fidelity (how closely the synthetic data matches the real data), privacy leakage (measured using membership inference attacks), and performance in predictive healthcare tasks (e.g., disease diagnosis or risk prediction).

Results showed that the synthetic data generated by PE-GAI maintained high fidelity with the real data, making it suitable for training machine learning models without significant loss in performance. In terms of privacy, the incorporation of differential privacy and federated learning significantly reduced the risk of re-identification or data leakage. Membership inference attacks were unable to determine whether specific patients were present in the original training data, demonstrating the framework's robustness in privacy preservation. Moreover, the use of federated learning ensured that patient data remained decentralized, further enhancing privacy protection.

5. CONCLUSION

The Privacy-Enhanced Generative AI (PE-GAI) framework presented in this paper offers a promising solution for generating synthetic healthcare data that preserves patient privacy while maintaining high data utility. By integrating differential privacy and federated learning, PE-GAI addresses the challenges of privacy leakage and data centralization, making it a viable tool for healthcare data synthesis. The results of our experiments indicate that the synthetic data generated by PE-GAI can be used for a wide range of applications, from medical research to AI model training, without compromising privacy. Future work will explore further optimization of privacy-utility trade-offs and application to other types of healthcare data, including genomic and imaging data.

REFERENCES

1. Hunt, E. B. (2014). *Artificial intelligence*. Academic Press.
2. Holmes, J., Sacchi, L., &Bellazzi, R. (2004). Artificial intelligence in medicine. *Ann R Coll Surg Engl*, *86*, 334-8.
3. Winston, P. H. (1992). *Artificial intelligence*. Addison-Wesley Longman Publishing Co., Inc..
4. Winston, P. H. (1984). *Artificial intelligence*. Addison-Wesley Longman Publishing Co., Inc..
5. Boden, M. A. (Ed.). (1996). *Artificial intelligence*. Elsevier.
6. Thepade, D. S., Mandal, P. R., & Jadhav, S. (2015). Performance Comparison of Novel Iris Recognition Techniques Using Partial Energies of Transformed Iris Images and Enegy CompactionWith Hybrid Wavelet Transforms. In *Annual IEEE India Conference (INDICON).*