

State-of-the-art review and critical analysis of emerging trends in generative artificial intelligence

Jyoti Gupta *

Department of Electronics and Communication Engineering

Bharati Vidyapeeth's College of Engineering

New Delhi 110063

India

Monica Bhutani [†]

Department of Electronics and Communication Engineering

Bharati Vidyapeeth's College of Engineering

New Delhi 110063

India

Mahesh Kumar [§]

Department of Information Technology Engineering

Bharati Vidyapeeth's College of Engineering

New Delhi 110063

India

Aman Dureja [‡]

Department of Information Technology

Bhagwan Parshuram Institute of Technology

New Delhi 110089

India

* E-mail: jyotig1989@gmail.com (Corresponding Author)

[†] E-mail: monica.bhutani@bharatividyaapeeth.edu

[§] E-mail: maresh.kumar1@bharatividyaapeeth.edu

[‡] E-mail: amandureja@gmail.com

Shyla Singh [@]

Mohit Dayal [#]

*Department of Information Technology Engineering
Bharati Vidyapeeth's College of Engineering
New Delhi 110063
India*

Abstract

Generative AI technology now leads the way as a transformative innovation that produces diverse, realistic content across various content modalities. This research extensively reviews generative AI by examining its core mechanisms, architectural improvements, and current breakthroughs in making images, text, audio, and videos. The paper discusses three main generative models, GANs, VAEs and diffusion models, to explain their distinctive advantages and technical constraints. The document demonstrates industrial uses in different sectors yet focuses on essential issues connected to biased data, unintelligible model configurations, and efficiency constraints. The paper focuses on upcoming developments of multimodal systems and environmentally friendly AI practices. The main emphasis lies in building generative models that demonstrate robustness, interpretability, and sustainability across the environment. The paper explores research directions for the future with a specific focus on integrating multiple modalities and establishing ethical practices during model design. Researchers, practitioners, and policymakers will find this review highly helpful in their tasks related to understanding and developing generative AI technologies.

Subject Classification: 68T05.

Keywords: *Generative AI, Deep learning, Neural networks, Multimodal generation, Diffusion models, Artificial intelligence, Machine learning, Sustainable AI.*

1. Introduction

Generative AI draws its roots from the early days of artificial intelligence research, explicitly involving probabilistic methods and representation learning. However, the arrival of deep learning technology and high-powered computational resources allowed generative AI to develop quickly. The first success stories were generative adversarial networks (GANs) and variational autoencoders (VAEs), showcasing impressive abilities to generate realistic images [1]. The intervening years have seen an explosion of development and research, with models becoming more advanced and the application space expanding.

[@] E-mail: shylasinghit@gmail.com

[#] E-mail: mohitdayal.md@gmail.com

Although considerable progress has been achieved, generative AI faces several challenges and research gaps. One major challenge is developing multi-modal models that produce high-fidelity, diverse, coherent samples across modalities. Moreover, there is great urgency in addressing the ethics of powered AI, including but not limited to concerns about biases and privacy alongside misuse issues. Considering the importance of improving trust and responsible development of AI, providing model interpretability will be crucial. In short, this paper purports to be the necessity of offering a complete perspective on current generative AI advancements and highlighting a few major recognized

Table 1
Comparing GANs, VAEs, Normalizing Flows, and Diffusion Models, highlighting their key components, training processes, and advantages/ disadvantages.

Feature	GANs	VAEs	Normalizing Flows	Diffusion Models
Key Components	Generator, Discriminator	Encoder, Decoder, Latent Space	Invertible Transformations	Forward and Reverse Diffusion Processes
Training Process	Adversarial Training	Variational Inference	Maximum Likelihood Estimation	Denoising Score Matching
Advantages	High-quality image generation, versatility	Probabilistic interpretation, smooth latent space	Exact log-likelihood, invertibility	High-quality and diverse samples
Disadvantages	Training instability, mode collapse	Blurry samples, lower quality	Computationally expensive	Slow sampling, complex training
Notable Variants/ Improvements	DCGAN, StyleGAN, BigGAN	Beta-VAE, VQ-VAE	Real NVP, Glow	DDPM, Latent Diffusion Models
Applications	Image generation, super-resolution, inpainting	Anomaly detection, image denoising, clustering	Density estimation, data compression	Text-to-image synthesis, image generation

research areas indicating potential for future deployment. A detailed analysis of the fundamental aspects, latest developments, applications and hurdles in generative AI has been attempted to provide an extensive overview of this ever-growing field, motivating further research and advancements.

Neural networks, particularly the deep learning architectures that some form, have transformed this field by providing highly flexible tools to learn complex patterns from data. These networks, which consist of multiple layers among neuronal connections computing derived features from the input data, act as functions that can learn these higher-level representations and treat them to model complex distributions. New data points can be produced by sampling from the latent space. Normalizing flows are when you can calculate the log density and generate samples from a distribution but want to embed it in something more powerful [2].

Table 1 compares GANs and VAEs with Normalizing Flows and Diffusion Models. It provides significant aspects of those respective containers, such as their training, benefits or disadvantages. GANs: The GAN framework contains a generator and discriminator in an adversarial environment to generate life-like data. However, the problem is that it creates great images and mostly unstable training. While VAEs encode data to a latent space, they decode data back, enabling smooth interpolation; however, the generated outputs are less realistic. Normalizing Flows (which use invertible transformations) have exact likelihood computation but are computationally expensive. Diffusion models progressively denoise data that produces high-quality samples with fast training and slow generation times.

2. Literature Review

Diffusion models have arisen as a potential solution to these problems. Naturally, they generate the image (i.e., gradient) that will make a data point less likely in the training distribution by injecting noise into some location starting from its original position and turning it back to random white pixels [3]. We show that diffusion models can also produce high-fidelity images that outperform prior art. More recently, Latent Diffusion Models (LDMs) have sustained high-resolution details in these images of unprecedented quality that look straight out of the real world.

Combining the knowledge of a language to generate images in multiple ways has brought new opportunities. For instance, there are text-to-image diffusion models [4], able to produce a very realistic image from

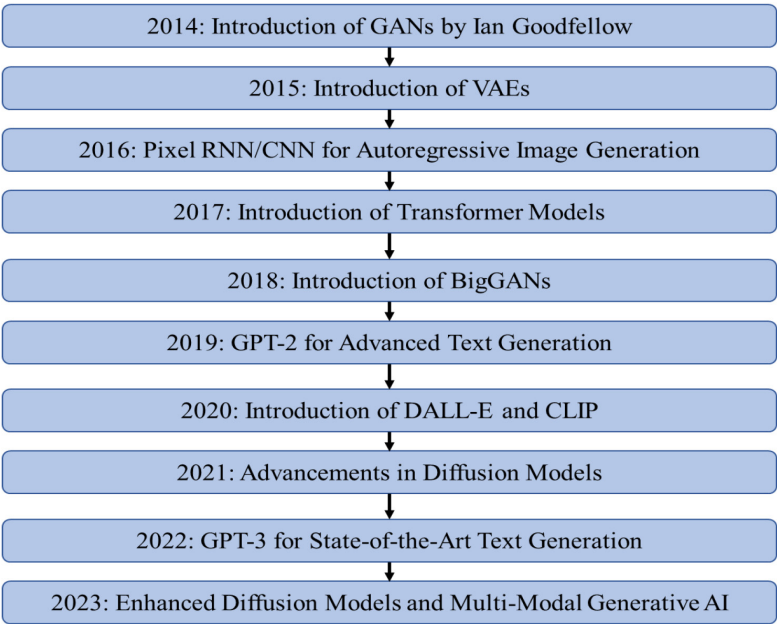


Figure 1
Key Advancements in Generative AI

the written description, which helps narrow the ambiguity between natural language and Visual representation. And that opens up many possibilities for many applications, from content generation to design and virtual reality. Despite these impressive advancements, several challenges remain, such as image quality, diversity, and controllability. Looking to the future, more efficient and stable generative models (if not an entirely new methodology) may be better technically suited for generating images of high quality or some out-of-the-box applications. Moreover, it is also his opinion about the deepfakes that any human minds behind the creation of these images are unfeasible for image generation to be sussed out completely and whether introducing something equivalent in hardware is possible.

Another stream of revolutionizing AI-based technology has been made in the area of generating texts, many of which can seem similar to those created by humans. A standard method to generate such text is utilizing a generative model, which learns the statistical patterns and relationships between words from a large corpus and then applies this gained knowledge to create new coherent text, from realistic news and articles to creative poems as shown in the Figure 1. In particular, text

generation via generative AI models made a significant leap forward with transformers [5], a specific type of neural network architecture. Such networks can adequately account for long-range dependencies within text, thus producing grammatically correct and semantically meaningful updates. In addition, Generative Pre-trained Transformers (GPT) models have shown outstanding results in bringing text generation to the next level and have demonstrated the technology to generate text of human-like quality that is often indistinguishable from human inputs. Moreover, conditional text generation allows models to generate text from a specific prompt, style, or topic.

Performance remains a problem even for more recent methods; key challenges include fidelity, timbre, style control, and diversity in generated audio. Dedicated future research should, therefore, explore new ways of dealing with the Design Aperture associated with audio data in general, albeit also excitingly across different machine learning structures and persevering controversies over rights due to copyright hustle, faux realities, and ethical issues concerning deepfakes. Another area in which the frontier of generative AI has dramatically expanded is video generation, as evidenced through recent years. Based on image and voice control, researchers have explored methods capable of generating high-quality video from multiple sources, from text descriptions to images and even other videos. In earlier methods, generation would scan across several consecutive frames by consecutive endeavours, but temporal continuity and performance efficiency were bottlenecked.

Recently, however, there have been advancements in video generation models. Recent years have witnessed a surge in progress within the field of video synthesis with innovations such as VideoGAN and ProGAN-Video using Generative Adversarial Networks (GANs) [6], which are capable of synthesizing diverse and authentic-seeming sequences consistently. Variational autoencoders (VAEs) have also been investigated for video generation, giving an expressive probabilistic model over the space of possible video data.

While these progress outstandingly in producing high-quality, long-range and/or semantically coherent videos, ensuring the generated videos are temporally consistent and diverse, work is also active research. With the maturation of video technology, vast possibilities exist for use in entertainment, education and VR applications.

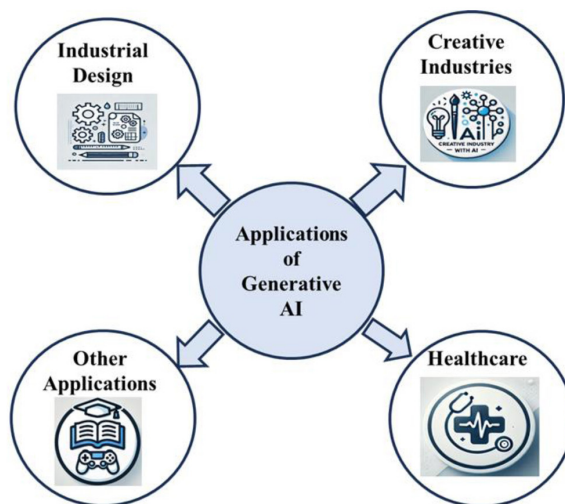


Figure 2
Applications of Generative AI

3. Applications of Generative AI

Generative Art, Design and Entertainment based on Figure 2. The role of generative AI in generating different kinds or styles of art to automate the design process and content generation [7]. In marketing, for example, generative AI can create catchy content that drives traffic to a website, personalize some campaigns or at least promote the product in different parts of an ad (co-constraints), and optimize strategies using natural language processing models. Generative AI is poised to transform the healthcare industry by helping speed up drug discovery by creating previously non-existent molecules that perform better than existing alternatives, enhancing medical image analysis for earlier disease detection and providing patient-specific data-driven personalisation of medicine. In particular, industrial design can use generative AI to streamline product designs, experiment with new form factors and speed up prototyping. However, aside from these sectors, generative AI is also improving education by offering tailored learning experiences for students and transforming gaming with dynamic content generation, among other fields. The applications of generative AI continue to grow as a pillar for innovative industries.

4. Challenges and Limitations

Generative AI is promising but also troublesome. These models are usually designed to learn from data, and that same being, the lifeblood can be biased, leading to drifting into amplified bias when generating outputs. All of this can reinforce dangerous stereotypes and inequalities. However, the enormous computational resources we need to train and run these models together are one of its biggest problems since those criteria prevent them from being accessible to all users or simply not serious for long-term work. Not to mention the profound ethical implications. Think of how easy it would be for someone to create deep-fake style machine-learned photographs, poorly imitated voices, potential privacy invasions into people's homes or other breaches such as infringing on copyrights [8]. Finally, the opacity of these models (black box issue) makes them difficult to insert in society. As our generative AI algorithms get more innovative, transparency and fairness become more critical.

Solving these challenges is crucial to unlock the power of generative AI. Well-thought data cleaning and preprocessing methods with an ethical essence are the need of the hour. Efficient algorithms and hardware optimizations can reduce computational constraints. Building transparent models in terms of code, logic, and moral principles to interpret them can increase trust while ensuring proper usage. Addressing these issues can help lay the groundwork for a growing generative AI that brings social value without eroding its values.

5. Future Prospects of Generative AI

Improving model robustness is a key part of moving the needle on generative AI. In this context, robustness means a model should still perform adequately under different and likely harsh conditions. This includes robustness against adversarial attacks, distribution shifts in data and other disturbances. Although progress has not been negligible, there is a fundamental challenge to making such advanced generative models, as discussed in [9], reliable for use cases that work best in practice.

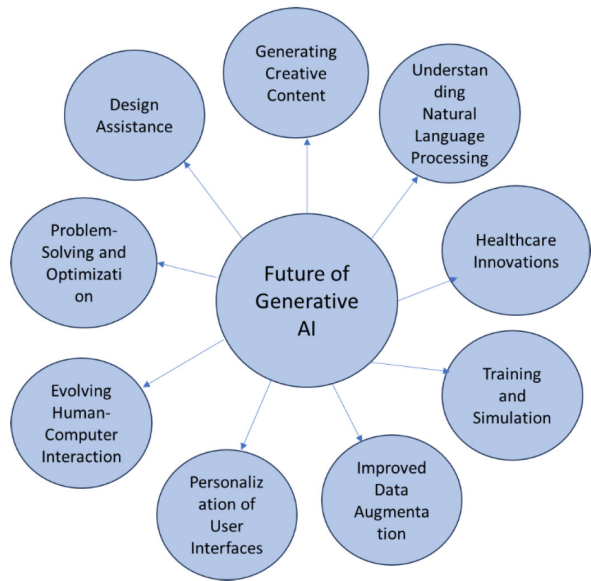


Figure 3
Future Prospects of Generative AI.

The most exciting option for improving robustness is training with adversarial examples, which means that models see “fake” images during the training to make them immune. This means they also become more resistant to attempts at “poisoning” them with adversarial inputs besides succeeding in telling whether an example is real. Further, as detailed in [10], we can also investigate methods to increase the interpretability of models, which will help us better understand vulnerabilities and craft tailored defences. Additionally, the study of data augmentation and a little more diversity could provide us with better models in return. Exposing models to more variable data, even in the form of synthetic examples, can help make them less vulnerable to degradation when presented with unknown data distributions. It also uses techniques like domain adaptation and transfer learning to improve the generalization capabilities of the model.

A critical frontier in generative AI is improving more interpretable and explainable models. Although these models have proven successful in generating photorealistic outputs, their black-box decision-making process makes it difficult for humans to trust and use them as illustrated in Figure 3. One solution to this problem is that researchers are trying out different

ways of interpreting generative models. One solution is to work on tools that can make the internal representations of these models more transparent and understandable. Inspection of latent spaces & features Researchers can also learn how the model processes information and outputs. Moreover, in [11], attention mechanisms are further used for surfacing the inputs that have a high bearing on a particular production, thereby providing partial explainability. A second encouraging one is offering post-hoc explanation techniques. Explainability methods aim to produce human-interpretable reasons for the model's prediction after training. With reason, methods such as LIME [12] and SHAP have successfully explained sophisticated models like deep neural nets. The use of these techniques in generative models can give a strong signal about how the model makes decisions and what biases or errors are related to it.

6. Emerging Trends and Unrevealed Research Directions

Generative AI is a fluid discipline where whole swathes are new frontiers. One of the trends is that generative models and other AI paradigms are converging. Combining, for example, reinforcement learning with generative models shows promise of creating intelligent agents that can act in a goal-directed manner autonomously. Second, causality is the key to unlocking many other areas of science and social sciences, e.g. it leads the way for counterfactual reasoning in economics or even basic research questions like what causes a particular disease. Meanwhile, multimodal generative models are also gaining momentum for generating various classes of data jointly, and their uses include virtual reality or drug discovery. As concerns arise regarding generating AI models, especially in the generative domain, it is necessary to research and develop fair, accountable, and visible integrity of generated models [13]. In the end, much of the real power and future opportunity for generative AI is harnessing it in service to human cognition and creativity, enabling new applications that can ultimately benefit society [14].

6.1 *Multimodal Generative Models*

Multimodal generative models such as this are an essential and emergent frontier in generative AI; those that can handle multiple data types like text, image, and audio. They can fuse different sensory modalities that hold promise to transform many fields. Multimodal generative models can learn complex relations between these data types and discover

new ideas, providing pathways to never-before-seen innovation. For instance, the technology of text-to-image generation has shown significant progress in creating almost photorealistic images based on textual description. Multimodal models can also be used in a task like a video synthesis when we want to generate proper content (both visual and auditory), given some textual prompts. This is a massive development for the entertainment, learning and advertising sectors.

One of the key challenges for multimodal models that generate data is to develop 'aligned' datasets where there are few-to-no datasets, often massive in scale, that span a variety of modalities. Manually going through, curating and annotating such datasets is an arduous process. Aligning the latent spaces of modalities is, in addition to being costly and complex given how disparate embeddings could be, rather nontrivial, resulting in much more advanced methods required to make suitable representations of them. For instance, deep learning architectures like transformers and convolutional neural networks have shown the potential to catch fine-grained correlations among different modalities. Unfortunately, combining these representations remains a tricky open research question.

Now comes the task of evaluating multi-modal state-space analyzers. Many evaluation metrics have been proposed for static images and video, but these are primarily designed in the context of a single modality and work scantily when it comes to multimodal settings as they tacitly depict only intrinsic quality. In the future, it is essential to define how to measure each modality and its interplay with other factors. Multi-modal generative models are challenging to interpret as well. Deciphering how these models arrive at decisions, in terms of input types and output specificity there are often complex interactions to this level.

However, the payoff for multimodal generative models can also be very high. In healthcare, these models help in medical image analysis, drug discovery and patient diagnosis by incorporating medical images, patients' records, and genetics. For example, in the entertainment industry, we see the use of multimodal generative models towards creating immersive virtual and augmented reality experiences, including realistic characters, environments and interactive narratives. Moreover, these models can transform education by offering personalized learning experiences per student while integrating different learning materials (text/images/videos).

6.2 Sustainable AI Practices

As the field of generative AI machines that can create human-like text and images continues to grow, it stands on a collision course with one of our most pressing environmental challenges. Training big generative models is a computationally expensive process that requires enormous energy, generating significant carbon emissions. This ecological cost is a major threat to the sustainability of AI, and tackling it will be key to its sustainable growth.

In search of a way to reduce the carbon footprint from generative AI, research and practices continue. This includes introducing an energy-aware line of model architectures and training algorithms. The computation of power-hungry models can be reduced through models like model pruning, quantization and knowledge distillation. In addition, you can keep looking for hardware alternatives like AI accelerators or energy-saving processors, which could help save some power.

Moreover, the selection of training data and the algorithms themselves can play a significant role in the environmental impact generated by generative models. Having datasets curated with the least redundancy and using energy-efficient training algorithms is another sustainable step forward. Yan adds that creating models that can learn from smaller and broader datasets can also reduce the computational load. In addition, model reuse and transfer learning can also reduce the training time for models from scratch.

7. Conclusion

This paper elaborates on the state-of-the-art on generative AI and aims to improve a deeper understanding of knowledge in this domain. It is a valuable resource for researchers, practitioners, and policy-makers who want to understand the map of generative AI better. Generative AI holds a great future. Therefore, multimodal generative models are proposed with the hope that they can concatenate the content of various data modalities as input and yield more advanced output for realistic synthesis. In addition, improvements in AI sustainability are critical to the ongoing practicality of this technology. With the continued maturation of generative AI, it should be built and used responsibly with ethical principles and societal well-being in mind. To summarise, generative AI is a transformative tech that can change the dimensions of verticals and human lives for good. And with continued innovation at breakneck speed,

that future looks bright. Solving the challenges mentioned above and investigating new directions would go a long way to unlock generative AI capabilities for researchers and practitioners that will achieve human-like performance, putting us on track toward creative machines working alongside humanity.

References

- [1] S. Balasubramaniam, V. Chirchi, S. Kadry, M. Agoramoorthy, S. P. Gururama, K. K. Satheesh, and T. A. Sivakumar, "The road ahead: Emerging trends, unresolved issues, and concluding remarks in generative AI—A comprehensive review," *Int. J. Intell. Syst.* (2024).
- [2] R. T. Hughes, L. Zhu, and T. Bednarz, "Generative adversarial networks-enabled human-artificial intelligence collaborative applications for creative and design industries: A systematic review of current approaches and trends," *Front. Artif. Intell.*, vol. 4 (2021).
- [3] S. S. Sengar, A. B. Hasan, S. Kumar, and F. Carroll, "Generative artificial intelligence: A systematic review and applications," *Multimed. Tools Appl.*, pp. 1–40 (2024).
- [4] A. Bhattacharya, R. Mittal, and A. Kumar, "A comprehensive survey on generative AI techniques and their applications across domains," *Appl. Intell.*, vol. 54, pp. 12130–12154 (2024).
- [5] R. Gozalo-Brizuela and E. C. Garrido-Merchan, "ChatGPT is not all you need. A state of the art review of large generative AI models," *arXiv preprint*, arXiv:2301.04655 (2023).
- [6] S. Bengesi, H. El-Sayed, M. K. Sarker, Y. Houkpati, J. Irungu, and T. Oladunni, "Advancements in generative AI: A comprehensive review of GANs, GPT, autoencoders, diffusion model, and transformers," *IEEE Access* (2024).
- [7] D. H. Hagos, R. Battle, and D. B. Rawat, "Recent advances in generative AI and large language models: Current status, challenges, and perspectives," *IEEE Trans. Artif. Intell.* (2024).
- [8] M. O. Ayemowa, R. Ibrahim, and M. M. Khan, "Analysis of recommender system using generative artificial intelligence: A systematic literature review," *IEEE Access* (2024).
- [9] R. Law, K. J. Lin, H. Ye, and D. K. C. Fong, "Artificial intelligence research in hospitality: A state-of-the-art review and future directions," *Int. J. Contemp. Hosp. Manag.*, vol. 36, no. 6, pp. 2049–2068 (2024).

- [10] A. D. Samala, S. Rawas, T. Wang, J. M. Reed, J. Kim, N.-J. Howard, and M. Ertz, "Unveiling the landscape of generative artificial intelligence in education: A comprehensive taxonomy of applications, challenges, and future prospects," *Educ. Inf. Technol.*, pp. 1–40 (2024).
- [11] R. Khera, E. K. Oikonomou, G. N. Nadkarni, J. R. Morley, J. Wiens, A. J. Butte, and E. J. Topol, "Transforming cardiovascular care with artificial intelligence: From discovery to practice: JACC state-of-the-art review," *J. Am. Coll. Cardiol.*, vol. 84, no. 1, pp. 97–114 (2024).
- [12] F. Naz, A. Kumar, A. Majumdar, and R. Agrawal, "Is artificial intelligence an enabler of supply chain resiliency post-COVID-19? An exploratory state-of-the-art review for future research," *Oper. Manag. Res.*, vol. 15, no. 1, pp. 378–398 (2022).
- [13] A. K. Dubey and V. Jain, "A review of face recognition methods using deep learning network," *J. Inf. Optim. Sci.*, vol. 40, no. 2, pp. 547–558 (2019).
- [14] S. Maurya and A. Jain, "Deep learning to combat phishing," *J. Stat. Manag. Syst.*, vol. 23, no. 6, pp. 945–957 (2020).

Received August, 2024