International Journal of Innovative Research and Scientific Studies, 8(5) 2025, pages: 286-293

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Generative adversarial networks for synthetic data generation: A systematic review of techniques, applications, and evaluation methods

Rajermani Thinakaran[1], Ram Kinkar Pandey[2*], Prabhat Kr Srivastava[3], Jyotsna[4], Sudhakar Madhavedi[5]

[1]*Faculty of Data Science and Information Technology, INTI International University, Malaysia.*
[2]*Department of Computer Science & Engineering, Research Fellow, INTI International University, Malaysia.*
[3]*Department of Computer Science & Engineering, IMS Engineering College, Ghaziabad, India.*
[4]*Vice Principal, Kshatriya School, Nizamabad, India.*
[5]*Dept. of Business Management Kshatriya College of Engineering (under JNTU, Hyderabad), India & Research Fellow, INTI International University, Malaysia.*

Corresponding author: Ram Kinkar Pandey (*Email:* *dr.ramkpandey@gmail.com*)

Abstract

Generative adversarial networks (GANs), which have emerged as one of the powerful frameworks for generating synthetic data, have proven remarkably capable across domains. This systematic review explores the rapidly evolving GAN landscape, particularly their applications for generating high-fidelity synthetic data that resemble real-world datasets' statistical properties. We comprehensively analyze recent literature to present the following key findings: 1. GANs' Capabilities: GANs have demonstrated significant potential across various fields, especially in creating synthetic data that mimic real-world datasets. 2. State-of-the-Art Architectures: Advanced GAN variants, such as Conditional GANs, Wasserstein GANs, and Cycle GANs, have shown great promise for transformation in sectors like healthcare, finance, and image processing. 3. Evaluation Methodologies: Metrics for assessing GAN-generated data include statistical similarity, downstream task performance, and privacy preservation, highlighting strengths and limitations in current evaluation paradigms. 4. Training Difficulties: GANs face challenges such as mode collapse, instability, and sensitivity to hyperparameters, which require further innovation and exploration. Additionally, we critically examine the methodologies used to evaluate the quality and utility of GAN-generated data. Metrics like statistical similarity, downstream task performance, and privacy preservation provide a broad view of current strengths and limitations. Besides synthetic data generation using GAN-based methods, this review discusses training difficulties and emerging directions aimed at mitigating issues like mode collapse, instability, and hyperparameter sensitivity. The findings emphasize significant progress in GAN-based synthetic data generation but underline the need for a robust, standardized evaluation framework and continued innovation in model architectures. 1. Robust Evaluation Framework: Developing a standardized evaluation framework for GAN-generated data is essential for advancing the field. 2. Model Architecture Innovation: Ongoing innovation in model architectures is necessary to overcome current limitations and enhance GAN performance. 3. Synthetic Data Generation: GANs hold great potential for generating synthetic data, which can address data privacy concerns, data scarcity, and data augmentation needs. This review aims to help researchers and practitioners understand the current state and future directions of GAN applications in synthetic data generation.

Keywords: Data privacy, Deep learning, Machine learning, Evaluation metrics, Mode collapse, Generative adversarial networks (GANs), Synthetic data generation, Data augmentation, Systematic review, Wasserstein GAN, Conditional GAN.

DOI: 10.53894/ijirss.v8i5.8655
Funding: This work is supported by INTI International University, Malaysia.
History: Received: 19 May 2025 / **Revised:** 24 June 2025 / **Accepted:** 26 June 2025 / **Published:** 18 July 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

1.1. The Rise of Synthetic Data and the Need for Advanced Generation Techniques

Data has been growing exponentially in the 21st century, fueling enormous breakthroughs in artificial intelligence (AI) and machine learning (ML). Nevertheless, there is an even greater challenge to this data-driven revolution: privacy concerns, scarcity of data in niche domains, and high costs involved with the need for data annotations. All these limitations have generated an increasing interest in synthetic data artificially created data that imitates real-world datasets' statistical properties while maintaining confidentiality and addressing many problems associated with the acquisition and use of real data. Synthetic data promises significant benefits for training robust AI models, particularly in situations where the acquisition of real data is sensitive, unavailable, or insufficient. For instance, in healthcare, it can be used to train diagnostic models without any confidentiality issues related to patient details [1]. In finance, the system can utilize synthetic transaction data when developing fraud detection models without exposing sensitive financial information [2].

These include rule-based systems and statistical modeling techniques, but they fail to capture the inherent complexities and high-dimensional distributions in real-world data [3, 4]. Such approaches may produce unreal data, lose subtle correlations, or exhibit bias that negatively affects the performance of models trained upon such data. However, the growing demand for high-fidelity synthetic data mandates new techniques that are more powerful and have fine precision, closely approximating real-data patterns and intricacies across various domains. Finally, privacy regulations have recently become even stricter through bodies such as the General Data Protection Regulation (GDPR) and CCPA, along with a generally increasing consumer interest, which necessitates alternative private access solutions rather than direct use of the same actual data set [5, 6].

1.2. Generative Adversarial Networks (GANs): A Paradigm Shift in Data Generation

The Goodfellow et al. [7] is an example of the newly developed Generative Adversarial Networks (GANs), which will revolutionize the generation of synthetic data. The GAN represents a paradigm shift and offers a very powerful framework to learn complex distributions of data as well as novel, realistic samples of data. Unlike approaches that depend on predefined rules or explicit statistical models, GANs are capable of exploiting deep neural networks combined with adversarial training for data-driven discovery of the underlying structure of real data. Fundamentally, the idea behind GANs lies in a two-player game involving two neural networks, namely, the generator and the discriminator. The generator is designed to produce synthetic data samples that are indistinguishable from real data, and the discriminator is supposed to differentiate between real and synthetic samples. This adversarial process enhances both networks iteratively: the generator learns to produce increasingly realistic data, and the discriminator becomes better at distinguishing subtle differences.

The GAN is learned through a dynamic interplay between the generator and the discriminator. As a result, it generates synthetic data that can capture the subtlest patterns, correlations, and other nuances present in real data. The ability of GANs to learn from data without explicit programming or feature engineering has made them a versatile tool capable of generating data in various formats, including images, text, tabular data, and time series. GANs have demonstrated remarkable success in producing high-fidelity synthetic data with similar statistical properties, visual appearance, and semantic coherence to actual data. This potential has garnered significant interest among academics and industry professionals regarding data augmentation, privacy preservation, and data sharing across different applications [8, 9].

1.3. Scope and Objectives of the Systematic Review

This systematic review aims to provide a comprehensive overview of the rapidly evolving landscape of GANs for synthetic data generation [10]. The review is divided into three key aspects: techniques, applications, and evaluation methods. First, we explore the diverse array of GAN architectures that have been developed since their inception [11].

The authors begin with the fundamental architecture of GANs and trace its development through advanced versions, including Conditional GANs (cGANs), which enable controlled data generation, and WGANs, which offer improved training stability [12, 13]. Later, we explore even recent architectures, such as CycleGANs and StyleGANs, for some unique features. We examine the potential of GANs in multiple domains, specifically in healthcare and finance, which are areas characterized by data privacy concerns and scarcity. Furthermore, the authors mention the application of GANs in image processing and other emerging fields [14]. Then, the authors critically analyze the methods applied to assess the quality and usability of GAN-generated data. Finally, they discuss metrics based on statistical fidelity, downstream task performance, and privacy preservation to provide a comprehensive view of the current evaluation landscape [15].

2. Evolution and Taxonomy of GAN Architectures

2.1. Foundational GAN Architectures: The Genesis of Adversarial Training

In a seminal work, Keerthana and Boopathi Raja [13] and Vaz and Figueira [16], introduced GANs as a new paradigm for generative modeling in 2014. The original architecture of GANs is a two-neural network construction known as a generator (*G*) and a discriminator (*D*), which play a minimax game. The generator generates artificial data samples based on the noise vector it is fed. On the other hand, the discriminator distinguishes the original data samples taken from the training dataset from synthetic data samples that have been produced by the generator [17, 18]. The entire process is known as adversarial training, in which G minimizes a value function while *D* attempts to maximize the same function; both networks are trained iteratively [19]. Early versions of GANs, though pioneering, suffered from unstable training, mode collapse (where the generator produced few varieties of samples), and the inability to produce high-resolution images [12, 17]. These factors pushed researchers into developing alternative architectures and training strategies.

2.2. Conditional GANs (cGANs) Steering Data Generation with Control

cGANs were the next step in the evolution of GANs and brought the idea of conditioning into the generative process [15, 20]. In cGANs, both the generator and discriminator are conditioned on additional information, such as class labels or other attributes, to control data generation. For example, when generating images of specific digits, the cGAN can be conditioned on the digit's label, enabling the generation of images of a particular digit [11]. This ability has proven very useful for various applications and image-to-image translations, where the generator takes input images as conditioned and outputs images into a different space [21]. The use of cGAN has become highly efficient in many research areas such as medicine and other diseases, imaging synthesis, and its respective applications within medical diagnosis research areas [6, 9].

2.3. Addressing Training Instability: Wasserstein GANs (WGANs) and their Variants

Training instability was also one of the main issues with the design of GANs. In practice, this meant mode collapse and vanishing gradients often presented serious challenges for convergence to high-quality samples [5, 10]. To help resolve these difficulties, Wasserstein GANs (WGANs), which leverage the Wasserstein distance (often referred to as Earth Mover's distance) rather than the Jensen-Shannon divergence (JSD), were later introduced in WGAN [22]. It turns out that the Wasserstein distance is an even more meaningful measure of the distance between real and generated data distributions, leading to far less unstable training and highly improved sample quality. WGANs, especially with a gradient penalty term, have become very popular nowadays for a wide variety of applications because they result in much steadier training and the generation of higher-quality samples [4, 23]. They have demonstrated potential in domains such as finance, where stable training is also crucial for producing realistic financial data [10, 24].

2.4. Advanced GAN Architectures: Cycle GANs, Style GANs, and Beyond

After GANs were introduced, they rapidly developed as a field with many high-order architectures that have been proposed specifically to solve certain challenges and increase the capability of generative models. Cycle GANs introduced a new approach for unpaired image-to-image translation, and the possibility was provided through their training of mappings between two domains without requiring paired training data [11, 25]. This is achieved using a cycle consistency loss; this loss is intended to ensure that an image translated from one domain into another and back again can be restored to its original form. The use of Cycle GANs has been applied to style transfer, object transfiguration, and even medical image conversion [14, 25].

StyleGANs, as developed by NVIDIA, transformed high-resolution image synthesis, clearly demonstrating that the ability to generate very realistic and diverse images is possible [8, 26]. StyleGANs utilize a style-based generator that permits finer control over various levels of details in the synthesized images, enabling manipulation of specific attributes, such as pose, expression, and style. These innovations open new opportunities for art, design, and entertainment [27].

Other than that, researchers have studied many other GAN architectures, such as BigGANs designed for large-scale image synthesis, SAGANs, which incorporate long-range dependencies in image generation, and PGGANs, which grow high-resolution images progressively [3, 18]. This expanded the horizon of GANs since they can now tackle more complicated problems in generating data from different domains. Recent studies have also included some recent works on the applications of GANs in agriculture, for instance, for image augmentation that improves the performance and robustness of the model in several applications of farming [18, 26]. These architectures opened new horizons for synthetic data generation and demonstrated how versatile and powerful the GAN framework is. They have also found numerous applications of GANs for synthetic data generation in the domain of anomaly detection [27].

3. Theoretical Foundations and Training Dynamics of GANs

GANs rely on foundations in game theory and information theory. Two neural networks are used to train these models in a complex interplay. This chapter briefly reviews the core theoretical ideas behind GANs and outlines some key optimization techniques employed during training [19].

3.1. The Adversarial Objective: A Minimax Game

The adversarial relationship between a generator, *G*, and a discriminator, *D*, forms the core of GANs. This is presented as a minimax game where G tries to create synthetic data close to real data, and *D* tries to distinguish between real and synthetic samples. The objective function *V(G, D)* is formulated in such a way that *D* attempts to maximize it by correctly identifying real and fake samples, and G attempts to minimize it by fooling *D*. In other words, the generator tries to make the discriminator lose its ability to distinguish its output from real data, which involves a dynamic, competitive learning process.

This setup is theoretically connected to the Jensen-Shannon Divergence (JSD) between the real and generated data distributions. Ideally, the generator should be able to produce samples that perfectly match the real data distribution by minimizing the JSD. However, mode collapse and vanishing gradients are practical challenges that may occur, which can make training difficult, as illustrated in Table 1 and Figure 1.

Table 1.
Performance Comparison of GAN Architectures on Image Datasets.

GAN Architecture	CIFAR-10 FID	ImageNet FID
DCGAN	21.7	65.4
WGAN-GP	20.5	58.2
SNGAN	18.6	27.6
GAN2 (ADA)	2.42	3.84

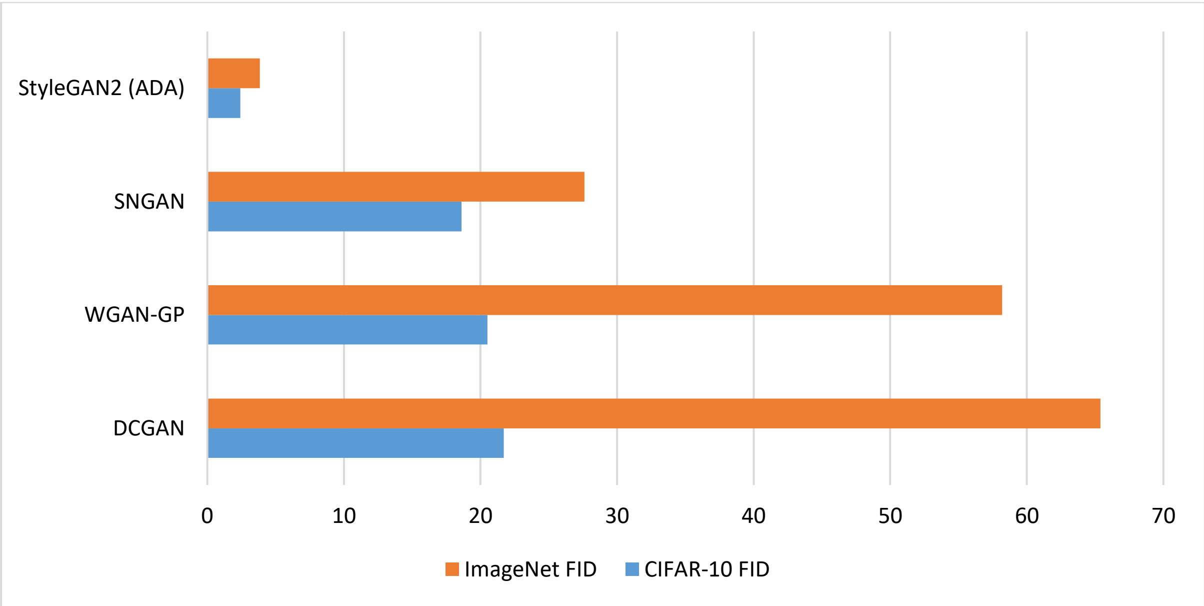

Figure 1.
Performance Comparison of GAN Architectures on Image Datasets.

3.2. Optimization Techniques: Algorithms for GAN Training

Training GANs is notoriously challenging, with the underlying optimization landscape often being complex and non-convex. Stochastic Gradient Descent (SGD) and its variants, Adam and RMSprop, are very popular update schemes for *G* and *D's* parameters. The optimizer and its hyperparameters significantly influence the stability of training. Balancing *G* and *D* is important. An overly powerful *D* leads to uninformative gradients for *G*, causing instability; a dominant *G* can hence exploit *D's* weaknesses and lead to mode collapse [20, 25].

To address these issues, several techniques have been developed. Feature matching stabilizes training by forcing *G* to match the intermediate feature representations in *D*, rather than their outputs. Minibatch discrimination helps the discriminator consider the relationships between the samples and helps avoid mode collapse. One-sided label smoothing reduces *D's* confidence in real samples to prevent it from overpowering *G*.

3.3. Enhancing Stability and Performance: Advanced Strategies

More advanced training techniques for GANs include gradient penalty, which was proposed in Wasserstein GAN with Gradient Penalty (WGAN-GP), by adding a Lipschitz constraint on *D* via the penalty on the norm of its gradient, thereby enhancing stability and sample quality, and spectral normalization, which limits the Lipschitz constant of *D* by normalizing the spectral norm of its weights, further improving the stability of the training process. These techniques, combined with good architecture design and hyperparameter tuning, have made GAN training much more feasible. Nonetheless, the field remains an active area of research aimed at developing more robust and efficient optimization algorithms to enable GANs to generate synthetic data with high fidelity. Studying these training dynamics is crucial for advancing the practical applications of GANs across various domains [21].

4. Applications of GANs in Synthetic Data Generation

GANs have proven surprisingly capable of creating high-fidelity synthetic data across a broad range of application domains. By learning complex distributions over data and synthesizing realistic samples, GANs have opened new opportunities where data privacy, scarcity, or augmentation pose significant challenges. The following presents some of the most important applications of GANs in synthetic data generation.

4.1. GANs in Healthcare: Revolutionizing Medical Data Synthesis

The healthcare industry, in particular, faces major problems related to the privacy of medical data and the scarcity of diverse medical data. GANs have emerged as a powerful tool for generating synthetic medical data to facilitate privacy-preserving data sharing and analysis [23]. In medical imaging, GANs have been used to synthesize realistic medical images, such as X-rays, CT scans, and MRIs, with specific pathologies or characteristics. For example, cGANs can be trained to generate synthetic chest X-rays with various lung diseases, providing valuable resources for training and validating diagnostic models. Such synthetic datasets can enhance existing real datasets, improving the robustness and generalizability of AI models for disease detection and diagnosis. GANs can also generate synthetic EHRs. which include patient demographics, diagnoses, medications, and lab results.

Synthetic EHRs can be used to advance research and development of predictive modeling and personalized medicine without compromising patient privacy. All such synthesized datasets created by researchers using GANs can be tested and further developed to predict disease risks or optimize treatment plans by signaling potential drug interactions. Such applications in healthcare can accelerate medical research, improve diagnostics, and potentially enhance patient care [28].

4.2. GANs in Finance, Image Processing, and Other Domains: Expanding the Impact of Synthetic Data

Apart from healthcare, GANs are applied in several other domains. In finance, GANs can generate synthetic financial data, such as time-series data for stock prices, transaction records, and credit risk profiles. These synthetic datasets are used for training and validating machine learning models for fraud detection, algorithmic trading, and risk management. The use of synthetic financial data can also facilitate the simulation of market scenarios and stress testing of underlying financial models under various economic conditions. Ensuring the privacy and security of data is paramount in this highly regulated industry.

GANs are also utilized in image processing tasks such as image inpainting, super-resolution, and data augmentation. For example, the GAN could be applied to fill out the missing parts or corrupted portions of images, or in restoring old photographs, generating high-resolution versions from low-resolution ones. These data augmentation techniques can also boost the performance of image classification and object detection models by generating additional samples with variations in lighting, pose, and background. Another domain where GANs have proved themselves is natural language processing: They can very well be applied to synthetic text data generation for training language models, and they can also generate synthetic network traffic data for intrusion detection systems. As discussed above, GANs offer versatility, hence are valuable in generating synthetic data across a wide spectrum of applications as depicted in Table 2 and Figure 2.

Table 2.
Usage of Synthetic Data in Healthcare.

Reason for Using Synthetic Data in Healthcare	Percentage of Respondents (%)
To comply with data privacy regulations (e.g., GDPR)	78
To increase the size of the training dataset	65
To access data that is difficult to obtain	58
To improve model performance	52
To reduce the cost of data acquisition	45

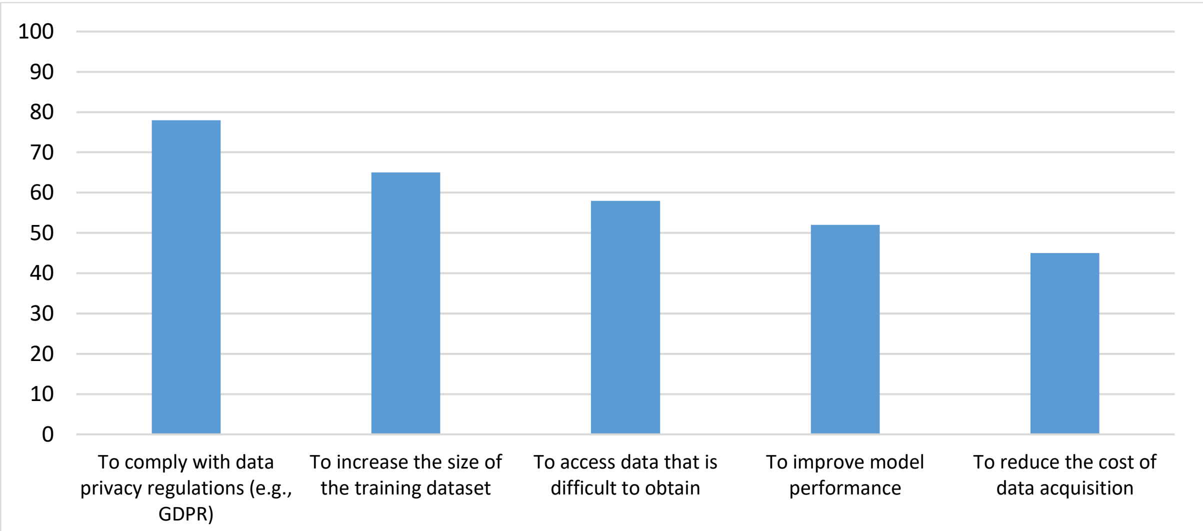

Figure 2.
Usage of Synthetic Data in Healthcare.

GANs are a highly promising technology for generating synthetic data, which has significant implications across various industries. The ability to produce high-fidelity, privacy-preserving synthetic data will transform sectors such as healthcare and finance, and potentially others, accelerating research and enhancing model performance through data-driven decision-making. We can anticipate numerous innovative applications in the coming years, driven by advancements in GAN research.

5. Challenges and Future Directions in GAN-Based Synthetic Data Generation

Although GANs have shown great promise for synthetic data generation, there are still challenges that remain, motivating continued research. Addressing these key issues is essential to ensure the full utilization of GANs and their responsible and effective application across various fields.

5.1. Addressing Persistent Challenges: Mode Collapse, Training Instability, and Hyper parameter Sensitivity

The primary challenge in GAN training is mode collapse, where the generator produces limited varieties of samples, failing to capture the full diversity of the real data distribution. This often occurs when the generator finds it easier to deceive the discriminator by producing a few realistic samples rather than learning the entire data distribution. Training instability, characterized by oscillations and a lack of convergence in adversarial training, is another concern. The root causes include the complexity of the minimax game dynamics and the non-convex nature of the objective function. Hyperparameter settings in GANs are particularly sensitive, making it challenging to identify the most suitable configuration for different datasets and architectures.

Various efforts have been concentrated on developing ways to overcome the issues. The most common way to address mode collapse is through alternative objective functions: that is, one uses Wasserstein distance applied to WGAN, or novel training methods such as unrolled GANs or minibatch discrimination. Techniques that could improve stability include gradient penalty and spectral normalization. Hyperparameter sensitivity is a subject of much current research work. Efforts focus on training algorithms that are more robust to changes and more automated forms of hyperparameter optimization.

5.2. Emerging Trends: Hybrid Models, Explainable GANs, and Ethical Considerations

Several emerging trends have defined the future of GAN-based synthetic data generation. Most recently, one popular trend has involved hybrid models integrating GANs with other architectures, for instance, VAEs, such that hybrid models take the benefits from both, enabling GAN sharp sample generation combined with the stability and representational properties in training a VAE model. The trend towards explainable GANs that aim to reveal the internal mechanisms of GANs and how factors impact the output data is increasing. Such research is very crucial in determining the limitations of GANs, identifying possible biases, and thus building trust in the output generated.

Furthermore, in light of the growing applications of GANs to synthetic data generation, there are important ethical concerns: bias, fairness, and misuse. It becomes critical that GANs do not perpetuate or enhance biases within training data and consequently produce unfair or discriminatory results. Researchers are also currently investigating detection and mitigation strategies for biases in GANs and best practices for the safe development and release of this technology. There are also concerns about the misuse of GANs to create deceptive or harmful material, such as deepfakes. It is therefore critical to address these ethical challenges to ensure that GANs are used for beneficial purposes and their impact on society is positive.

In conclusion, GANs, with their potential to offer a powerful framework for synthetic data generation, must overcome challenges and continue research on emerging trends in this area. Further work on mitigating mode collapse, training stabilization, hybrid models, explainable GANs, and ethics will shape the future of this field, enabling it to realize its full potential in synthetic data generation using GAN-based methodologies.

6. Conclusion

This review has drawn attention to the rapidly evolving landscape of GANs in synthetic data generation. Our analysis covered theoretical foundations, architecturally diverse models, key applications, and evaluation methods that define this transformative technology. From the original GAN to advanced versions such as Conditional GANs, Wasserstein GANs, Cycle GANs, and Style GANs, significant progress has been made within a relatively short period. These advancements have substantially enhanced the capabilities of GANs in learning complex data distributions and generating high-fidelity synthetic data across various domains. Exploration of the theoretical underpinnings, from the minimax game to optimization techniques, has illuminated the intricacies of GAN training and the primary challenges in achieving stable and effective results. In health and finance applications, GANs have demonstrated considerable potential for transforming data-driven research and development, particularly in sharing, augmenting, and analyzing data within privacy-preserving boundaries.

The synthesis of findings in this review indicates the significant potential of GANs for addressing critical issues related to data privacy, scarcity, and the need for diverse, representative datasets. GANs have demonstrated effectiveness in generating synthetic data, which can enhance existing datasets and improve the robustness of machine learning models; they also assist in research areas with limited or sensitive real data. However, the review also highlights ongoing challenges in GAN training, including mode collapse, instability, and sensitivity to hyperparameters. Considering these challenges alongside emerging ethical concerns related to bias, fairness, and potential misuse, further research and responsible development are necessary. Exploring new architectures, hybrid models, explainable GANs, and robust evaluation metrics is essential to advance state-of-the-art developments and fully realize the potential of synthetic data generation by GANs.

GANs represent a revolutionary shift in the approach to the generation and utilization of data. Their ability to learn from data and produce synthetic samples that closely mimic original samples has opened vast possibilities across multiple domains. This capability could easily solve many long-standing problems in data-driven research and development. We hope that continued progress in this area will occur, and that GANs will become increasingly involved in the future of AI, helping to facilitate discoveries and inspire innovation in general, ultimately contributing to the development of more resilient, equitable, and beneficial AI systems. Ongoing efforts to address remaining challenges and explore new frontiers in GAN research are critical for unlocking the full potential of this transformative technology and ensuring it serves a positive purpose in society. The journey of GANs from a theoretical concept to a practical tool for generating synthetic data reflects the power of innovation and relentless pursuit to advance the boundaries of artificial intelligence.

References

[1] K. H. Bhuiyan, S. Ahmed, and I. Jahan, "Consumer attitude toward using artificial intelligence (AI) devices in hospitality services," *Journal of Hospitality and Tourism Insights,* vol. 7, no. 2, pp. 968-985, 2024. https://doi.org/10.1108/JHTI-08-2023-0551

[2] P. Roy, B. Ghose, P. K. Singh, P. K. Tyagi, and A. Vasudevan, "Artificial intelligence and finance: A bibliometric review on the trends, influences, and research directions," *F1000Research,* vol. 14, p. 122, 2025.

[3] A. M. Al-Ghaili *et al.*, "A review: Image processing techniques' roles towards energy-efficient and secure IoT," *Applied Sciences,* vol. 13, no. 4, p. 2098, 2023. https://doi.org/10.3390/app13042098

[4] S. Sinha and S. B. Verma, "A systematic review on generative adversarial networks (GANs) for biometrics," *Library Progress International,* vol. 44, no. 3, pp. 8182–8149, 2024.

[5] A. Figueira and B. Vaz, *Survey on synthetic data generation, evaluation methods and GANs*. Porto, Portugal: CRACS-INESC TEC, University of Porto, 2022.

[6] P. N. Kiss, "Generative adversarial networks (GANs) for synthetic data generation: Applications and challenges," *International Journal of Statistical Computing and Simulation,* vol. 15, no. 1, pp. 1–21, 2023.

[7] I. Goodfellow *et al.*, "Generative adversarial nets," *Advances in Neural Information Processing Systems,* vol. 27, pp. 2672–2680, 2014.

[8] A. Goncalves, P. Ray, B. Soper, J. Stevens, L. Coyle, and A. P. Sales, "Generation and evaluation of synthetic patient data," *BMC Medical Research Methodology,* vol. 20, no. 1, p. 108, 2020. https://doi.org/10.1186/s12874-020-00977-1

[9] U. B. Bhadange, S. Jadhav, B. Jadhav, S. Ghatol, and P. Kahale, "Comprehensive review of synthetic data generation techniques and their applications in healthcare, finance, and marketing," *International Journal of Advanced Research in Science, Communication and Technology,* vol. 4, no. 1, p. 379, 2024.

[10] V. C. Pezoulas *et al.*, "Synthetic data generation methods in healthcare: A review on open-source tools and methods," *Computational and Structural Biotechnology Journal,* vol. 23, pp. 2892-2910, 2024. https://doi.org/10.1016/j.csbj.2024.07.005

[11] F. Ramzan, C. Sartori, S. Consoli, and D. Reforgiato Recupero, "Generative adversarial networks for synthetic data generation in finance: Evaluating statistical similarities and quality assessment," *AI,* vol. 5, no. 2, pp. 667-685, 2024. https://doi.org/10.3390/ai5020035

[12] M. Goyal and Q. H. Mahmoud, "A systematic review of synthetic data generation techniques using generative AI," *Electronics,* vol. 13, no. 17, p. 3509, 2024. https://doi.org/10.3390/electronics13173509

[13] V. Keerthana and S. Boopathi Raja, "GANs for synthetic data generation: Advancements and challenges using machine learning," *International Journal of Advanced Research in Innovative Ideas and Technology,* vol. 10, no. 1, pp. 285-293, 2024.

[14] D. C. Asimopoulos, M. Nitsiou, L. Lazaridis, and G. F. Fragulis, "Generative adversarial networks: A systematic review and applications," *SHS Web Conf.,* vol. 139, p. 03012, 2022. https://doi.org/10.1051/shsconf/202213903012

[15] M. H. Akpinar *et al.*, "Synthetic data generation via generative adversarial networks in healthcare: A systematic review of image- and signal-based studies," *IEEE Open Journal of Engineering in Medicine and Biology,* vol. 6, pp. 183-192, 2025. https://doi.org/10.1109/OJEMB.2024.3508472

[16] B. Vaz and Á. Figueira, "GANs in the panorama of synthetic data generation methods," *ACM Transactions on Multimedia Computing, Communications, and Applications,* vol. 21, no. 1, p. Article 3, 2024. https://doi.org/10.1145/3657294

[17] A. Aggarwal, M. Mittal, and G. Battineni, "Generative adversarial network: An overview of theory and applications," *International Journal of Information Management Data Insights,* vol. 1, no. 1, p. 100004, 2021. https://doi.org/10.1016/j.jjimei.2020.100004

[18] I. H. Rather and S. Kumar, "Generative adversarial network based synthetic data training model for lightweight convolutional neural networks," *Multimedia Tools and Applications,* vol. 83, no. 2, pp. 6249-6271, 2024. https://doi.org/10.1007/s11042-023-15747-6

[19] Y. Lu, D. Chen, E. Olaniyi, and Y. Huang, "Generative adversarial networks (GANs) for image augmentation in agriculture: A systematic review," *Computers and Electronics in Agriculture,* vol. 200, p. 107208, 2022. https://doi.org/10.1016/j.compag.2022.107208

[20] E. Brophy, Z. Wang, Q. She, and T. Ward, "Generative adversarial networks in time series: A systematic literature review," *ACM Comput. Surv.,* vol. 55, no. 10, p. Article 199, 2023. https://doi.org/10.1145/3559540

[21] L. Michelutti *et al.*, "Generative adversarial networks (GANs) in the field of head and neck surgery: Current evidence and prospects for the future—a systematic review," *Journal of Clinical Medicine,* vol. 13, no. 12, p. 3556, 2024. https://doi.org/10.3390/jcm13123556

[22] L. Chen *et al.*, "Generative adversarial synthetic neighbors-based unsupervised anomaly detection," *Scientific Reports,* vol. 15, no. 1, p. 16, 2025. https://doi.org/10.1038/s41598-024-84863-6

[23] C. Little, M. Elliot, R. Allmendinger, and S. S. Samani, "Generative adversarial networks for synthetic data generation: A comparative study," *arXiv preprint arXiv:2112.01925,* 2021.

[24] D. Garcia Torres, "Generation of synthetic data with generative adversarial networks," Degree Project, KTH Royal Institute of Technology, Stockholm, Sweden, 2018.

[25] Y. Lu *et al.*, "Machine learning for synthetic data generation: A review," *arXiv preprint arXiv:2302.04062,* 2023.

[26] Z. U. Rahman, M. S. M. Asaari, H. Ibrahim, I. S. Z. Abidin, and M. K. Ishak, "Generative adversarial networks (GANs) for image augmentation in farming: A review," *IEEE Access*, vol. 12, pp. 179912-179943, 2024. https://doi.org/10.1109/ACCESS.2024.3505989

[27] P. Sharma, M. Kumar, H. K. Sharma, and S. M. Biju, "Generative adversarial networks (GANs): Introduction, taxonomy, variants, limitations, and applications," *Multimedia Tools and Applications*, vol. 83, no. 41, pp. 88811-88858, 2024. https://doi.org/10.1007/s11042-024-18767-y

[28] X. Guo and Y. Chen, "Generative ai for synthetic data generation: Methods, challenges and the future," *arXiv preprint arXiv:2403.04190*, 2024.