STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 13, Issue 1, 2026
January–April
DOI Journal: 10.37591/JoAIRA

https://journals.stmjournals.com/joaira

Review JoAIRA

A Comparison of Different Generative AI Models

Arya Jayendra Bandal[1,*], Samiya Kasid Attar[1], Sakshi Avinash Badak[1]

Abstract

Generative models have significantly advanced the field of artificial intelligence by allowing machines to produce complex and realistic outputs such as images, text, and other forms of data. Among the leading frameworks in this domain are generative adversarial networks (GANs), variational autoencoders (VAEs), and architectures based on Transformers. Each model offers specific benefits and drawbacks concerning design structure, training demands, and range of applications. This paper provides a detailed comparison of these generative techniques, with a focus on core aspects like accuracy, computational cost, and usability in real-world scenarios. We delve into the foundational concepts behind each model, evaluate their performance using widely accepted metrics, and analyze their effectiveness across tasks such as image generation, language modeling, and anomaly identification. The study outlines the trade-offs between flexibility, robustness, and scalability, aiming to guide practitioners in choosing the best-suited model for particular use cases. Finally, the paper discusses prospective research pathways to further enhance the power and versatility of generative models.

Keywords: Generative models, generative adversarial networks (GANs), variational autoencoders (VAEs), Transformers, neural networks, artificial intelligence, image synthesis, machine learning models, computational efficiency, likelihood estimation, performance evaluation

INTRODUCTION

A class of artificial intelligence (AI) models is designed to create new data that is based on the data used for training that they have learned from, called generative models. Applications of these models have spread widely in the fields of image generation, text synthesis, and data augmentation. Generative models learn the underlying patterns from the data and generate new samples that resemble real-world examples, whereas traditional machine learning models focus on classification and prediction.

Generative adversarial networks (GANs), variational autoencoders (VAEs), and transformer-based models are the most prominent generative models. In GANs, the generator and discriminator are two neural networks. These two networks are trained in opposition, which results in the creation of realistic data by the generator, and the discriminator distinguishes real and generated samples. Data are converted into a latent space using probabilistic techniques and then decoded back into a new representation in VAEs. In recent years, transformer-based models have gained popularity because they rely on self-attention mechanisms to generate coherent and contextually rich sequences, making them particularly effective in text and image generation [1].

Every generative model has its own advantages and disadvantages, but by accurately choosing the compatible model, generative tasks can be completed. GANs excel at producing visually

***Author for Correspondence**
Arya Bandal
E-mail: aryabandal257@gmail.com

[1]Student, Department of Electronics and Telecommunication, Rajgad Dnyanpeeth's Shri Chhatrapati Shivaji Maharaj College of Engineering, Savitribai Phule Pune University, Maharashtra, India

Received Date: April 12, 2025
Accepted Date: September 11, 2025
Published Date: February 19, 2026

Citation: Arya Jayendra Bandal, Samiya Kasid Attar, Sakshi Avinash Badak. A Comparison of Different Generative AI Models. Journal of Artificial Intelligence Research & Advances. 2026; 13(1):16–22p.

convincing images but can suffer from instability during training. VAEs provide structured latent spaces that help in the interpolation and manipulation of data, although they often generate blurrier samples. Transformers have revolutionized AI applications, especially in natural language processing and image generation, but they require large amounts of computational resources [2].

The main objective of this study is to compare different generative AI models, such as GANs, VAEs, and Transformers, based on their performance, efficiency, and applications. By analyzing the strengths and weaknesses, we aim to find the most compatible model for different generative tasks. In addition, this paper highlights recent advancements in generative AI and discusses future research directions to further improve these models [3].

BRIEF HISTORY OF GENERATIVE MODELS

The development of generative models in AI has progressed over several years with significant advancements in probabilistic modeling, deep learning, and neural networks. Early generative models were based on statistical methods, but the introduction of deep learning has revolutionized the field, enabling models to generate highly realistic data [4].

Early Generative Models (Pre-2000s)

The foundations of generative models are Gaussian mixture models (GMMs) and hidden Markov models (HMMs), which are probabilistic and statistical methods. The primary uses of these models include speech recognition, natural language processing, and other structured data tasks. Another early approach was the Boltzmann Machine, introduced in the 1980s, which laid the groundwork for energy-based generative models [5].

Rise of Neural Network-Based Generative Models (2000s–2010s)

In the early 2000s, the development of more powerful generative models was possible only because of the emergence of deep learning. Hinton et al. introduced the first deep learning-based generative models, named restricted Boltzmann machines (RBMs) and deep belief networks (DBNs). These models have limited scalability but assured improved feature learning [6].

The introduction of variational autoencoders (VAEs) in 2013 by Kingma and Welling [1] was a major breakthrough. VAEs provide a probabilistic approach for learning latent representations to make them useful for generating structured and interpolatable data. In 2014, around the same time, Ian Goodfellow et al. [3] introduced GANs. GANs revolutionize image generation with the help of adversarial training, where a generator competes against a discriminator to generate highly realistic samples.

Transformers and the Modern Era of Generative AI (2017 to Present)

In 2017, Vaswani et al. [4] introduced transformer architecture, which resulted in another turning point in generative AI. Models such as generative pre-trained transformer (GPT), BERT, DALL·E, and Stable Diffusion, which were designed for natural language processing, Transformers demonstrated their capability in text as well as image generation. Unlike VAEs and GANs, Transformers rely on self-attention mechanisms that allow them to capture long-range dependencies in data.

In today’s era, generative models are widely used in various applications such as creative AI, image generation, text synthesis, and drug discovery. These fields continue to evolve, along with researchers working on various aspects such as improving model efficiency, enhancing interpretability, and reducing biases.

OVERVIEW OF PREVIOUS COMPARATIVE STUDIES

Several studies have been conducted to compare different generative models, with a focus on their performance, architecture, and applications. GANs, variational autoencoders (VAEs), and transformer-based models have been analyzed by various researchers based on key factors such as computational efficiency, image quality, training stability, and real-world usability [7].

The first widely used deep generative models were GANs and VAEs; hence, comparative studies have primarily focused on them. Sharper and more realistic images are produced by GANs, but they suffer from mode collapse and training instability. In contrast, VAEs provide structured latent representations, making them useful for interpolation and feature learning, although the generated samples are often blurred.

Transformers have demonstrated superior performance in text and image generation tasks, as inferred in recent studies. Comparisons between GANs and Transformers suggest that GANs excel at generating high-resolution images, whereas transformer-based models (such as DALL·E and Stable Diffusion) are better suited for controllable and diverse content generation. To achieve optimal results, Transformers require high computational power and large-scale datasets.

The introduction of hybrid approaches, combining elements of VAEs, GANs, and Transformers in some studies, tends to overcome their individual limitations. For instance, VQ-VAE-2 integrates VAE-based encoding with transformer-based sequence modeling, leading to an improved generation quality while retaining a structured latent space.

Overall, previous research highlights the trade-offs between efficiency, quality, and computational demands across these models. Although GANs remain dominant in high-quality image synthesis, VAEs are preferred for structured generative tasks, and Transformers have redefined text and multimodal generation. This study builds on these comparisons by providing a more detailed evaluation of their strengths, weaknesses, and suitability for various applications.

OVERVIEW OF GENERATIVE MODELS

To generate new data similar to training data, a class of AI techniques called generative models was designed. These models are useful for tasks such as text generation, image synthesis, and data augmentation. These models also learn the underlying data distributions and create new samples. Well-known generative models include GANs, variational autoencoders (VAEs), and transformer-based models [8].

Generative Adversarial Networks

Ian Goodfellow introduced GANs in 2020 [3], which consist of two neural networks:

- *Generator*: This helps in creating synthetic data samples.
- *Discriminator*: This helps in identifying whether the generated samples are real or fake.

Here, the generator produces realistic samples, and the discriminator distinguishes real data from the generated ones. Hence, these two networks compete in adversarial processes. Over multiple iterations, the generator learns to create highly realistic outputs [9].

Applications of GANs

- Image synthesis (e.g., generating photorealistic images).
- Style transfer (e.g., transforming images into artistic styles).
- Data augmentation (e.g., generating synthetic data for training models).

Variational Autoencoders

VAEs, introduced by Kingma and Welling in 2013 [1], are built on an encoder-decoder architecture:

- *Encoder*: The input data is compressed into a lower-dimensional latent space
- *Decoder*: Reconstructs data from this latent representation.

A key feature of VAEs is their probabilistic latent space representation, which allows for controlled data generation and interpolation. However, VAE-generated images often appear blurrier than GANs because of their design.

Applications of VAEs

- Data compression (reducing the size of large datasets).
- Anomaly detection (identifying outliers in data).
- Medical imaging (generating synthetic medical scans for analysis).

Transformers (GPT, DALL·E, Stable Diffusion)

In 2017, Vaswani et al. [4] introduced Transformers, which are a class of models that are based on the concept of a self-attention mechanism, allowing them to capture complex relationships between a wide range of data points efficiently. Unlike GANs and VAEs, Transformers do not rely on adversarial training or explicit latent spaces; instead, they use attention mechanisms to process large data sequences [10].

Modern transformer-based generative models undergo large-scale pre-training on massive datasets, enabling them to generate highly coherent text, images, and multimodal content. Notable examples include the following.

- GPT models for text generation (e.g., chatbots, content creation).
- DALL·E and Stable Diffusion for AI-generated art and image synthesis.

Applications of Transformers

- Text generation (e.g., automated content writing, chatbots).
- AI art (e.g., generating creative images from text descriptions).
- Multimodal AI (e.g., models that generate both text and images).

Different applications are possible because each generative model has unique strengths. While GANs excel at generating high-quality images, VAEs provide structured latent representations that are useful for feature learning, and Transformers dominate large-scale text and multimodal generation.

COMPARISON CRITERIA

A comparison of GANs, variational autoencoders (VAEs), and transformer-based models must be based on several key factors. These include performance metrics, training complexity, computational requirements, and real-world applications.

Performance Metrics

To assess the quality and accuracy of the generated samples, it is necessary to evaluate generative models. The specific metrics are as follows:

- *Fréchet inception distance (FID score):* The goal was to accurately measure the similarity between the generated and real images. Lower FID scores indicated better quality. They are primarily used for GANs and VAEs.
- *Likelihood estimation*: The data distribution is evaluated using this metric and is generally used for VAEs. Better generalization is achieved when the likelihood is higher.
- *Perplexity*: Used for text-based models such as Transformers, measures how well a model predicts a sequence. Lower perplexity indicates better fluency and coherence.

Training Complexity

Different generative models require varying levels of effort to train effectively:

- *GANs*: Owing to the adversarial process between the generator and discriminator, training is challenging. However, this model has issues such as collapse and instability.
- *VAEs*: They are easier to train compared to GANs because they rely on probabilistic encoding, but they may produce lower-quality samples.
- *Transformers*: Requires extensive pre-training on large datasets, making training time-consuming and computationally expensive.

Computational Requirements

The hardware and scalability of generative models impact their real-world usability:

- *GANs*: Requires powerful graphics processing unit (GPUs) but can be optimized for specific tasks such as image generation.
- *VAEs*: These models are less resource-intensive; hence, they are more accessible for tasks such as anomaly detection and data compression.
- *Transformers*: Demand large-scale hardware (Tensor processing units (TPUs), high-end GPUs) and significant memory owing to self-attention mechanisms and large parameter sizes.

Use Cases and Industry Adoption

Various applications are possible by using different models:

- *GANs*: Best for high-quality image synthesis, style transfer, and deepfake generation. Used in entertainment, gaming, and content creation.
- *VAEs*: Ideal for data compression, anomaly detection, and medical imaging. Applied in healthcare, cybersecurity, and scientific research.
- *Transformer*s: Most effective for text generation, AI art, and multimodal applications. It has been widely adopted in chatbots, creative AI, and natural language processing.

Different tasks are possible because each model has unique strengths and trade-offs. This comparison is helpful in selecting the right model based on various parameters such as performance, complexity, and industry needs.

STRENGTHS AND WEAKNESSES OF EACH MODEL

Each generative model, GANs, VAEs, and Transformers, has distinct advantages and limitations based on performance, efficiency, and application suitability [11].

Generative Adversarial Networks

- *Strengths*: It is adaptable for generating high-quality, realistic images. It is widely used for image synthesis, deep fake generation, and artistic style transfers.
- *Weaknesses*: Owing to adversarial learning, the risk of mode collapse (generator produces limited variations), high computational demands, and training instability.

Variational Autoencoders

- *Strengths*: It provides structured latent spaces to make them useful for controlled data generation, anomaly detection, and feature learning. It is easier to train than GANs.
- *Weaknesses*: Often, blurry or less realistic images are produced because of the nature of probabilistic modeling. However, this is not as effective for high-resolution image synthesis.

Transformers (GPT, DALL·E, Stable Diffusion)

- *Strengths*: Excellent in handling sequential and multimodal data. State-of-the-art results have been achieved in text generation, AI-driven art, and natural language processing.
- *Weaknesses*: Requires massive training data and computational resources. High memory consumption and increased environmental costs are due to energy-intensive training.

FUTURE SCOPE IN GENERATIVE AI

The evolution of generative AI is rapidly happening with several key advancements shaping the future.

- *Hybrid models:* Combining elements of GANs, VAEs, and Transformers to leverage their strengths and overcome their limitations (e.g., VQ-VAE-2).
- *More efficient training*: Research into reducing computational costs through optimized architecture, quantization, and knowledge distillation.
- *Ethical and bias reduction*: Addressing challenges related to biased data generation and misinformation by improving model transparency and fairness.
- *Interactive and personalized AI*: Models capable of generating personalized content tailored to individual user preferences.

Suggested Improvements and Future Research Directions

- *Improving training stability*: Techniques such as adversarial training improvements (e.g., Wasserstein GANs) and enhanced regularization methods can help stabilize the GANs.
- *Enhancing sample quality*: VAEs can benefit from improved latent space representations to produce sharper and more realistic output.
- *Reducing computational costs*: Exploring lightweight architectures and more efficient self-attention mechanisms in Transformers to make them accessible for broader applications.
- *Ethical considerations*: Developing robust methods to detect and prevent the misuse of generative AI, particularly in deepfakes and misinformation.

As generative AI continues to advance, future research should focus on making models more efficient, accessible, and ethically responsible, thereby ensuring broader adoption in various industries.

CONCLUSION

This study compared GANs, VAEs, and transformer-based models based on their performance, training complexity, computational requirements, and real-world applications. GANs excel in high-quality image generation; VAEs are useful for structured latent representation, and Transformers dominate text and multimodal generation. However, each model presents trade-offs in terms of stability, scalability, and efficiency.

Declaration of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this study.

Acknowledgement

The authors would like to express their gratitude to researchers and contributors in the field of generative AI, whose work laid the foundation for this study. We also extend our appreciation to our mentors, peers, and the academic community for their valuable insights and discussions on GANs, VAEs, and Transformers. Additionally, we acknowledge the resources, datasets, and tools that facilitated this study. Finally, we thank our family and friends for their continuous support and encouragement throughout this study.

REFERENCES

1. Kingma DP, Welling M. Auto-encoding variational Bayes. [Preprint]. 2013. arXiv:1312.6114. doi:10.48550/arXiv.1312.6114.
2. Saul LK, Weiss Y, Bottou L. Advances in Neural Information Processing Systems 17: Proceedings of the 2004 Conference. Cambridge (MA): MIT Press; 2005.
3. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM. 2020;63(11):139–144. doi:10.1145/3422622.
4. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: von Luxburg U, Guyon I, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017). Red Hook (NY): Curran Associates Inc.; 2017. p. 5998–6008.
5. Ahmad Z, Jaffri ZA, Chen M, Bao S. Understanding GANs: fundamentals, variants, training challenges, applications, and open problems. Multimed Tools Appl. 2025;84(12):10347–10423. doi:10.1007/s11042-024-19361-y.
6. Wang J, Yang C, Xu Y, Shen Y, Li H, Zhou B. Improving GAN equilibrium by raising spatial awareness. In: Proc IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR). 2022. p.11275–11283. doi:10.1109/CVPR52688.2022.01100.
7. Odaibo S. Tutorial: deriving the standard variational autoencoder (VAE) loss function. [Preprint]. 2019. arXiv:1907.08956. doi:10.48550/arXiv.1907.08956.
8. Thrun S, Saul L, Schölkopf B. Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference. Cambridge (MA): MIT Press; 2004. p.47–110.

9. Margossian CC, Blei DM. Amortized variational inference: when and why? [Preprint]. 2023. arXiv:2307.11018. /doi:10.48550/arXiv.2307.11018.
10. Zhang C, Bütepage J, Kjellström H, Mandt S. Advances in variational inference. IEEE Trans Pattern Anal Mach Intell. 2018;41(8):2008–2026. doi:10.1109/TPAMI.2018.2889774.
11. Hacker P, Engel A, Mauer M. Regulating ChatGPT and other large generative AI models. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23); 2023 Jun 12–15; Chicago, IL, USA. New York (NY): Association for Computing Machinery; 2023. p. 1112–1123. doi:10.1145/3593013.3594067.

