Babylonian Journal of Artificial Intelligence
Vol. **(2023)**, 2023, **pp**. 10–14
DOI: https://doi.org/10.58496/BJAI/2023/003 ISSN: xxxxxx
https://mesopotamian.press/journals/index.php/BJAI

Research Article

A Survey of Generative Artificial Intelligence Techniques

Tam Sakirin[1,*], Siddartha Kusuma [2]

[1] *Computer Science Department, Faculty of science and information technology, International University, Phnom Penh, Cambodia*
[2] *Balaji Institute of Technology and Science, India*

Article info

Article History
Received 22 Dec 2022
Accepted 11 Feb 2023
Published 10 Mar 2023

Keywords
Generative Adversarial Networks (GANs)
Variational Autoencoders (VAEs)
Diffusion Models
Image Synthesis

Abstract

Generative artificial intelligence (AI) refers to algorithms capable of creating novel, realistic digital content autonomously. Recently, generative models have attained groundbreaking results in domains like image and audio synthesis, spurring vast interest in the field. This paper surveys the landscape of modern techniques powering the rise of creative AI systems. We structurally examine predominant algorithmic approaches including generative adversarial networks (GANs), variational autoencoders (VAEs), and autoregressive models. Architectural innovations and illustrations of generated outputs are highlighted for major models under each category. We give special attention to generative techniques for constructing realistic images, tracing rapid progress from early GAN samples to modern diffusion models like Stable Diffusion. The paper further reviews generative modeling to create convincing audio, video, and 3D renderings, which introduce critical challenges around fake media detection and data bias. Additionally, we discuss common datasets that have enabled advances in generative modeling. Finally, open questions around evaluation, technique blending, controlling model behaviors, commercial deployment, and ethical considerations are outlined as active areas for future work. This survey presents both long-standing and emerging techniques molding the state and trajectory of generative AI. The key goals are to overview major algorithm families, highlight innovations through example models, synthesize capabilities for multimedia generation, and discuss open problems around data, evaluation, control, and ethics. Please let me know if you would like any clarification or modification of this proposed abstract.

1. INTRODUCTION

The ability to automatically generate novel, realistic digital content is an emerging hallmark of artificial intelligence.[1] Generative algorithms create original images, audio, video, and 3D models that exhibit remarkable verisimilitude to actual data distributions. The outputs emerge synthesized completely from scratch rather than recasting existing samples. Spurred by open-source models like DALL-E 2 [2]painting images from text and stable diffusion delivering general image editing abilities surpassing prior expectations, public enthusiasm surrounds creative AI. Simultaneously, rapid research progress expands the horizons of what generative models can craft and constrain.

This survey charts the landscape of modern techniques molding generative model capabilities[3]. We focus on three major algorithmic[4] families powering generative systems today: 1) Generative adversarial networks which pit models against one another to enrich output quality, 2) Auto encoders that compress inputs into latent variables from which new samples are constructed through decoding, and 3) Autoregressive models predicting sequential output tokens based on learned conditional probabilities. Highlighting innovations in loss formulations, neural architectures, sampling methods, and pre-training procedures, we track progress in generating increasingly realistic image, audio, video and 3D model outputs. Diffusion models for images and Transformers for text notably illustrate breakthroughs realized by stacking components from across algorithm families. We further discuss the role of datasets in training as well as evaluation shortcomings for generatively produced content of indeterminate ground truth quality. Beyond synthesizing existing techniques, we outline newly emerging directions around control, multimodality, and responsible deployment that compose the future frontiers molding generative AI.

*Corresponding author. Email: erame2019ert@gmail.com

By reviewing fundamental approaches, paradigm-shifting models, tangible applications and open questions, this paper surveys the machinery powering generative AI'[5]s ascendance as one of the most rapidly developing and seminal fields within modern artificial intelligence research.

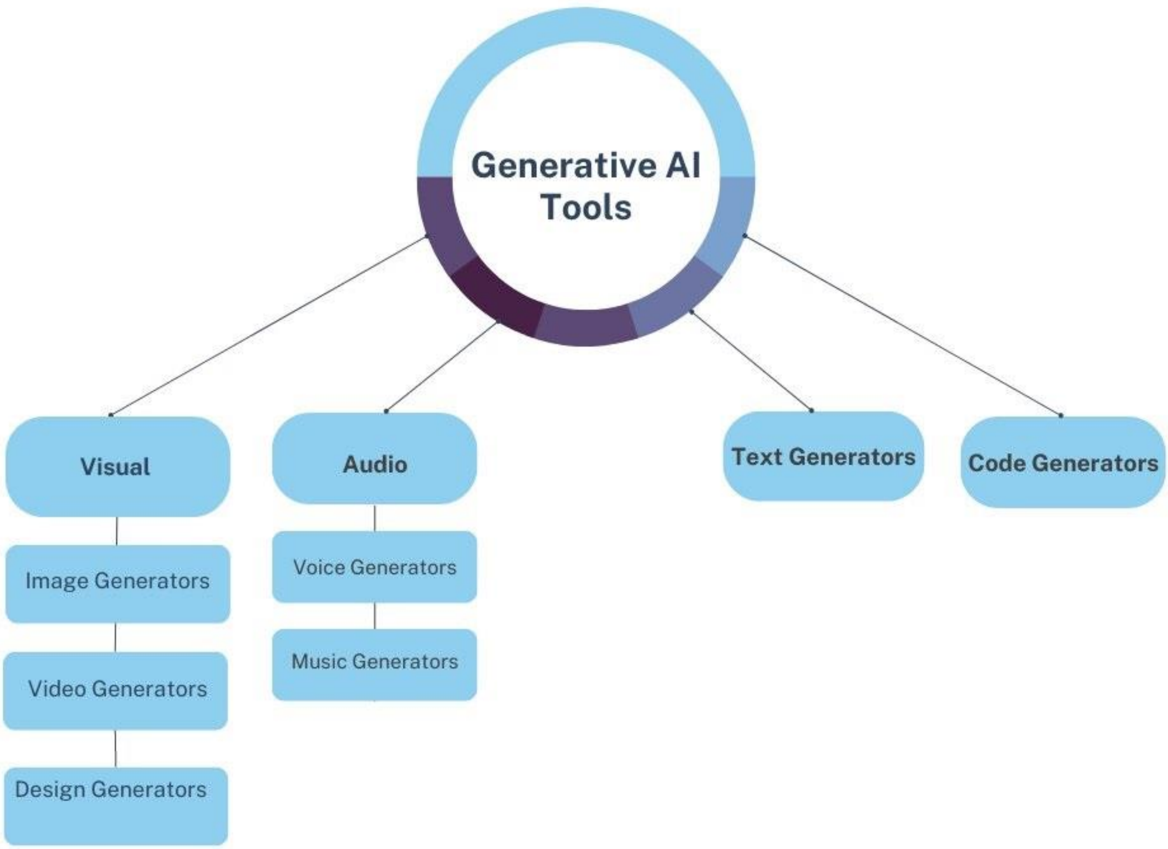

Fig. 1. Generative AI Tools

2. ALGORITHMIC APPROACHES

There are three predominant algorithm families that provide frameworks for developing generative AI systems - generative adversarial networks, auto encoders[6], and autoregressive models. Each approach has spawned numerous innovations addressing challenges of representation learning and sampling quality for synthesized content. We review foundational and emerging techniques within each category.

1. Generative Adversarial Networks

Generative adversarial networks (GANs)[7] compose generative models as a two-player mini-max game between competing networks. The generator attempts to produce novel samples resembling the target data distribution. It is pitted against an adversary - the discriminator - that classifies outputs as either real or synthetically generated. The networks evolve through this adversarial competition to enrich output quality. Architectural innovations enable generators to create strikingly realistic images, audio, video, and 3D models.

Early GAN formulations suffered from problems like mode collapse with limited output diversity. Enhancements like Wasserstein distance losses and progressive growing of generators address these. Conditional GANs incorporate external information to guide synthesis. GAN training remains intrinsically tricky, but approaches like style-based generators (Karras et al., 2020) sidestep typical generator architecture design challenges.

2. Autoencoders

Auto encoders compress input data into lower dimensional latent representations and then reconstruct inputs from these embedding. The encoder and decoder components are jointly trained. Variational autoencoders (VAEs)[8] impose additional constraints for latent vectors to follow desired distributions. Once trained, the decoder portion generates new samples by sampling latent vectors from distribution priors and mapping encodings to outputs.VAEs present a more stable approach over GANs but tend to produce blurrier outputs lacking high frequency details. Very Deep VAEs begin ameliorating such deficiencies. Adversarially Regularized Auto encoders augment VAE training with adversarial losses for refined realism.

3. Autoregressive Models

Unlike VAEs and GANs [9]which map latent vectors to outputs in parallel, autoregressive models factorize joint probabilities to predict portions of output sequences iteratively based on previously generated sections. Transformers shine light on capturing long range dependencies. Autoregressive models produce strong results across modalities but requireexpensive iterative calculations during sampling. Efforts like Parallel WaveNet accelerate audio synthesis. Masked autoencoders like BERT distill bidirectional representations useful for conditional generative tasks. Models including GPT-3 actively employ Transformers to attain state-of-the-art natural language generation, illustrating the scalability of autoregressive approach.

3. GENERATING REALISTIC IMAGES

The ability to automatically create and edit photorealistic images presents a seminal challenge for generative algorithms. Impressive progress has been attained through combining adversarial objectives, encoder-decoder architectures, and sequential modeling under a common framework - diffusion models. Meanwhile, progress continues along numerous fronts to enrich detail, resolution, and fidelity of image synthesis.

1. Diffusion Models

Diffusion models gradually corrupt data then train networks to reverse that process towards pristine outputs. DDPM and DDIM establish strong performance by modeling stochastic differential equations governing corrupting Markov chains. Latent vector concatenation and adversarial losses in ADM boost sample quality. GLIDE and DALL-E pretrain text-to-image generators as masked autoencoders before fine-tuning on diffusion objectives.

Stable Diffusion adapts latent text encodings from GLIDE into the sampling loop of DDPM image synthesis. This composability of diverse algorithmic techniques under the diffusion framework drives Stable Diffusion's flexible manipulation capabilities. Meanwhile, edge-generation refinement networks in DALL-E 2 and imaginative text-conditioning in Parti enrich creative expression.

2. Ongoing Directions

Active progress continues in numerous areas. GANs like StyleGAN-3 generate intricate high-resolution images through adaptive discriminator augmentations and noise conditioning. BigGANs and StackGANs demonstrate scaling up GAN image synthesis with added layers and parameters. Techniques for semantic image editing manipulate latent spaces for changes aligned with human-intuitive features. Efforts also accelerate training on huge visual datasets. These extend the horizon of photorealistic image synthesis.

3. Generating Audio, Video and 3D Renderings

While image generation garners significant attention, generative models also empower realistic synthesis across modalities like audio, video, and 3D shapes which introduce new intricacies around capturing temporal dynamics and high dimensionality.

4. Audio Synthesis

Neural audio synthesis models generate raw waveform samples directly. WaveNet first demonstrated this with an autoregressive approach. Follow ups like WaveGAN and WaveRNN improved efficiency. Recent models leverage adversarial training like GanSynth producing minutes long piano music coherent to human judgement. Parallel approaches also help scale up synthesis, evidenced by Jukebox (Dhariwal et al., 2020) generating genre-specific music conditioning on metadata.

5. Video Generation

Synthesizing realistic videos poses complex challenges of coordinating coherent visual features and dynamics across frames. Progress has occurred through extending GAN architectures. MoCoGAN decomposes motion and content into latent representations for video generation. StyleGAN-V adapts image based stylegan for video by predicting motion between static frames. Video prediction models like DVD-GAN-FP forecast plausible future frames by learning temporal dynamics.

6. Model Generation

Generative techniques for 3D content create usable models for simulation, gaming, VR and 3D printing applications. Variational approaches like GRAF (Schwarz et al., 2020) couple a graph-network encoders for shape structure with mesh decoders. Adversarial training improves outputs for domain specific 3D generation like cars, chairs, or clothing. Part assembly models like GAF construct 3D objects out of salient component pieces. Scalability however remains an open challenge. As generative models expand across modalities, key issues around temporal consistency and evaluation rigor become more prominent to address. Tackling these areas could enable generative AI to meaningfully impact a wider range of industries and applications.

4. DATASETS AND ETHICS

Behind groundbreaking generative models lie massive datasets advancing model capabilities. However, reliance on large corpora of data also introduces questions around ethical sourcing, biases, and malicious use of synthetic media.

1. Driving Datasets

Many influential datasets underpin domains like image and language. FFT-based CLIP leverages 400 million image-text pairs from internet sources to excel at image classification. LAION-5B's 5 billion image-text tuples (Schuhmann et al., 2022) fuel the DALL-E model family. For audio, LibriTTS's 585,000 hours of labeled speech (Zen et al., 2019) powers various vocoders. And Common Voice's 10,000 hours of donation-based recordings demonstrate more ethical curation. These datasets drove innovation but some prompted appropriation concerns, highlighting the need for responsible collection and documentation. Continual progress will require high-quality, ethically sourced datasets, necessitating standardized guidelines.

2. Emerging Issues

While generative models produce new data combinations, risks around data bias and misuse of synthetic content remain. Documented failures of toxic text detection in language models prove that harmful biases persist in training data. Deepfakes manipulating images and videos also introduce threats like political disinformation. Usage policies and technical protections are active discussion points. There also exists potential for generative models to help, such as using synthetic data to augment minority groups lacking representation. Overall there is a crucial need for ethical foresight to guide responsible advancement of generative AI alongside societal readiness. Open questions cover a wide range swath from the nature of data curation powering future progress to governance of applications. Principled stances steering access, monitoring for harms, and inclusive development carry models beyond technical capabilities alone towards enacting positive real-world change.

5. CURRENT FRONTIERS AND FUTURE OUTLOOK

While modern techniques already enable remarkable generative applications, many compelling research questions remain unresolved on the path towards more advanced creative AI systems. Several open challenges currently confront the field across dimensions of quality, capability, and control. Improved evaluation metrics measuring output realism and diversity lag behind synthesis advances. Training procedures for multimodal inputs and outputs remain less explored, though promising for more representational embodiments of creativity. And dynamics governing emerging behavior and aligned value systems pose complex questions around controlling model development and deployment directions.

In the near term, techniques blending algorithmic innovations across families could compound strengths while ameliorating limitations of individual approaches. For instance, autoregressive guidance of diffusion models to sharpen text-conditional image generation based on progressive feedback. Dynamic composability also promotes accessibility to guide rapid innovation. In the longer term, lifecycle considerations around data curation, carbon footprint, commercialization and collaboration must complement technical progress. Real-world integration further necessitates international standards, safety guidelines and monitoring against potential harms.Ultimately, generative AI's profound creative potential compels informed, inclusive advancement of capabilities driven by humanistic values. Technique is not an endpoint - it must enable positive real-world impact. This survey highlights achievements so far while framing directions needed to responsibly realize AI creativity for social good. With ethical foundations and inspired vision, generative models hold immense promise to enhance how we communicate, create and build mutual understanding.

6. CONCLUSION

This survey has traced the current landscape and frontiers of generative artificial intelligence - where algorithms autonomously synthesize novel realistic content without simply recasting existing samples. We structured an analysis across three predominant algorithm families powering advances in this machine creativity. Generative adversarial networks were shown to produce striking creations across modalities like images, audio, and video by pitting generator models against adversarial discriminators. Variational autoencoders enable sampling from learned latent spaces using encoder-decoder neural network pipelines. Autoregressive models iteratively predict sequence outputs while capturing long-range dependencies. Technique consolidations under the diffusion model framework meanwhile drive recent versatile image manipulation breakthroughs. We highlighted innovations in loss formulations, neural architectures, and pre-training strategies that enriched generative modeling capabilities over time. Performance improvements were charted through example models like StyleGAN, WaveNet, and GPT-3 which represent paradigm shifts for image realism, audio generation, and language modeling respectively. Beyond common datasets powering progress, crucial open questions around evaluation practices, multimodal synthesis, responsible model development, and real-world integration remain active frontiers requiring interdisciplinary perspectives. Ultimately, accelerating generative AI's capabilities syncs tightly with deliberating its ethically guided integration towards positive change. While technical machinery molded the era of machine creativity we surveyed here, it is proactive translation of these tools to advance societal benefit that defines the vital path ahead. Generative AI thereby warrants informed, inclusive progress in line with human values - for technological potential is only optimally realized through respecting the many dimensions of lives it stands to transform.

Funding

None.

Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The authors would like to express their gratitude to the Al-Iraqia University, the Informatics Institute for Postgraduate Studies, and the Al Salam University College for their moral support. Please accept my sincere gratitude for the useful recommendations and constructive remarks provided by the anonymous reviewers.

References

[1] D. Baidoo-Anu and L. O. Ansah, "Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning," *Journal of AI*, vol. 7, no. 1, pp. 52-62, 2023.

[2] G. Harshvardhan, M. K. Gourisaria, M. Pandey, and S. S. Rautaray, "A comprehensive survey and analysis of generative models in machine learning," *Computer Science Review*, vol. 38, p. 100285, 2020.

[3] S. Noy and W. Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Available at SSRN 4375283*, 2023.

[4] K. Wach *et al.*, "The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT," *Entrepreneurial Business and Economics Review*, vol. 11, no. 2, pp. 7-30, 2023.

[5] M. Castelli and L. Manzoni, "Generative models in artificial intelligence and their applications," vol. 12, ed: MDPI, 2022, p. 4127.

[6] Z. Bahroun, C. Anane, V. Ahmed, and A. Zacca, "Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis," *Sustainability*, vol. 15, no. 17, p. 12983, 2023.

[7] D. K. Kanbach, L. Heiduk, G. Blueher, M. Schreiter, and A. Lahmann, "The GenAI is out of the bottle: generative artificial intelligence from a business model innovation perspective," *Review of Managerial Science*, pp. 1-32, 2023.

[8] V. Ratten and P. Jones, "Generative artificial intelligence (ChatGPT): Implications for management educators," *The International Journal of Management Education*, vol. 21, no. 3, p. 100857, 2023.

[9] 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," *Frontiers in artificial intelligence*, vol. 4, p. 604234, 2021.