International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(8s) 2026 pp. 927-941
https://doi.org/10.70917/ijcisim-2026-4133
Article

GENERATIVE AI A COMPREHENSIVE EXAMINATION OF TECHNIQUES, APPLICATIONS, AND DIFFICULTIES

Sanat Jain[1], Mehbob Ali[2], Mohd Rafee[3], Swati Mahesh Patil[4], Vaibhav Panwar[5], Homera Halvadi[6]

[1] School of Computing Science and Engineering, VIT Bhopal University.
Email: jain.sanat@gmail.com
[2] University of Ladakh, Kargil Campus.
Email: mehbobaali@gmail.com
[3] University of Ladakh.
Email: mohdrafee7270@gmail.com
[4]Department of AI & ML, Bharati Institute of Technology (Polytechnic), Belapur CBD, Navi Mumbai – 400614.
Email: swati24mtit@student.mes.ac.in
[5] Manipal University Jaipur.
Email: Panwar.vaibhav@gmail.com
[6]Department of CSE, SVET College, Saurashtra University, Jamnagar, Gujarat, India.
Email: hnhalvadi@gmail.com

Abstract: Generative Artificial Intelligence (Generative AI) has emerged as a transformative technology capable of generating realistic text, images, audio, video and code in a wide range of application areas. The review covers several important Generative AI methodologies, such as Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Transformer-based models, Diffusion Models, and Hybrid Architectures. This also delves into their applications in the fields of healthcare, NLP, computer vision, finance, cybersecurity, climate science, and digital twin technologies. Also covered are key ethical, social, and economic concerns, and new research areas like multimodal AI, explainable AI, AI safety, and foundation models. This research paper reveals the latest developments, current challenges, and future research directions in creating reliable, effective, and ethical Generative AI systems.

Keywords: Generative AI, VAEs, GANs, Transformers, Diffusion Models, Large Language Models, Deep Learning.

1. INTRODUCTION

AI (Artificial Intelligence) is one of the most transformative technologies of the digital age, which allows the machine to execute processes that would normally necessitate human intelligence: learning, reasoning, decision making, perception, and problem-solving. The capabilities of AI have grown tremendously in recent years, especially in the field of deep learning and neural networks, which enable intelligent systems to process vast amounts of data and automate complex tasks in different areas. Generative Artificial Intelligence (Generative AI) is a significant leap forward in AI, moving beyond analysis and classification to generate new and original content.

Whereas traditional AI systems are mostly used to predict, recognize, or support decision making, Generative AI uncovers the patterns and probability distributions in training data to create authentic text, images, audio, video, code, and more. This ability has revolutionized the way intelligent systems interact with their users and real-world problems are solved. From its initial probabilistic models and rudimentary neural networks, Generative AI has advanced to complex deep learning architectures that can generate highly realistic outputs. Key advancements such as Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Transformer-based architectures, Large Language Models (LLMs), and Diffusion Models have enhanced the quality, diversity, and reliability of AI-generated content.

These technologies have led to the significant progress in natural language processing, computer vision, healthcare, robotics, digital media, scientific research, and intelligent automation. In today's digital economy, Generative AI is a key component in driving innovation, productivity, and cutting down creation time and expenses for content and decision-making. Generative AI is rapidly becoming a staple tool in healthcare, pharmaceuticals, software, finance, cybersecurity, education, virtual assistants, design, and industrial automation, among other applications. It can also be used to create synthetic but realistic data, which is useful when dealing with data scarcity and privacy concerns in machine learning applications. While its potential applications are vast, Generative AI also poses a number of ethical, societal, and technical challenges, such as algorithmic bias, misinformation, deepfakes, copyright concerns, privacy issues, computational demands, and the need for responsible AI governance. It is important to address these challenges in order to ensure the trustworthy, transparent and sustainable deployment of generative models.

This paper offers an extensive exploration of the concept of Generative AI, covering its underlying principles, key methods, mathematical underpinnings, various application areas, ethical and societal implications, and the latest advancements and future prospects. The research underscores the ongoing impact of emerging generative models on industries and science and provides insights into the opportunities and challenges of the next generation of intelligent systems.

2. REVIEW OF LITERATURE

Sengar, et,al (2025) In recent years, the study of artificial intelligence (AI) has undergone a paradigm shift. This has been propelled by the groundbreaking capabilities of generative models both in supervised and unsupervised learning scenarios. Generative AI has shown state-of-the-art performance in solving perplexing real-world conundrums in fields such as image translation, medical diagnostics, textual imagery fusion, natural language processing, and beyond. This paper documents the systematic review and analysis of recent advancements and techniques in Generative AI with a detailed discussion of their applications including application-specific models. Indeed, the major impact that generative AI has made to date, has been in language generation with the development of large language models, in the field of image translation and several other interdisciplinary applications of generative AI. Moreover, the primary contribution of this paper lies in its coherent synthesis of the latest advancements in these areas, seamlessly weaving together contemporary breakthroughs in the field. Particularly, how it shares an exploration of the future trajectory for generative AI. In conclusion, the paper ends with a discussion of Responsible AI principles, and the necessary ethical considerations for the sustainability and growth of these generative models.

Balasubramaniam et,al (2024) The feld of generative artifcial intelligence (AI) is experiencing rapid advancements, impacting a multitude of sectors, from computer vision to healthcare. Tis paper provides a comprehensive review of generative AI’s evolution, signifcance, and applications, including the foundational architectures such as generative adversarial networks (GANs), variational autoencoders (VAEs), autoregressive models, fow-based models, and difusion models. We delve into the impact of generative algorithms on computer vision, natural language processing, artistic creation, and healthcare, demonstrating their revolutionary potential in data augmentation, text and speech synthesis, and medical image interpretation. While the transformative capabilities of generative AI are acknowledged, the paper also examines ethical concerns, most notably the advent of deepfakes, calling for the development of robust detection frameworks and responsible use guidelines. As generative AI continues to evolve, driven by advances in neural network architectures and deep learning methodologies, this paper provides a holistic overview of the current landscape and a roadmap for future research and ethical considerations in generative AI.

Yazdani et,al (2025) In recent years, deep learning based generative models, particularly Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Diffusion Models (DMs), have been instrumental in generating diverse, high-quality content across various domains, such as image and video synthesis. This capability has led to widespread adoption of these models and has captured strong public interest. As they continue to advance at a rapid pace, the growing volume of research, expanding application areas, and unresolved technical challenges make it increasingly difficult to stay current. To address this need, this survey introduces a comprehensive taxonomy that organizes the literature and provides a cohesive framework for understanding the development of GANs, VAEs, and DMs, including their many variants and combined approaches. We highlight key innovations that have improved the quality, diversity, and controllability of generated outputs, reflecting the expanding potential of generative artificial intelligence. In addition to summarizing technical progress, we examine rising ethical concerns, including the risks of misuse and the broader societal impact of synthetic media. Finally, we outline persistent challenges and propose future research directions, offering a structured and forward looking perspective for researchers in this fast evolving field.

Elhanashi, et,al (2026) Generative artificial intelligence and foundation models have changed machine learning by allowing systems to produce readable text, realistic images, and other multimodal content with little direct input from a user. Foundation models are large neural networks trained on very large and varied datasets, and they form the core of many current generative AI (GenAI) systems. Their rapid development has led to major advances in areas like natural language processing, computer vision, multimodal learning, and robotics. Examples include GPT, LLaMA, and diffusion-based architectures, such as models often used for image generation. Systems such as Stable Diffusion show this shift by illustrating how AI can interpret information, draw basic inferences, and produce new outputs using more than one type of data. This review surveys common foundation model architectures and examines what they can do in generative tasks. It reviews Transformer, diffusion, and multimodal architectures, focusing on methods that support scaling and transfer across domains. The paper also reviews key approaches to pretraining and fine-tuning, including self-supervised learning, instruction tuning, and parameter-efficient adaptation, which support these systems' ability to generalize across tasks. In addition to the technical details, this review discusses how GenAI is being used for text generation, image synthesis, robotics, and biomedical research. The study also notes continuing challenges, such as the high computing and energy demands of large models, ethical concerns about data bias and misinformation, and worries about privacy, reliability, and responsible use of AI in real settings. This review brings together ideas about model design, training methods, and social implications to point future research toward GenAI systems that are efficient, easy to interpret, and reliable, while supporting scientific progress and ethical responsibility.

Sai et,al (2024) Generative artificial intelligence (GAI) can be broadly described as an artificial intelligence system capable of generating images, text, and other media types with human prompts. GAI models like ChatGPT, DALL-E, and Bard have recently caught the attention of industry and academia equally. GAI applications span various industries like art, gaming, fashion, and healthcare. In healthcare, GAI shows promise in medical research, diagnosis, treatment, and patient care and is already making strides in real-world deployments. There has yet to be any detailed study concerning the applications and scope of GAI in healthcare. Addressing this research gap, we explore several applications, real-world scenarios, and limitations of GAI in healthcare. We examine how GAI models like ChatGPT and DALL-E can be leveraged to aid in the applications of medical imaging, drug discovery, personalized patient treatment, medical simulation and training, clinical trial optimization, mental health support, healthcare operations and research, medical chatbots, human movement simulation, and a few more applications. Along with applications, we cover four real-world healthcare scenarios that employ GAI: visual snow syndrome diagnosis, molecular drug optimization, medical education, and dentistry. We also provide an elaborate discussion on seven healthcare-customized LLMs like Med-PaLM, BioGPT, DeepHealth, etc.,Since GAI is still evolving, it poses challenges like the lack of professional expertise in decision making, risk of patient data privacy, issues in integrating with existing healthcare systems, and the problem of data bias which are elaborated on in this work along with several other challenges. We also put forward multiple directions for future research in GAI for healthcare.

3. METHODOLOGIES AND ARCHITECTURES

Generative AI uses a variety of mathematical models to approximate complex data distributions and to generate high fidelity artificial data. It provides a careful examination of the core generative methods such as VAEs, GANs, Transformer-Based Generative Models, Diffusion Models, Hybrid Architectures and Energy-Based Models (EBMs) / Score-Based Generative Models.

TABLE 1. Evolution of Generative AI: Key Models and Contributions

Year	Model/Architecture	Key Contributions	Strengths	Limitations
2006	Restricted Boltzmann Machines (RBMs), Deep Belief Networks (DBNs)	Introduced deep probabilistic generative modeling using energy-based learning	Hierarchical feature learning; unsupervised representation learning	Training instability; limited scalability to complex datasets
2013	Variational Autoencoders (VAEs)	Introduced probabilistic latent-space learning for data generation	Efficient training; structured latent	Generated samples often appear blurry; limited visual fidelity

			representation; smooth interpolation	
2014	Generative Adversarial Networks (GANs)	Introduced adversarial learning for realistic data generation	High-quality image synthesis; realistic outputs	Mode collapse; unstable training; sensitive to hyperparameter selection
2015–2019	DCGAN, WGAN, Progressive GAN (PGGAN), StyleGAN	Improved GAN stability, resolution, and controllability for image synthesis	High-resolution image generation; improved stability; controllable image styles	High computational cost; difficult training and tuning
2017	Transformer Architecture	Introduced the self-attention mechanism for efficient sequence modeling	Parallel computation; captures long-range dependencies; foundation of modern LLMs	Computationally expensive; high memory consumption for long sequences
2018–2020	GPT Family (GPT, GPT-2, GPT-3) and BERT	Large-scale Transformer pre-training advanced language generation (GPT) and language understanding (BERT)	Strong contextual representations; excellent transfer learning; scalable performance	Large training cost; bias; hallucination in generative models; BERT is not a generative model
2020–2022	Diffusion Models (DDPM, Imagen, Stable Diffusion)	Introduced iterative denoising for high-fidelity image generation	Stable training; photorealistic outputs; diverse image synthesis	Slow inference; computationally intensive sampling
2022	ChatGPT	Popularized conversational Generative AI through instruction-tuned large language models	Human-like dialogue; broad public adoption; versatile real-world applications	Hallucinations; factual inaccuracies; significant computational requirements
2023	GPT-4, LLaMA, SDXL, Consistency Models	Advanced multimodal reasoning, open foundation models, and faster diffusion-based generation	Improved reasoning; multimodal capabilities; higher image quality; faster inference	High computational requirements; safety, bias, and copyright concerns
2024	GPT-4o, Sora	Introduced native multimodal AI and advanced text-to-video generation	Real-time multimodal interaction; high-quality video generation; improved efficiency	High infrastructure cost; ethical, copyright, and misuse concerns

- **Variational Autoencoders**

Variational Autoencoders (VAEs) are a class of probabilistic generative models which learn a compact representation of input data and also generate new data. Unlike traditional autoencoders, VAEs encode every input as a probability distribution, enabling the model to produce a variety of output images that are more realistic. VAEs are widely used in image synthesis, anomaly detection, data compression and representation learning because of their efficient latent-space learning.

Given an input x, the encoder learns an approximate posterior distribution over the latent variable z, expressed as

$$q_{\phi}(z \mid x) = \mathcal{N}\left(\mu_{\phi'}(x)\sigma_{\phi}^{2}(x)I\right)$$

where $\mu_\phi(x)$ and $\sigma_\phi(x)$ denote the mean and standard deviation of the latent distribution. The decoder reconstructs the input by estimating

$$p_\theta(x \mid z)$$

The model is trained by maximizing the Evidence Lower Bound (ELBO), defined as

$$\mathcal{L}_{VAE} = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta (x \mid z)] - D_{KL}\left(q_\phi(z \mid x) \parallel p(z)\right)$$

where D_{KL} regularizes the latent distribution toward the prior $p(z)$. To enable backpropagation through the stochastic latent variables, the reparameterization trick is applied as

$$z = \mu + \sigma \odot \epsilon\,,\ \epsilon \sim \mathcal{N}(0, I)$$

The VAE architecture consists of an encoder, a latent space, and a decoder that together learn meaningful data representations and generate new samples.

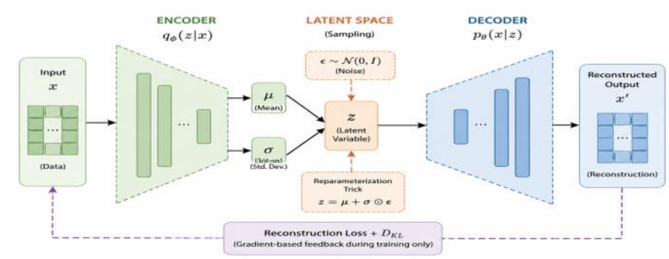

Figure 1. Variational Autoencoder (VAE) Architecture

The stability of the training and structured latent representations is also appealing properties of VAEs for applications like medical imaging, anomaly detection, and synthetic data generation. The generated outputs are, however, smoother/blurrier than outputs from GANs or diffusion models, especially in the synthesis of high-resolution images.

- **Generative Adversarial Networks**

Generative Adversarial Networks (GANs) are deep generative models using adversarial learning to produce a realistic synthetic data set. A GAN is composed of two neural networks, one that generates synthetic samples from random latent vectors (generator) and the other that determines if the sample is real or synthetic (discriminator). When training, both of the networks compete with each other, so the generator can increasingly create more and more realistic results. Medical imaging, style transfer and data augmentation are common applications of GANs in image synthesis.

The objective function of a GAN is defined as

$$\min_G \max_D V(D, G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D\,(x)] + \mathbb{E}_{z\sim p(z)}\left[\log\left(1 - D\big(G(z)\big)\right)\right]$$

where $p_{data}(x)$ represents the real data distribution and $p(z)$ denotes the latent distribution. The generator aims to fool the discriminator, while the discriminator learns to distinguish real samples from generated ones.

To improve training stability, variants such as WGAN replace the original loss with the Wasserstein distance:

$$W(p_{data}, p_G) = \inf_{\gamma\in\Pi(p_{data}, p_G)} \mathbb{E}_{(x,y)\sim\gamma}[\parallel x - y \parallel]$$

where p_G represents the generated data distribution.

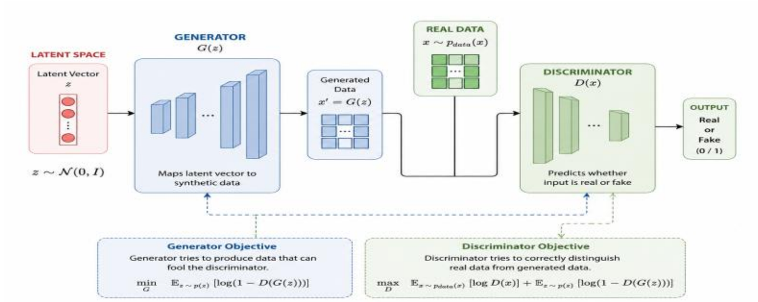

Figure 2. Architecture of Generative Adversarial Networks (GANs)

GANs can produce very realistic synthetic data and have been one of the most popular generative models for computer vision and multimedia applications. But they suffer from problems like training instabilities, mode collapse, and computational demands, and thus more stable variants have been produced, like WGAN, StyleGAN and CycleGAN.

- **Transformer-Based Generative Models**

The ability of transformer-based generative models to efficiently process sequential data with the self-attention mechanism has led to the development of modern Generative AI. Transformers enable parallel computation and effectively capture long-range dependencies, which make them very useful for large-scale language and multimodal tasks, as opposed to recurrent models. The GPT model, BART, T5, and LLaMA are examples of transformer models that are commonly used for tasks like text generation, translation, summarization, and conversational AI.

The self-attention mechanism is computed as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, and V represent the query, key, and value matrices, respectively, and d_k denotes the dimensionality of the key vectors. This mechanism enables the model to identify the importance of each input token while generating contextual representations.

For autoregressive generation, the probability of generating a sequence is expressed as

$$P(x) = \prod_{t=1}^{T} P\,(x_t, | , x_1, x_2 \ldots x_{t-1})$$

where each token is predicted based on the previously generated tokens.

Transformer-based models have dramatically enhanced the performance of text generation and multimodal applications by learning contextually-aware relationships from large-scale datasets. They are known for their scalability and ability to understand complex texts and have become the foundation of modern LLM architectures. These models will, however, need significant computational resources and might generate output that is biased or factually wrong, which is why it is crucial to keep researching efficient and reliable Transformer architectures.

- **Diffusion Models**

Diffusion models are deep generative models which produce high-quality synthetic data through a gradual noise addition and noise removal process. These models can be used to effectively generate and edit images because they learn to generate original data by reversing a diffusion process. New models like DDPM, Stable Diffusion, and SDXL have proven to be excellent generators of realistic and varied images.

The forward diffusion process gradually adds Gaussian noise to the input sample and is defined as

$$q(x_t \mid x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t};\ x_{t-1}\beta_t I\right)$$

where x_t is the noisy sample at time step t, β_t controls the amount of noise added, and I denotes the identity matrix.

The reverse diffusion process learns to remove the added noise and recover the original data distribution. The reverse transition probability is expressed as

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}\left(x_{t-1}; \mu_\theta; (x_t, t)\Sigma_\theta(x_t, t)\right)$$

where μ_θ and Σ_θ represent the predicted mean and covariance of the reverse process.

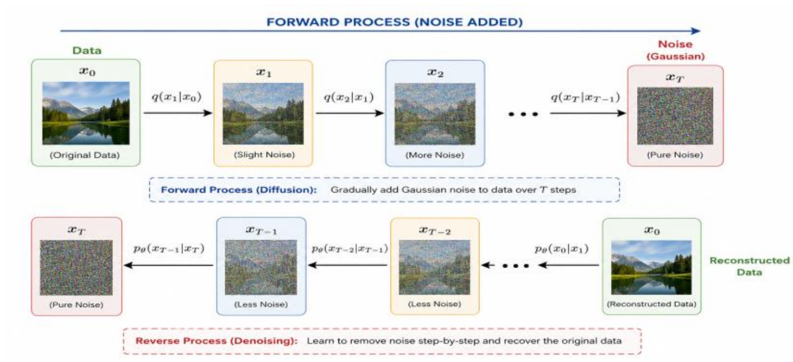

Figure 3. Diffusion Model process

The image generation process is stable, and diffusion models can also generate high-quality and diverse images, which is why they have been one of the preferred methods. These are applied extensively in image synthesis, image editing, text to image and content creation. They perform iterative denoising which has higher computation cost and slower inference than multiple other generative models, however.

- **Hybrid Architectures**

Hybrids are architectures that use two or more generative models together to take advantage of their complementary characteristics. The architectures combine various models, including VAEs, GANs, Transformers, and Diffusion Models, to enhance the quality of generations, their robustness, and learning efficiency. Recently, hybrid methods have been receiving a lot of attention in such applications as image synthesis, text generation, healthcare, multimodal learning, and scientific research.

The overall objective of a hybrid generative model can be represented as a weighted combination of multiple learning objectives:

$$\mathcal{L}_{Hybrid} = \lambda_1 \mathcal{L}_{VAE} + \lambda_2 \mathcal{L}_{GAN} + \lambda_3 \mathcal{L}_{Task}$$

where $\mathcal{L}_{VAE}$ represents the variational reconstruction loss, $\mathcal{L}_{GAN}$ denotes the adversarial loss, $\mathcal{L}_{Task}$ corresponds to a task-specific objective, and λ_1, λ_2, and λ_3 are weighting parameters that balance the contribution of each component during optimization.

Examples range from latent representation learning and adversarial training models like VAE–GANs for image generation to realistic images to Transformer–Diffusion models that fuse contextual understanding with high-quality image synthesis. In modern multimodal foundation models, hybrid architectures are used that process both text, images, audio and video in a unified framework. The hybrid architectures are designed to take advantage of the benefits of several models, and they are used to build strong and efficient generative systems. They have demonstrated their performance improvement in various applications, especially in multimodal content generation and complex real-life tasks. But using several architectures will lead to a higher level of computation complexity, a longer training time, and a more complex optimization process, which will need to be designed with care and optimized efficiently.

4. APPLICATIONS OF GENERATIVE AI ACROSS VARIOUS DOMAINS

Generative AI has become a groundbreaking technology, applicable in a wide range of fields, including healthcare, natural language processing, computer vision, climate science, aerospace, autonomous systems, creative industries, finance, cybersecurity, and intelligent search. Generative AI can produce realistic data, provide intelligent decision support, automate processes and forecast the future, all thanks to the use of advanced architectures like Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Transformers, Diffusion Models and Hybrid Architectures. They are not only making the operations more efficient but also making scientific discoveries faster and supporting customization of services and innovative solutions in research and industry. Let's briefly review the key applications of Generative AI.

A. Healthcare and Biomedical Research

The use of Generative AI has revolutionized healthcare through advancements in medical imaging, drug discovery, protein structure prediction, and personalized medicine. The GANs and Diffusion Models are used to improve MRI, CT, and PET images for super-resolution, image reconstruction, and synthetic data generation, which can boost the accuracy of diagnosis without relying on large annotated datasets. VAEs/DMs have the potential to speed up drug discovery by creating new molecular structures and improving candidate molecules. Biomedical research and precision medicine rely on accurate protein structure prediction, which can be achieved by transformer-based models such as AlphaFold and ESMFold. While these developments are encouraging, there are still research areas of clinical validation, regulatory compliance and data privacy.

B. Natural Language Processing and Text Generation

Transformer architecture for language models with human-like understanding and generation has transformed Natural Language Processing (NLP). GPT, BERT, T5, and LLaMA are models that are capable of machine translation, text summarization, question answering, content generation, code generation and conversational AI. They enhance context understanding, semantic reasoning, and multilingual communication, and are useful in software development, healthcare, and customer service, among other domains, in the educational environment. But problems like hallucination, bias, misinformation are still active lines of research.

C. Computer Vision and Image Synthesis

Generative AI has significantly improved computer vision by creating lifelike images, enhancing images, super-resolution, inpainting, style transfer, and generating images from text. GANs and Diffusion Models are used to create cutting-edge systems like Stable Diffusion and DALL·E for generating high-quality images from textual inputs. High-resolution 3D scene reconstruction is also important for virtual reality, augmented reality, robotics and digital content creation, and Neural Radiance Fields (NeRF) are an important addition to this. Recent studies are aimed at improving the quality of images, efficiency of computations, and the control of produced images.

D. Climate Science and Environmental Modeling

By enabling weather forecasting, climate simulation, environmental monitoring, and disaster prediction, generative AI plays a key role in climate science. Physics-informed generative models can help with climate analysis, and generative models based on GANs can better help in remote sensing, such as deforestation monitoring, land use, biodiversity, and natural disaster detection. Diffusion Models are also used for designing materials for carbon capture and sustainable environmental technologies. These applications will serve to better understand the development of more accurate climate modeling and to help with evidence based environmental decision-making.

E. Aerospace Engineering and Space Exploration

In the field of aerospace engineering, generative AI aids in designing aircrafts, developing spacecrafts, planning trajectories, and autonomous navigation. GANs are effective for aerodynamic shape optimization, and Diffusion Models and reinforcement learning-based simulations are effective in mission planning and robotic explorations. Super-resolution techniques enhance the resolution of satellite images and astronomical observations, which helps in the identification and analysis of distant celestial objects. The technologies help to make aerospace systems more efficient and future space missions.

F. Reinforcement Learning and Autonomous Systems

Generative AI complements and enriches reinforcement learning by providing simulated environments with realistic scenarios for autonomous decision-making and policy optimization. Predicting future environmental states is crucial for robotics, autonomous driving, intelligent navigation and industrial automation, and World Models and Transformer-based reinforcement learning help achieve this. These simulation-based techniques are used to decrease training expenses, enhance sample efficiency, and boost the reliability of autonomous systems in real-world settings.

G. Art, Music, and Creative Industries

The creative industries have been revolutionized by generative AI, which has helped create AI algorithms for art, music, video generation, animation, and digital story creation. Style transfer, procedural content generation and multimedia synthesis can be achieved using GANs, VAEs, Diffusion Models and Transformer-based models. Some models, like MuseNet and Jukebox, can produce original music compositions, while the newer image generation models can help artists, designers, filmmakers, and game developers speed up their creative processes. Copyright and ownership and authenticity are still topics of discussion.

H. Metaverse and Digital Twin Technologies

The potential applications of Generative AI in the Metaverse and Digital Twin are significant, especially when it comes to creating realistic virtual spaces, smart avatars, procedural content, and VR/AR experiences. Digital Twins are AI-driven virtual representations of physical facilities like factories, smart cities, transportation systems, and healthcare centers, which help in predictive maintenance, real-time monitoring, and optimizing operations. These technologies help make the best decision and minimize operating costs.

I. Finance and Market Simulation

Financial forecasting, fraud detection, risk assessment, algorithmic trading, market simulation are all areas of financial applications that are supported by generative AI. Time-series generative models like TimeGAN create accurate financial time series data for investment analysis, stress testing and portfolio optimization. Multi-agent simulations enhance trading strategies by simulating interactions between investors and market conditions. Financial institutions leverage Generative AI to keep customer information confidential when creating synthetic data, for credit scoring, and to prevent money laundering.

J. Security and Cyber Threat Detection

Some of the ways generative AI enhances cybersecurity are by enabling better anomaly detection, intrusion detection, malware analysis, cyber deception, and threat intelligence. GAN based models detect abnormal network activities and create synthetic attack scenarios for security testing. With enhanced capabilities in deepfake detection and adversarial defense, these advanced techniques contribute to organizations becoming more cyber resilient and able to withstand emerging digital threats. Finding the right balance between the good and the bad use of Generative AI is a big research challenge.

K. LLM-Augmented Search Engines

Large Language Model (LLM) enhanced search engines utilize Transformer models for language understanding and incorporate traditional search algorithms to provide conversational and context-aware search results. Retrieval-Augmented Generation (RAG) combines retrieved documents with language generation to generate more accurate, explainable and contextually relevant responses. The modern systems enhance the comprehension of queries, document summarization and interactive knowledge discovery while diminishing the amount of reasoning that users have to put in. However, some challenges with next-generation AI-powered search systems remain: ensuring the information is factual, reducing hallucinations, improving source attribution, and maintaining transparency.

TABLE 2. Comparative Analysis of Generative AI Applications Across Different Domains

Application Domain	**Representative Generative Models**	**Primary Applications**	**Key Advantages**	**Major Challenges**
Healthcare and Biomedical Research	GAN, Diffusion Models, VAE, AlphaFold, ESMFold	Medical image enhancement, disease diagnosis, drug discovery, protein structure prediction, personalized medicine	Improved diagnostic accuracy, faster drug discovery, privacy-preserving synthetic medical data	Data privacy, limited labeled datasets, clinical validation, regulatory approval

Natural Language Processing and Text Generation	GPT-4, GPT-3.5, LLaMA, T5, BART, Gemini	Machine translation, text summarization, chatbots, question answering, content generation, code generation	Human-like text generation, multilingual support, improved productivity	Hallucination, bias, misinformation, high computational cost
Computer Vision and Image Synthesis	Stable Diffusion, DALL·E, Midjourney, StyleGAN2, CycleGAN, NeRF	Text-to-image generation, image editing, super-resolution, inpainting, 3D reconstruction	High-quality image generation, realistic visual content, creative design support	Copyright concerns, fake image generation, computational requirements
Climate Science and Environmental Modeling	Physics-informed GANs, Diffusion Models, ClimateGAN, Generative Earth System Models (GESMs)	Weather prediction, climate simulation, disaster forecasting, environmental monitoring, carbon capture material design	Better environmental forecasting, improved satellite image analysis, sustainability research	Data uncertainty, model complexity, large-scale computation
Aerospace Engineering and Space Exploration	Diffusion Models, GANs, Reinforcement Learning World Models, Transformer-based planners	Aircraft design optimization, spacecraft design, trajectory planning, autonomous navigation, satellite image enhancement	Reduced design time, improved fuel efficiency, better mission planning	Safety validation, simulation reliability, expensive computational resources
Reinforcement Learning and Autonomous Systems	DreamerV3, MuZero, World Models, Transformer-RL, GAN-based simulation	Robot learning, autonomous driving, motion planning, intelligent decision-making, digital simulation	Sample-efficient learning, realistic simulation, reduced training cost	Sim-to-real transfer, high training complexity, safety constraints
Art, Music and Creative Industries	StyleGAN, MuseNet, Jukebox, DALL·E, Stable Diffusion, Runway Gen	AI-generated art, music composition, video generation, animation, game asset creation	Faster creative workflows, personalized content, enhanced digital creativity	Copyright issues, originality concerns, ethical misuse
Metaverse and Digital Twin Technologies	NeRF, Instant-NGP, DigitalTwinGAN, Gaussian Splatting, GANs	Virtual avatars, immersive VR/AR, digital twins, smart cities, industrial simulation	Realistic virtual environments, predictive maintenance, real-time monitoring	High computational cost, scalability, interoperability
Finance and Market Simulation	TimeGAN, TabDDPM, GAN-Fin, Variational Recurrent Autoencoders	Fraud detection, risk analysis, algorithmic trading, portfolio optimization, synthetic financial data generation	Better financial forecasting, privacy-preserving synthetic data, improved risk management	Regulatory compliance, data security, market uncertainty
Security and Cyber Threat Detection	GAN-based anomaly detection, Vision Transformers (ViT),	Intrusion detection, malware analysis, cyberattack simulation,	Early attack detection, proactive cybersecurity,	Adversarial attacks, evolving threats, false positives

	DeepFakeGAN, Transformer models	deepfake detection, threat intelligence	stronger digital defense	
LLM-Augmented Search Engines	GPT-4, Gemini, Claude, RAG, PaLM, ReRankers, Retrieval-Augmented LLMs	Conversational search, enterprise search, document summarization, question answering, knowledge retrieval	Context-aware responses, grounded information, reduced hallucination, improved user experience	Retrieval quality, latency, source reliability, hallucination control

5. CHALLENGES AND DIFFICULTIES IN GENERATIVE AI

Generative AI has revolutionized various sectors by facilitating intelligent content creation, automation, and decision-making. Yet, its widespread uptake poses a number of ethical, social, and economic concerns that need to be resolved to ensure responsible and trustworthy use of AI. Issues are algorithmic bias, misinformation, intellectual property issues, transformation of the workforce, governance, and economic inequality.

A. Ethical Considerations

• Bias and Fairness in AI-Generated Content

Generative AI models are trained on large datasets that may contain historical and societal biases. Consequently, the learned conditional distribution $p_\theta(y \mid x)$can reinforce existing discrimination in applications such as hiring, healthcare, finance, and criminal justice. Fairness-aware learning techniques reduce these biases by incorporating a fairness regularization term into the optimization objective.

The fairness-aware optimization objective is expressed as

$$\min_\theta \mathbb{E}_{p(x)} \left[L(f_{\theta'}(x)y) + \lambda D_{KL}\left(p_\theta(y \mid x) \parallel p_{fair}(y)\right)\right]$$

where $L(f_{\theta'}(x)y)$represents the prediction loss, $p_{fair}(y)$denotes the fairness-aware target distribution, and λcontrols the fairness constraint.

• Deepfakes and Misinformation

Highly realistic synthetic images, videos and speech can be generated by generative AI, which can be used in ways that lead to deepfakes and misinformation. This has the potential to compromise digital trust, cybersecurity, political stability and personal privacy. The existing mitigation methods are AI watermarking, blockchain-based content authentication, digital forensics, and deepfake detection models.

The discriminator used in deepfake detection is trained using the binary classification objective

$$\max_\theta \mathbb{E}_{x\sim p_{real}}[\log D_\theta(x)] + \mathbb{E}_{x\sim p_{fake}}\left[\log\left(1 - D_\theta(x)\right)\right]$$

where $D_\theta(x)$estimates whether an input sample is authentic or AI-generated.

• Intellectual Property and Copyright Issues

Generative AI systems are based on vast amounts of text, images, music and other digital information. This poses serious copyright and owner issues due to the possibility of AI output being similar to other copyrighted works. Clear licensing policy, transparency of training data, and new copyright laws are necessary to strike a balance between innovation and creators' rights.

• AI Deployment in Safety-Critical Systems

Healthcare, aviation, autonomous systems in cars and industry rely on the critical need for safety, robustness, and explainability when using generative AI. Failure in these environments can result in serious consequences, making it critical for verification, human oversight, explainable AI (XAI), and adherence to safety standards to ensure reliable deployment.

B. Societal Implications

• Labor Market Disruptions and Workforce Transformation

Generative AI can be used to automate repeat cognitive and creative tasks across various industries. While automation can eliminate some roles, it is also creating opportunities in the field of AI development, auditing, and human-AI collaboration. There is therefore a need for workforce reskilling and perpetual learning.

The estimated labor displacement is represented as

$$L_{displaced} = \sum_{i=1}^{N} \alpha_i E_i$$

• AI Governance and Regulatory Challenges

The increasing deployment of generative AI requires effective governance frameworks that promote transparency, fairness, accountability, and privacy. To create risk-based governance and responsible deployment of AI in various application domains, international regulations like the EU AI Act and national AI policies are being drafted.

• Cognitive and Psychological Effects of LLM Usage

Regular usage of Large Language Models (LLMs) can diminish analytical thinking, critical thinking, and problem-solving skills. Excessive trust in AI generated outputs by the human user is also a possibility. Hence, AI systems need to give explanations, levels of confidence and make sure that there is responsible human supervision.

• Education and AI Literacy

AI literacy allows people to be aware of the capabilities and limitations of generative AI. Students, professionals and policymakers use educational programs to assess AI-created content critically, identify misinformation, and responsibly employ AI.

• Human-AI Collaboration

Collaboration between humans and AI involves using computational intelligence, but also engaging in human creativity, judgment, and ethical considerations. Instead of replacing human knowledge, generative AI is used more as an intelligent assistant, enhancing productivity, creativity, and decision making in a variety of industries.

C. Economic Implications

• Productivity and Business Transformation

The generative AI helps in increasing productivity by automating document creation, software development, customer support, design, and business analytics. The production function is a useful model to illustrate the economic value of AI.

$$Y = A \cdot G(K, L)$$

• AI-Driven Innovation and Market Growth

In healthcare, generative AI can streamline drug discovery and development, while in finance, it can improve risk assessment for investment decisions.In education, AI can generate content for textbooks and course materials, and in manufacturing, it can create product designs. In entertainment and scientific research, AI can enhance content creation and research. It does improve productivity and economic growth but it might also raise economic inequality due to the dominance of large technology companies in the market.

The AI-induced change in income inequality is represented as

$$G_{new} = G_{old} + G(AI)$$

• Cost, Energy Consumption, and Infrastructure Challenges

The training of modern generative AI models consumes vast computational power, such as high-performance GPUs, cloud computing resources, and energy. These requirements impose restrictions on the finances for SMEs and academic institutions. Parameter-efficient fine-tuning, quantization, Mixture-of-Experts (MoE), and lightweight open-source models are employed to mitigate the computational costs, while enhancing accessibility.

6. EMERGING TRENDS AND FUTURE RESEARCH DIRECTIONS

- Multi-Modal Generative AI: Future generative AI systems will be more likely to combine text, images, audio and video into a single system, so that content can be generated across multiple modalities and understood across multiple modalities. They are used in various applications, including text-to-image generation, image captioning, speech generation, video generation, robotics, healthcare, and scientific discovery. The joint generation process can be represented as:

$$p_\theta\left(x_{image}, x_{text}, x_{audio}\right) = p_\theta\left(x_{image} \mid x_{text}\right) \cdot p_\theta\left(x_{audio} \mid x_{image}\right)$$

- Reinforcement Learning-Enhanced Generative AI: Integrating Reinforcement Learning (RL) with generative models to optimize decisions based on rewards. AI systems learn from user feedback and environmental interactions, making them ideal for robotics, autonomous systems, recommendation engines, and intelligent assistants, as opposed to static outputs.

$$\theta^* = \arg\max_\theta \mathbb{E}_{x\sim p_\theta(x)}[R(x)]$$

- Explainable and Trustworthy Generative AI: Future research is geared toward enhancing the transparency, interpretability, and controllability of generative models. Explainable AI (XAI) techniques are employed to explain the output, making critical applications like health services, financial services, and science more trustworthy.

$$p_\theta(x \mid z) = p_\theta(x, \mid, z_1, z_2 \ldots z_k)$$

- AI Safety and Alignment: Ensuring that generative AI systems generate safe, reliable, and ethically appropriate outputs is a critical research challenge. Recent efforts focus on minimizing hallucination, robustness in the face of adversarial attacks, adding Reinforcement Learning from Human Feedback (RLHF), addressing bias, and building governance frameworks to ensure responsible use of AI.

- Efficient and Sustainable Generative AI: The demand for more computing power and energy consumption of these large AI models has sparked efforts to make them more efficient, including Mixture of Experts (MoE), Low-Rank Adaptation (LoRA), model quantization, model pruning, and knowledge distillation. All the approaches proposed in this article are low computing-cost and high model-performance.

$$f_\theta(x) = \sum_{i=1}^{M} g_i\,(x)\, f_{\theta_i}(x)$$

- Foundation Models and General Artificial Intelligence: Next-generation AI systems are based on foundation models and general AI, trained on vast and large amounts of data. Researchers will strive to create more general models that can reason, continuously learn, understand long context, solve problems independently, and transfer knowledge across domains towards Artificial General Intelligence (AGI).

- Future Research Opportunities: Future research to explore lightweight and energy-efficient generative models, improve multimodal reasoning, enhance explainability and trustworthiness, further enhance AI safety and alignment, preserve privacy, prevent synthetic data contamination, enable scientific discovery, expand human-AI collaboration, and build robust ethical and regulatory frameworks. The research avenues listed above will enhance the scalability, reliability, and real-world adoption of generative AI in various application areas.

7. DISCUSSION

The study showcases the emergence of Generative AI from traditional generative models to sophisticated architectures like Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Transformer-based models, Diffusion Models, and Large Language Models (LLMs), broadening its use in healthcare, natural language processing, computer vision, climate science, finance, cybersecurity, autonomous systems, and digital twin technologies. This paper presents a comprehensive overview, as opposed to the existing surveys, which typically focus on a particular model or application domain; it brings together methods, mathematical foundations, a variety of real-world applications, ethical issues and future research directions under a unified framework. The review emphasizes the current state-of-the-art models for high-quality text, image, audio, and multimodal content generation are based on Transformers and Diffusion, and LLMs are revolutionizing human-computer interaction and intelligent information retrieval. However, there are still several research challenges to address, such as the explainability of the models, mitigation of biases, computational efficiency, preserving privacy, energy consumption, reducing hallucinations, and creating trustworthy and sustainable AI systems. Additionally, the ubiquity of Generative AI presents ethical, social and economic issues concerning misinformation, copyright, technology influence on the workforce and AI governance. This paper being a review-based study is limited to the reliance on published work and the absence of experimentation and empirical verification. The next generation research and development tasks should be directed towards creating more transparent, efficient, secure, and human-centered Generative AI models aimed at supporting the deployment of Generative AI responsibly into various real-world applications.

8. CONCLUSION

Generative AI has become one of the most revolutionary technologies in the field of AI, allowing developers to generate realistic text, images, audio, video and more, with its uses spanning across numerous application areas. This review covered and discussed the various methods extensively, such as Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Transformer-based models, Diffusion Models, and Hybrid Architectures, which were presented along with their working principles, mathematical foundations, benefits, drawbacks, and applications. The study also examined the expanding role of Generative AI in the healthcare industry, NLP, computer vision, climate science, aerospace, finance, digital twins and intelligent search systems in cybersecurity. In addition to these developments, critical ethical, societal, and economic concesrns were explored, including algorithmic bias, misinformation, copyright concerns, governance of AI, transforming workforces, and computational costs. The rapidly evolving landscape of Generative AI is marked by several emerging trends such as multimodal AI, explainable AI, integration with reinforcement learning, the creation of AI safety and sustainable AI development. Overall, although considerable advances have been made, future research should focus on creating trustworthy, efficient, transparent and human-centred Generative AI systems to promote responsible innovation and maximise societal benefits.

References

1. Sai, S., Gaur, A., Sai, R., Chamola, V., Guizani, M., & Rodrigues, J. J. (2024). Generative AI for transformative healthcare: a comprehensive study of emerging models, applications, case studies, and limitations. Ieee Access, 12, 31078-31106.
2. Elhanashi, A., Essahraui, S., Dini, P., Paolini, D., Zheng, Q., & Saponara, S. (2026). Generative AI and the Foundation Model Era: A Comprehensive Review. Big Data and Cognitive Computing, 10(3), 94.
3. Yazdani, S., Singh, A., Saxena, N., Wang, Z., Palikhe, A., Pan, D., ... & Zhang, W. (2025). Generative AI in depth: A survey of recent advances, model variants, and real-world applications. Journal of Big Data, 12(1), 230.
4. Balasubramaniam, S., Chirchi, V., Kadry, S., Agoramoorthy, M., Gururama, S. P., Satheesh, K. K., & Sivakumar, T. A. (2024). The road ahead: emerging trends, unresolved issues, and concluding remarks in generative AI—a comprehensive review. International Journal of Intelligent Systems, 2024.
5. Sengar, S. S., Hasan, A. B., Kumar, S., & Carroll, F. (2025). Generative artificial intelligence: a systematic review and applications. Multimedia Tools and Applications, 84(21), 23661-23700.
6. Nova, K. (2023). Generative AI in healthcare: Advancements in electronic health records, facilitating medical languages, and personalized patient care. Journal of Advanced Analytics in Healthcare Management, 7(1), 115–131.
7. Sehwag, V., Hazirbas, C., Gordo, A., Ozgenel, F., & Ferrer, C. C. (2022). Generating high fidelity data from low-density regions using diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 11482–11491).
8. Foster, D. (2022). Generative deep learning (2nd ed.). O'Reilly Media.
9. Soltmann, R. J. R. B. (2024). Reinforcement learning enhanced generative replay for continual learning in diffusion models: A comprehensive study of existing and novel methodologies. Department of Electrical Engineering, University of Chile.
10. Xu, D., Fan, S., & Kankanhalli, M. (2023). Combating misinformation in the era of generative AI models. In Proceedings of the 31st ACM International Conference on Multimedia (pp. 9291–9298).
11. Wang, X., & Wu, Y. C. (2024). Balancing innovation and regulation in the age of generative artificial intelligence. Journal of Information Policy, 14, 385–416.
12. Becker, F., Skirzyński, J., van Opheusden, B., & Lieder, F. (2022). Boosting human decision-making with AI-generated decision aids. Computational Brain & Behavior, 5(4), 467–490.
13. Gatla, R. K., Gatla, A., Sridhar, P., Kumar, D. G., & Rao, D. S. N. M. (2024). Advancements in generative AI: Exploring fundamentals and evolution. In Proceedings of the International Conference on Electronics, Computing, Communication and Control Technology (ICECCC) (pp. 1–5).
14. Gatta, F., Giampaolo, F., Prezioso, E., Mei, G., Cuomo, S., & Piccialli, F. (2022). Neural networks generative models for time series. Journal of King Saud University – Computer and Information Sciences, 34(10), 7920–7939.

15. Fischer, A., & Igel, C. (2012). An introduction to restricted Boltzmann machines. In Proceedings of the 17th Iberoamerican Congress on Pattern Recognition (pp. 14–36). Springer.
16. Hinton, G. (2009). Deep belief networks. Scholarpedia, 4(5), 5947.
17. Béreux, N., Decelle, A., Furtlehner, C., Rosset, L., & Seoane, B. (2025). Fast training and sampling of restricted Boltzmann machines. In Proceedings of the 13th International Conference on Learning Representations (ICLR).
18. Béreux, N., Decelle, A., Furtlehner, C., & Seoane, B. (2023). Learning a restricted Boltzmann machine using biased Monte Carlo sampling. SciPost Physics, 14(3), Article 32.
19. Böhm, V., & Seljak, U. (2020). Probabilistic autoencoder (arXiv:2006.05479). arXiv.
20. Leeb, F., Bauer, S., Besserve, M., & Schölkopf, B. (2021). Exploring the latent space of autoencoders with interventional assays. In Advances in Neural Information Processing Systems (pp. 21562–21574).
21. Moghaddam, M. M., Boroomand, B., Jalali, M., Zareian, A., Daeijavad, A., Manshaei, M. H., & Krunz, M. (2023). Games of GANs: Game-theoretical models for generative adversarial networks. Artificial Intelligence Review, 56(9), 9771–9807.
22. Welfert, M., Kurri, G. R., Otstot, K., & Sankar, L. (2024). Addressing GAN training instabilities via tunable classification losses. IEEE Journal on Selected Areas in Information Theory, 5, 534–553.
23. Nankani, D., & Baruah, R. D. (2020). Investigating deep convolution conditional GANs for electrocardiogram generation. In Proceedings of the International Joint Conference on Neural Networks (IJCNN) (pp. 1–8).
24. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved training of Wasserstein GANs. In Advances in Neural Information Processing Systems (Vol. 30, pp. 5769–5779).
25. Raner, V. V., Joshi, A. D., Sawant, S. T., & Tamizharasan, P. (2025). Progressive growing of generative adversarial networks (PGGAN) approach to synthesize medical images. In The Impact of Algorithmic Technologies on Healthcare (pp. 157–176). Wiley.