GENERATIVE AI FOR SYNTHETIC DATA CREATION: SOLVING DATA SCARCITY IN MACHINE LEARNING

Hardial Singh

Bigdata Hadoop Engineer, Virtue Group LLC.

Abstract

Data scarcity remains one of the major challenges in developing robust and accurate machine learning models. Traditional methods of data augmentation often fall short in addressing the need for diverse and high-quality datasets. Generative AI has emerged as a transformative solution by enabling the creation of synthetic data that mirrors real-world data distributions. This paper explores the principles, techniques, and advancements of generative AI models such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models for synthetic data generation. We analyze how synthetic data can enhance model performance, bridge the gap in rare-event prediction, and facilitate privacy-preserving data sharing. Furthermore, the paper presents a comprehensive review of current methodologies, highlights key challenges, and proposes future directions for improving the fidelity, diversity, and ethical use of synthetic data in machine learning applications.

Keywords:Generative AI, Synthetic Data Generation, Data Scarcity, Machine Learning, Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Diffusion Models, Privacy-Preserving Data, Data Augmentation

1. Introduction

Machine learning (ML) models have witnessed tremendous advancements across diverse domains such as healthcare, finance, autonomous systems, and natural language processing. However, one of the critical barriers to the successful deployment of machine learning solutions is the availability of large volumes of high-quality, diverse, and representative data. In many real-world scenarios, collecting sufficient labeled data is expensive, time-consuming, subject to privacy regulations, or practically impossible due to rare event occurrences. This problem, known as data scarcity, leads to challenges such as overfitting, biased model behavior, and poor generalization to unseen data.

To address these limitations, synthetic data generation has emerged as a powerful strategy. Synthetic data mimics the statistical properties of real-world datasets while allowing for greater control over data distribution and privacy protection. Traditional data augmentation methods, such as image rotations, scaling, and flipping, provide limited diversity and are often insufficient in complex applications where new, unseen scenarios must be simulated.

The advent of Generative AI, particularly models like Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and more recently, diffusion models, has revolutionized synthetic data creation. These models are capable of learning intricate data distributions and generating high-fidelity synthetic samples across various data types, including images, text, audio, and structured tabular data. Generative AI not only enriches datasets but also opens up new possibilities for training machine learning models where real data is limited or sensitive.

This paper aims to explore how Generative AI can effectively solve the problem of data scarcity in machine learning. We review the state-of-the-art techniques for synthetic data generation, discuss the underlying working principles of generative models, evaluate their effectiveness, and identify the key challenges and limitations associated with their adoption. Furthermore, we outline future enhancements

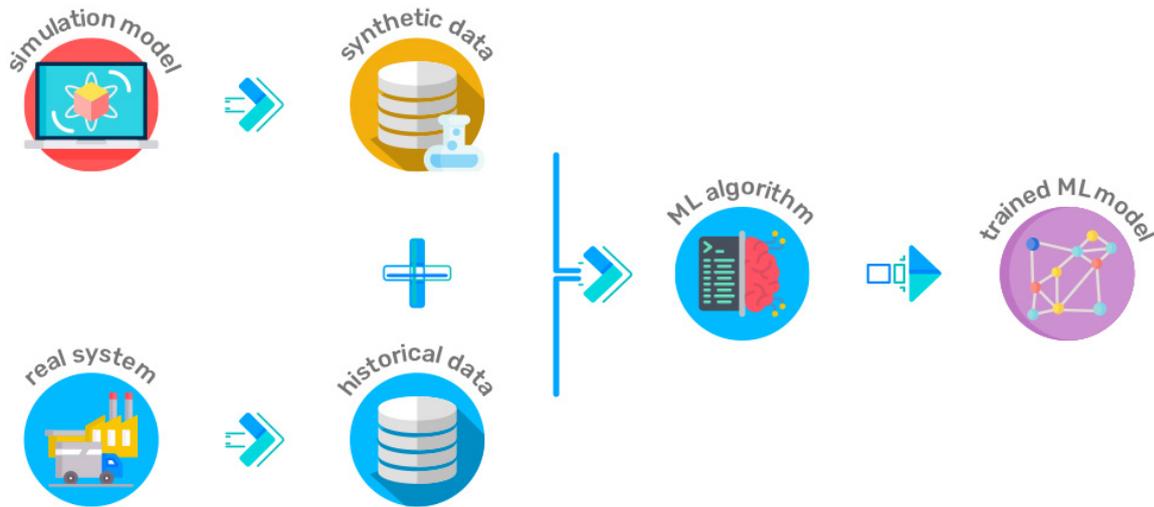needed to maximize the utility of synthetic data in next-generation AI systems.

Figure 1: Synthetic Data Generation: Transforming Analytics with Artificial Datasets

1.1 Data Scarcity Challenges in Machine Learning

Data scarcity presents a significant obstacle in the development and deployment of effective machine learning models. High-performing models, particularly those based on deep learning architectures, typically require extensive and diverse datasets to learn generalized patterns. However, in many domains, collecting large volumes of labeled data is prohibitively expensive, time-consuming, or impractical. Domains such as healthcare, finance, autonomous driving, and rare event detection often suffer from limited access to annotated datasets due to privacy concerns, regulatory restrictions, or the inherent rarity of certain phenomena. Insufficient data not only impedes model training but also increases the risk of overfitting, where the model memorizes the training data rather than learning to generalize. Furthermore, biased, imbalanced, or incomplete datasets can lead to models that perform poorly on underrepresented classes or in real-world settings, making data scarcity a critical problem to solve for robust AI development.

1.2 Emergence of Synthetic Data as a Solution

In response to data scarcity challenges, the use of synthetic data has emerged as a promising alternative. Synthetic data refers to artificially generated information that closely resembles real-world data in terms of statistical properties and diversity. Unlike traditional data augmentation techniques, which apply simple transformations to existing data, synthetic data generation creates entirely new data points, often tailored to meet specific requirements. This approach enables researchers and practitioners to create datasets that are larger, more balanced, and free from privacy or regulatory constraints. Synthetic data generation not only helps in expanding the training datasets but also allows the simulation of rare or edge-case scenarios that may be difficult to capture naturally. Consequently, synthetic data has become a critical enabler for advancing machine learning research, facilitating model validation, improving robustness, and accelerating the deployment of AI systems in sensitive and complex environments.

1.3 Role of Generative AI in Synthetic Data Creation

Generative AI has revolutionized synthetic data creation by enabling models to learn complex data distributions and generate high-quality, realistic samples. Unlike traditional rule-based synthetic data generation techniques, generative models such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models use deep learning to capture intricate patterns within the data. GANs, for instance, employ a dual-network architecture where a generator and discriminator compete, leading to the production of highly realistic synthetic samples. VAEs model data generation as a probabilistic process, allowing for smooth latent space interpolation and robust data synthesis. More

recently, diffusion models have demonstrated superior capabilities in producing detailed and diverse synthetic data by gradually transforming noise into coherent data samples. Generative AI has thus become a foundational technology in addressing data scarcity, providing scalable, privacy-preserving, and versatile solutions that are increasingly being adopted across industries to train, validate, and stress-test machine learning models.

2. Literature Survey

The challenge of data scarcity has long been a limiting factor in the performance and generalization capabilities of machine learning models. Researchers have increasingly turned to synthetic data generation as a solution, with Generative Artificial Intelligence (Generative AI) playing a pivotal role. This section reviews the key advancements, methodologies, and gaps in the current body of literature on synthetic data generation using Generative AI techniques. Early approaches to synthetic data generation relied on traditional data augmentation techniques such as rotation, flipping, scaling, and cropping, particularly in computer vision tasks (Shorten &Khoshgoftaar, 2019). However, these methods were often limited to minor variations of existing data and could not introduce fundamentally new data samples, thus restricting their ability to mitigate severe data scarcity.

The introduction of Generative Adversarial Networks (GANs) by Goodfellow et al. (2014) marked a significant breakthrough, enabling the creation of entirely new and realistic data samples. GANs have been extensively used across various domains, including image synthesis, text generation, and healthcare data creation. Subsequent enhancements, such as Conditional GANs (CGANs) and Wasserstein GANs (WGANs), have improved the stability and quality of synthetic data generation (Mirza &Osindero, 2014; Arjovsky et al., 2017).

Variational Autoencoders (VAEs), proposed by Kingma and Welling (2013), offered another pathway for synthetic data creation by learning latent representations of input data. Unlike GANs, VAEs maximize a lower bound on data likelihood, resulting in more stable but slightly less sharp synthetic samples. Despite their differences, both GANs and VAEs have been widely adopted for solving data scarcity issues in fields such as biomedical imaging (Frid-Adar

et al., 2018) and autonomous driving (Tremblay et al., 2018).

Recent advances have also explored diffusion models and transformer-based architectures for synthetic data generation, providing greater flexibility and improved fidelity, particularly in high-dimensional data spaces (Dhariwal & Nichol, 2021). These models have demonstrated the ability to generate diverse and high-quality synthetic datasets without overfitting to training data, thus better preserving data privacy — an increasingly important concern.

Despite these advances, challenges remain. Issues such as mode collapse in GANs, difficulties in evaluating synthetic data quality, and the risk of inadvertently encoding biases from training data into synthetic datasets have been widely acknowledged (Karras et al., 2019; Borji, 2020). Moreover, the lack of standardized evaluation metrics complicates the assessment of synthetic data effectiveness in downstream machine learning tasks.

In summary, Generative AI has emerged as a promising solution to the problem of data scarcity in machine learning. While substantial progress has been made in generating realistic and diverse synthetic data, ongoing research is needed to address challenges related to quality evaluation, bias mitigation, and domain-specific customization.

2.1 Traditional Data Augmentation Techniques

Traditional data augmentation has long been employed to enhance dataset diversity and mitigate the problem of data scarcity, particularly in computer vision and natural language processing tasks. Techniques such as rotation, scaling, cropping, flipping, translation, and noise addition create modified versions of existing data, thereby increasing the size and variability of training sets (Shorten &Khoshgoftaar, 2019). While effective to a certain extent, these methods primarily introduce small perturbations to original samples without generating fundamentally novel data. As a result, they have limited capability to address deep data insufficiencies, especially in cases where completely new patterns or classes are needed.

2.2 Evolution of Generative AI Models

The advent of Generative AI has significantly expanded the possibilities for synthetic data creation. Generative Adversarial Networks (GANs), introduced by Goodfellow et al.

(2014), revolutionized synthetic data generation by enabling the creation of entirely new, realistic data samples through adversarial training. Following this, Conditional GANs (CGANs), Wasserstein GANs (WGANs), and StyleGANs introduced mechanisms to control data generation and enhance stability (Mirza &Osindero, 2014; Arjovsky et al., 2017; Karras et al., 2019). Parallelly, Variational Autoencoders (VAEs) emerged as another generative paradigm focusing on learning probabilistic representations of input data (Kingma & Welling, 2013). More recently, diffusion models and transformer-based generators have showcased state-of-the-art results, particularly in high-fidelity image synthesis and large-scale text generation (Dhariwal & Nichol, 2021), further solidifying the role of Generative AI in overcoming data scarcity challenges.

2.3 Applications of Synthetic Data in Key Domains

Synthetic data generated through Generative AI has found applications across a range of industries. In healthcare, synthetic medical images and patient records have enabled model training without violating patient privacy (Frid-Adar et al., 2018). In autonomous driving, synthetic scenes and scenarios have been used to improve the robustness of self-driving algorithms (Tremblay et al., 2018). Financial services have leveraged synthetic transaction data for fraud detection model development, especially where real-world data is highly sensitive. Moreover, in cybersecurity, synthetic attack datasets help train intrusion detection systems without exposing systems to actual threats. These applications highlight the versatility and transformative potential of synthetic data across domains that traditionally struggle with data accessibility.

2.4 Comparative Analysis of Existing Methods

While various generative models offer synthetic data solutions, each comes with distinct advantages and limitations. GANs, while capable of producing highly realistic samples, suffer from training instability and mode collapse issues, where the generator produces limited varieties of outputs (Arjovsky et al., 2017; Borji, 2020). VAEs offer more stable training but often at the cost of less sharp and detailed synthetic outputs. Diffusion models, despite their success in generating high-quality

data, are computationally intensive and require longer training times (Dhariwal & Nichol, 2021). Evaluation of synthetic data also remains complex, with no universally accepted metrics to assess quality, diversity, and downstream task utility. Thus, selecting an appropriate generative model largely depends on the specific domain requirements and computational constraints.

2.5 Identified Research Gaps and Motivations

Despite notable advancements, significant research gaps persist in the domain of synthetic data generation. First, existing methods often struggle to ensure fairness and avoid bias replication from real datasets, raising ethical concerns. Second, the lack of standardized evaluation frameworks hinders consistent benchmarking and model comparison. Third, domain adaptation remains a challenge, where synthetic data needs to generalize across varying real-world environments. Finally, there is a need for lightweight and computationally efficient models capable of generating synthetic data on resource-constrained devices. Motivated by these gaps, this study aims to explore novel Generative AI architectures and evaluation mechanisms that can produce high-quality, unbiased synthetic datasets across multiple domains while ensuring computational efficiency.

3. Core Mechanisms Behind Generative AI-Based Synthetic Data Creation

Generative AI models operate on the fundamental principle of learning the underlying patterns, distributions, and structures of real-world datasets to create synthetic data that is both diverse and realistic. These models typically leverage advanced deep learning architectures such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and more recently, diffusion models. The primary working mechanism begins with the model analyzing the input data and encoding its key features into a latent space representation. This latent space acts as a compressed, high-level abstraction of the data distribution, from which new samples can be generated. In adversarial frameworks like GANs, two networks—the generator and the discriminator—engage in a competitive training process where the generator tries to produce data indistinguishable from real samples, and the discriminator aims to correctly identify synthetic from real data. This dynamic pushes

the generator towards producing increasingly authentic synthetic samples over time. VAEs, on the other hand, focus on probabilistic reconstruction by sampling from learned latent distributions to create new data points. Diffusion models approach the problem by progressively adding and removing noise to model the data generation process in a stable and structured manner. Regardless of the architecture, a critical aspect of these working principles is maintaining diversity without replicating specific training instances, thus ensuring both quality and privacy in the synthetic data. Evaluation metrics such as Fréchet Inception Distance (FID) and downstream task performance are often employed to assess the effectiveness of the generated data. Ultimately, these mechanisms collectively enable Generative AI systems to address the critical issue of data scarcity in machine learning by providing scalable, privacy-preserving, and high-quality synthetic datasets across various domains.

3.1 Fundamentals of Generative AI for Synthetic Data

Generative AI for synthetic data creation revolves around the concept of learning the distribution of real-world data and using this understanding to generate entirely new, realistic samples. The main models employed for this purpose are Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and the more recent diffusion models. Each of these models follows a distinct methodology but shares the common goal of expanding limited datasets without compromising authenticity or diversity.

Generative Adversarial Networks (GANs) consist of two neural networks—the generator and the discriminator—that are trained simultaneously through an adversarial process. The generator attempts to create synthetic data that is indistinguishable from real data, while the discriminator evaluates whether a given sample is real or generated. Through this continuous competition, the generator progressively improves, resulting in highly realistic synthetic datasets. GANs have been particularly influential in areas like image synthesis, medical imaging, and text-to-image generation due to their ability to produce fine-grained, detailed outputs.

Variational Autoencoders (VAEs), in contrast, rely on probabilistic modeling rather than adversarial training. VAEs encode input data into a structured latent space and then decode this representation to reconstruct the original input. During this process, they enforce a distribution over the latent variables, usually a Gaussian, enabling the generation of new data points by sampling from this latent space. VAEs are known for producing diverse, continuous variations of data, making them useful for applications requiring controlled sample generation and interpretability.
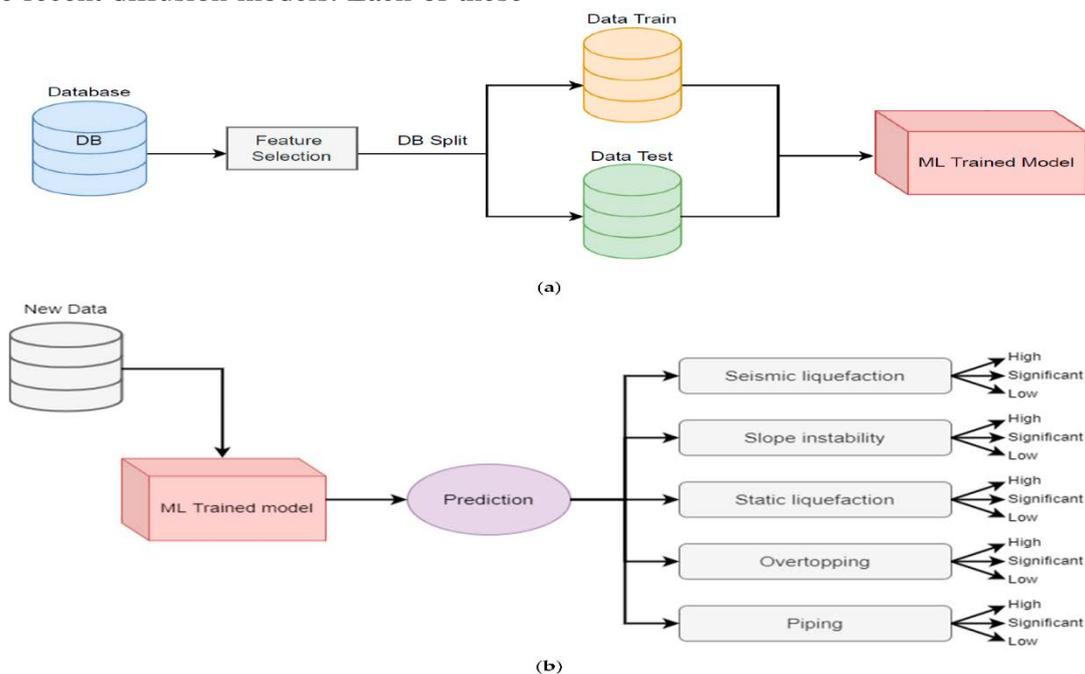

Figure 2: Generation of Synthetic Data for the Analysis of the Physical Stability of Tailing Dams through Artificial Intelligence

Diffusion models and other emerging approaches represent a newer class of generative models that have recently gained significant attention. Diffusion models generate data by gradually adding random noise to the input and learning to reverse this process step by step, effectively denoising to retrieve the original data distribution. These models have demonstrated exceptional performance in generating high-quality images and are considered more stable during training compared to GANs. Other emerging techniques, including transformer-based generative models and energy-based models, continue to push the boundaries of what is achievable in synthetic data generation, offering promising directions for future research and application.

3.2 Synthetic Data Generation Workflow

The workflow for synthetic data generation using Generative AI models follows a structured sequence of stages to ensure that the generated data is realistic, diverse, and useful for downstream applications. The process begins with data preprocessing and feature engineering, where the real-world data is cleaned, normalized, and transformed into a format suitable for model consumption. This stage may involve handling missing values, scaling numerical features, encoding categorical variables, and selecting relevant attributes that capture the most critical information from the original dataset. Proper preprocessing ensures that the generative models receive high-quality inputs that facilitate effective learning of data distributions.

Following preprocessing, the focus shifts to model architecture design and training strategies. Depending on the nature of the data and the specific requirements of the application, appropriate generative models such as GANs, VAEs, or diffusion models are selected. The architecture is carefully designed, specifying the number of layers, activation functions, loss functions, and optimization algorithms. During training, the models learn to map random inputs or latent vectors to meaningful synthetic outputs that closely resemble real-world samples. Strategies such as data augmentation during training, progressive growing of networks, and hyperparameter tuning are often employed to enhance model performance and stability. Regular evaluation during training using metrics like loss curves, Fréchet Inception

Distance (FID), and visual inspections helps in monitoring the quality of the synthetic outputs.

Once the models are trained, post-processing and data validation become critical to ensure that the generated synthetic data meets the desired standards of utility, fairness, and privacy. Post-processing steps may include filtering out low-quality samples, balancing the distribution of generated data, and introducing controlled variations if necessary. Data validation involves assessing the synthetic data's statistical similarity to real data, evaluating its performance on downstream machine learning tasks, and verifying that no sensitive or identifiable information from the original dataset has been inadvertently leaked. This comprehensive workflow ensures that synthetic data produced through Generative AI not only addresses data scarcity but also adheres to high standards of quality, ethical compliance, and practical applicability.

3.3 Evaluation of Synthetic Data

The evaluation of synthetic data is a critical step to ensure that the generated datasets are not only realistic and diverse but also effective for use in real-world machine learning applications. Authenticity and quality metrics are the primary indicators used to assess how closely the synthetic data resembles the original data. Metrics such as Fréchet Inception Distance (FID), Inception Score (IS), Precision-Recall curves, and Kernel Inception Distance (KID) are commonly employed to quantitatively measure the visual and statistical similarities between synthetic and real samples. A lower FID score, for instance, indicates that the generated data points are close to the true distribution of the original data. Visual inspections, t-SNE visualizations of feature spaces, and domain expert reviews also complement these metrics to provide qualitative assessments of authenticity and variability in the synthetic datasets.

Beyond authenticity, it is crucial to evaluate the usability of synthetic data and its impact on machine learning models. Usability evaluation focuses on how well models trained with synthetic data perform compared to those trained with real data. Metrics such as model accuracy, precision, recall, F1-score, and ROC-AUC are analyzed when using synthetic data either independently or in combination with real data. Effective synthetic datasets should

enhance model training, particularly in scenarios involving class imbalance, rare event prediction, or privacy-sensitive applications. Furthermore, synthetic data should not introduce biases or reduce the generalization ability of models. A successful evaluation framework ensures that synthetic data maintains high fidelity to the original distributions while providing tangible benefits to machine learning performance, thereby validating its role as a viable solution to data scarcity challenges.

3.4 Challenges and Limitations

Despite the significant potential of Generative AI for synthetic data creation, several challenges and limitations persist that must be carefully addressed. One major concern is the issue of bias and lack of diversity in the generated data. If the training data is imbalanced or contains inherent biases, generative models tend to replicate and even amplify these issues, leading to synthetic datasets that are not truly representative or fair. Ensuring diversity and mitigating biases in synthetic data generation remains an open research challenge that requires the careful design of training strategies, model architectures, and evaluation frameworks.

Overfitting and mode collapse are additional technical challenges that often arise during the training of generative models. Overfitting occurs when a model memorizes specific samples from the training data instead of learning the underlying distribution, leading to poor generalization and potential privacy risks. Mode collapse, particularly prevalent in GANs, happens when the generator produces limited varieties of outputs, failing to capture the full diversity of the original data distribution. These issues compromise the quality and utility of synthetic data, making it less effective for augmenting real datasets or training robust machine learning models.

Ethical, privacy, and legal concerns further complicate the deployment of synthetic data in sensitive applications. Although synthetic data is often promoted as privacy-preserving, there remains a risk that the generated data might inadvertently reveal sensitive information or enable re-identification attacks, especially if overfitting occurs. Moreover, the regulatory landscape for synthetic data usage is still evolving, raising questions around ownership, accountability, and compliance with standards such as GDPR and HIPAA. Addressing these ethical and legal challenges requires a combination of robust technical safeguards, transparent model design, and adherence to established data protection principles to ensure that synthetic data serves as a truly safe and equitable solution.

4. Conclusion

Generative AI has emerged as a transformative solution to address the longstanding challenge of data scarcity in machine learning by enabling the creation of high-quality synthetic datasets. Through the exploration of advanced models such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models, it is evident that synthetic data can effectively mimic real-world distributions while enhancing the scalability, privacy, and diversity of data used for AI training. The structured workflow encompassing preprocessing, model training, and post-processing ensures that synthetic data maintains authenticity and utility for practical applications. Nevertheless, challenges such as bias propagation, overfitting, mode collapse, and ethical concerns highlight the need for careful model design, rigorous evaluation, and responsible deployment practices. As research continues to evolve, overcoming these limitations will be critical to fully realizing the potential of synthetic data generation. By addressing these challenges, Generative AI can pave the way for more inclusive, ethical, and high-performing machine learning systems across a wide range of domains.

5. Future Enhancements

As Generative AI continues to evolve, several future enhancements can be pursued to further refine synthetic data generation and expand its applications. One key area for improvement is the development of more robust and diverse generative models that can handle increasingly complex data types, including multimodal data and real-time data streams. This would allow for broader use cases, such as in dynamic environments where data constantly changes. Additionally, integrating fairness and bias mitigation techniques directly into the model training process could help address ethical concerns and ensure that the generated data is equitable and representative of diverse populations. Another promising direction is the incorporation of explainability into generative models, making the generation process more transparent and interpretable, which is crucial

for high-stakes applications in sectors like healthcare and finance. Furthermore, the application of synthetic data in federated learning environments, where data privacy and security are paramount, presents an exciting opportunity for enhancing machine learning without compromising data confidentiality. Finally, the continuous refinement of privacy-preserving techniques, such as differential privacy, will help mitigate the risk of inadvertent data leakage, ensuring that synthetic data remains safe and compliant with global data protection regulations. These future advancements will help further unlock the potential of generative AI in overcoming data scarcity while addressing the evolving challenges in the field.

5.1 Advancements in Federated and Privacy-Preserving Synthetic Data

One of the most promising advancements in synthetic data generation is the integration of **federated learning** and **privacy-preserving techniques** into the generative models. Federated learning enables collaborative model training without the need to share sensitive data, making it a critical tool for industries dealing with highly confidential information, such as healthcare and finance. By combining federated learning with generative models, synthetic data can be generated in a decentralized manner, enhancing privacy without sacrificing model performance. Additionally, techniques like **differential privacy** can be further incorporated into synthetic data creation, ensuring that generated datasets do not inadvertently leak sensitive information from the original training data. This combination can provide a robust and scalable framework for privacy-preserving data generation, addressing both regulatory and security concerns in sensitive sectors.

5.2 Enhancing the Generalization Ability of Generative Models

Generative models, particularly GANs and VAEs, can sometimes face challenges in **generalization** when trained on limited datasets or overly specific distributions. To improve their effectiveness, future advancements should focus on techniques that enhance the generalization ability of these models. This could involve leveraging **transfer learning** or **few-shot learning** strategies, allowing generative models to learn better representations with minimal data while maintaining the ability to generate diverse and realistic synthetic datasets. Furthermore, incorporating techniques like **meta-learning** could enable generative models to quickly adapt to new, unseen data distributions, leading to better performance across various domains and applications. Enhancing generalization is crucial for ensuring that synthetic data can be used effectively in a wide range of real-world scenarios.

5.3 Real-Time Synthetic Data Creation for Adaptive Systems

As industries increasingly adopt **adaptive systems** that require continuous data updates, the ability to generate **real-time synthetic data** will become an essential capability. Real-time data creation can enable systems to operate in dynamic environments where the data changes frequently, such as in autonomous vehicles, real-time medical diagnostics, and industrial IoT systems. To achieve this, generative models will need to be optimized for rapid data synthesis, reducing latency while maintaining data quality. Incorporating **streaming data processing** and **online learning** into generative models will be key to enabling real-time, continuous synthetic data generation. This would allow adaptive systems to quickly respond to new information, ensuring their decision-making processes are based on the most up-to-date data available.

5.4 Building Ethical and Regulatory Frameworks

As synthetic data becomes more prevalent, the need for clear **ethical guidelines** and **regulatory frameworks** will become more pressing. Future work should focus on creating robust frameworks that ensure the responsible use of synthetic data, particularly in sensitive sectors like healthcare, finance, and law enforcement. These frameworks should address key concerns, such as data ownership, consent, accountability, and the potential for misuse. In addition, regulatory bodies may need to develop new standards specifically tailored to synthetic data, ensuring compliance with data protection laws such as **GDPR** and **HIPAA**. Ethical frameworks will also play a vital role in ensuring that synthetic data does not perpetuate biases or lead to discriminatory outcomes in machine learning applications.

5.5 Exploring Multimodal and Cross-Domain Synthetic Data Opportunities

With the growing complexity of modern AI applications, there is a significant opportunity to

explore the creation of **multimodal** and **cross-domain synthetic data**. Multimodal synthetic data involves generating data that spans multiple types of input—such as images, text, and audio—that can be used to train models for applications like autonomous driving or multi-sensory analysis in healthcare. Similarly, cross-domain synthetic data generation seeks to bridge gaps between distinct domains, enabling models trained in one field (e.g., medical imaging) to be applied effectively in another (e.g., agricultural sensing). Advancements in this area could unlock new opportunities for data sharing, collaboration, and cross-disciplinary research, leading to the creation of more robust AI systems that can operate across diverse contexts and domains.

References:

1. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … & Bengio, Y. (2014). Generative adversarialnets. In Advances in Neural Information Processing Systems (NIPS), 27, 2672-2680.

2. Kingma, D. P., & Welling, M. (2014). Auto-Encoding Variational Bayes. In International Conference on Learning Representations (ICLR).

3. Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. In International Conference on Machine Learning (ICML), 1-10.

4. Bengio, Y. (2013). Learning deep architectures for AI. Foundations and Trends®inMachineLearning,2(1),1-127.

5. Ramya, R., and T. Sasikala. "Experimenting biocryptic system using similarity distance measure functions." In 2014 Sixth International Conference on Advanced Computing (ICoAC), pp. 72-76. IEEE, 2014.

6. Ramya, R. "Evolving bio-inspiredrobots for keep away soccer through genetic programming." In INTERACT- 2010, pp. 329-333. IEEE, 2010.

7. Salimans, T., Goodfellow, I. J.,Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved Techniques for Training GANs. In Advances in Neural Information Processing Systems(NIPS),29,2234-2242.

8. Larsen, A. B. L., Søndberg, L. B., & Kosiorek, A. R. (2016). Autoencoding Variational Bayes: A Comprehensive Review. In Proceedings of the International Conference on Machine Learning (ICML), 1-12.

9. Makhzani, A., Shlens, J., Jaitly, N., & Goodfellow, I. (2015). Adversarial Autoencoders. In International ConferenceonLearningRepresentations (ICLR).

10. Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433-460.

11. Sutton, R. S., & Barto, A. G. (1998). ReinforcementLearning: An Introduction. MIT Press.

12. Yegnanarayana, B. (2009). Artificial Neural Networks. PHI Learning Pvt.Ltd.

13. Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on Machine Learning (ICML), 37-44.

14. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

15. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., &Manzagol, P. A. (2008). Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th International Conference on Machine Learning (ICML), 1096-1103.

16. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer VisionandPatternRecognition(CVPR), 770-778.

17. Hinton,G.E.,Osindero,S.,&Teh,Y. W. (2006). A Fast Learning Algorithm for Deep Belief Nets. Neural Computation, 18(7), 1527-1554