

Generative Artificial Intelligence: A Systematic Review and Applications

Sandeep Singh Sengar^{1*}, Affan Bin Hasan¹, Sanjay Kumar²,
Fiona Carroll¹

¹Cardiff School of Technologies, Cardiff Metropolitan University,
Cardiff, CF5 2YB, United Kingdom.

²Department of Computer Science and Engineering, Delhi Technological
University, New Delhi, 110042, India.

*Corresponding author(s). E-mail(s): SSSengar@cardiffmet.ac.uk,
Phone: +44-7780390651;

Contributing authors: ahasan@cardiffmet.ac.uk;
sanjay.kumar@dtu.ac.in; fc Carroll@cardiffmet.ac.uk;

Abstract

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.

Keywords: Generative Artificial Intelligence, Generative Adversarial Networks, Diffusion, Segmentation, Variational Autoencoder, Transformers.

1 Introduction

The recent advancement in Artificial Intelligence has been mainly the result of Generative Artificial Intelligence (often referred to as Generative AI or GenAI) being introduced. Generative AI encompasses artificial intelligence systems with the ability to create text, images, or various forms of media through the utilization of generative models. These models acquire an understanding of the underlying patterns and structures within their training data, subsequently producing fresh data that share similar traits and characteristics. The motivation of this systematic review is to gather, evaluate, and synthesize existing research on a GenAI. This paper presents a systematic review that highlights key applications and variations of the architecture of Generative Artificial Intelligence models and their performance. We conducted this review to (a) understand the state-of-the-art generative AI techniques including summarizing key methodologies, algorithms, and findings across a range of studies (b) systematically review a large body of literature, which includes emerging trends, common challenges, and recurring patterns in the development and application of generative AI techniques (c) compare and contrast different generative AI approaches, such as Autoencoders, Generative Adversarial Networks, Transformers, and Diffusion models (d) explore successful applications of generative AI such as image translation, video synthesis and generation, natural language processing, knowledge graph generation, etc. (e) identify ethical challenges and propose solutions for responsible AI development.

In this research, we outline the most recent research and advancement in the field of Generative Artificial Intelligence. It details the approach used to navigate and analyze cutting-edge developments, ensuring a comprehensive and insightful review of the current landscape in Generative AI. The following criteria were applied for searching the used research papers.

Time Period: This paper presents a comprehensive overview of the advancements and applications of Generative AI, focusing on significant developments between 2018 and 2023. Additionally, it offers a concise historical perspective, tracing the evolution of foundational models from 2012 to 2018, which laid the groundwork for the current state of Generative AI techniques. This historical context enriches the understanding of the field's rapid progression and its burgeoning applications.

Keywords: This paper employs a targeted keyword search strategy, incorporating specific terms such as 'Generative Adversarial Networks', 'Transformers', 'Variational Autoencoders', and 'Diffusion Models'. This approach also includes searching for advancements in 'image translation', 'video synthesis', and various applications of Generative AI in 'natural language processing' and 'knowledge graph generation'. This methodology ensures a focused and comprehensive review of the latest developments in the field of Generative AI.

Databases: The work primarily sources relevant literature from Google Scholar, focusing on the specified timeframe. It selectively includes research that showcases advancements in generative models. This criteria ensures the inclusion of studies where developed models were rigorously tested on well-recognized datasets, and where results are communicated effectively and clearly. This approach guarantees that the paper presents a detailed and credible overview of significant developments in the field of Generative AI.

Inclusion Criteria: This work exclusively incorporates peer-reviewed papers, conference, and journal papers that are written in English. It emphasizes studies that highlight either significant advancements or innovative applications in the realm of Generative AI, ensuring that the focus remains on cutting-edge and impactful developments within this field.

Exclusion Criteria: This paper meticulously filters its sources, excluding non-peer-reviewed materials, papers not written in English, and studies that fall outside the 2012-2023 timeframe. Additionally, it deliberately omits any papers that do not directly contribute to the advancement or understanding of Generative AI, ensuring a focused and relevant academic discourse.

Evaluation Criteria: In each subsection evaluating the advancements in Generative AI techniques, the paper compares the performance of various models using standardized datasets commonly cited in the field. This comparison focuses on how different state-of-the-art models perform on these datasets, providing a clear and consistent basis for assessing the progress and effectiveness of these techniques in their respective domains.

The following is the summary of the major contributions of our work-

- **Paradigm Shift in Artificial Intelligence:** The paper discusses the paradigm shift in artificial intelligence and highlights the significant impact of generative models in the field of machine learning.
- **Historical Context:** The paper includes a section that gives a straightforward overview of how key AI models have developed from 2012 to 2018, helping to better understand how the field has grown and changed over time.
- **Real-World Uses of Generative AI:** The paper describes how Generative AI is used in different areas like image translation, diagnosing medical conditions, combining text and images, processing natural language, etc.
- **Systematic Review of Generative AI:** The work provides a comprehensive review and analysis of recent advancements in Generative AI, focusing on techniques and applications, including application-specific models. We have also provided information on relevant datasets for each used application.
- **Impact on Language and Image Translation:** The paper discusses the major impact of generative AI in language generation with large language models and in the field of image translation.
- **Responsible AI Principles:** The paper ends with a discussion on Responsible AI principles and ethical considerations necessary for the sustainability and growth of generative models.

After this introductory section, Sections 2 and 3 review the basic early architecture of Generative adversarial Networks and their variants. Section 4 deeply explores the recent applications and the advancements in Generative AI application-specific techniques. Section 5 provides the Challenges and opportunities of Generative AI. Lastly, Section 6 concludes the work and highlights the future directions of generative AI.

2 What is Generative Artificial Intelligence?

As discussed Generative Artificial Intelligence refers to artificial intelligence systems with the capability to create text, images, or other forms of media through the utilization of generative models. These models acquire an understanding of patterns and structures within their training data, subsequently generating novel data with akin characteristics. Generative Artificial Intelligence encompasses various types, each tailored for specific tasks or forms of media generation. The following are some of the more well-known types: Generative Adversarial Networks (GANs) [71], Transformer-based Models (TRMs) [133], Variational Autoencoders (VAEs) [62], and Diffusion models (DMs) [68], to name a few. The following sections will discuss these in more detail.

2.1 Generative Adversarial Networks

A generative adversarial network (GAN) is a class of machine learning framework and a prominent framework for approaching generative AI. The aspect that is novel in this generative adversarial network set-up is that it does not depend upon heavily annotated training data. Moreover, the architecture that it affords is quite unique from the conventional Deep Neural Networks [27]. Indeed, it consists of two major components named *Generator* and *Discriminator*. The main operation of the generator is to keep on generating the fake data using the noise while the purpose of the discriminator is to distinguish whether the generated image is real or fake. The discriminator is trained using the real images of the domain that the generator is trying to synthetically produce and the discriminator's sole purpose is to identify whether the output produced by the generator is fake or not. The overall system is based on the zero-sum game dynamics, the winner will remain unchanged and the loser model each time has to modify its parameters, it will keep on doing this until the discriminator is unable to detect whether the generator output is fake or not [44]. The sole purpose of this is to build a powerful generator model that generates synthetic data that looks real.

Fig. 1 demonstrates how the generator and discriminator work together. The generator aims to deceive the discriminator by providing the synthetically generated image with the objective that it is proven real. The discriminator discerns between genuine and counterfeit images and generates the output signal. This output signal then goes to both the generator and discriminator, allowing the generator to produce better synthetic output. And, in case the discriminator fails to prove the image is fake, it also uses the signal to change its weights to give better predictions. In this entire architecture, it is important to note that only the discriminator has access to the real image, synthetic image, and its own signal output while the generator only learns from the output signal of the discriminator [27].

During the initial stages of development, Generative Adversarial Network (GAN)-based models encountered significant challenges in their training process. These difficulties primarily revolved around issues like training divergence and model collapse [87]. Training divergence refers to situations where the GAN's generator and discriminator fail to achieve a stable equilibrium during training, leading to oscillations and unreliable model outputs. This problem results in inconsistent and sub-optimal generation performance, hindering the GAN's ability to produce high-quality samples. On

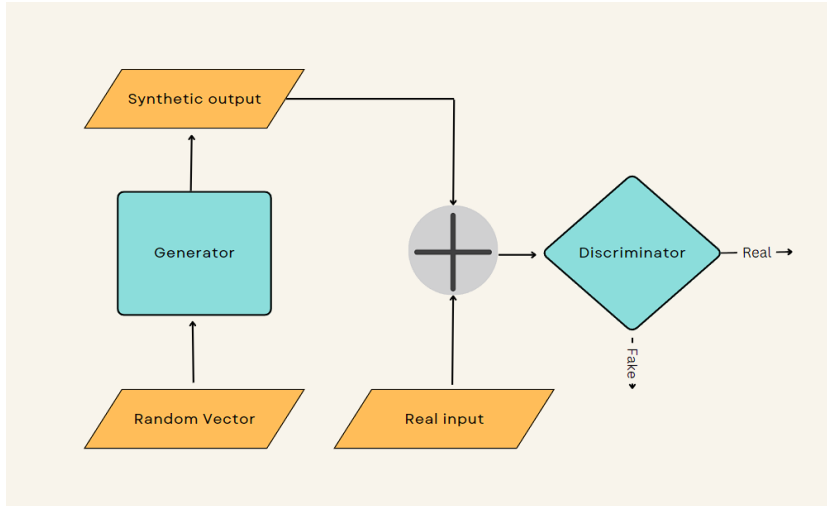


Fig. 1 GAN- Generator and discriminator working

the other hand, model collapse occurs when the GAN's generator produces limited and repetitive outputs, ignoring a large portion of the data distribution. This phenomenon causes the generator to focus on a small subset of data points, resulting in a lack of diversity and novelty in the generated samples. Addressing these challenges has been a main focus in the advancement of GAN-based models, with numerous research efforts aimed at improving stability, convergence, and diversity during the training process. As a result, substantial progress has been made, leading to the development of more robust and effective GAN architectures, which have significantly enhanced the performance and applicability of these generative models.

2.2 Transformers

Further Generative AI techniques called *Transformers* were introduced by Vaswani et al. [134]. This breakthrough architecture laid the foundation for various tasks, including machine translation and language generation, and it continues to influence subsequent neural network designs. The paper's emphasis on attention mechanisms highlighted their pivotal role in sequence-to-sequence tasks, advancing the state of the art. Transformers use both the self-attention and Multi-Head Attention mechanisms to learn the dependencies between the objects regardless of the distance between them and to learn the different relations and patterns between the input respectively. Often in Natural Language Processing, these methods are combined with positional encoding added to the input sequence to make the transformers keep track of the position of a specific word in an input sequence. Transformers are commonly used to build Generative AI Models such as Generative Pre-trained Transformers (GPT) models which are capable of generating coherent and contextually relevant text [109]. Bidirectional Encoder Representations from Transformers (BERT) and Open AI GPT are based on transformers.

2.3 Variational Autoencoders

Another model in the field of generative AI is Variational Autoencoders (VAEs), introduced by Kingma et al. [65]. As the name suggests the VAEs consist of an encoder and a decoder. The purpose of an encoder is to encode the given input in a lower dimension called *latent space* and the decoder decodes that latent output of the encoder into its original input shape. During this whole process, variation is introduced to the latent space by using the standard Gaussian distribution. The main goal is to achieve the output with a similar mean and variance as the given input after the introduction of the variance. This provides a structured way to learn meaningful representations of data and then generate new samples from that data distribution.

2.4 Diffusion Models

The Diffusion Models have been designed to improve the performance of the Simple Generative Adversarial Network, the technique was introduced by Salimans et al [115]. At a later stage, Kingma et al [64] introduced a variant of the diffusion model called Inverse Autoregressive Flow (IAF) as a building block for generative models. IAF is a type of normalizing flow. This is a type of generative model that aims to learn complex probability distributions by transforming a simple base distribution into the target distribution through a series of invertible transformations.

3 Evolution of Generative AI Models: A Look at Earlier Variants

3.1 Earlier GAN Variants

In the early stages of the introduction of Generative Artificial Intelligence models the major issue that researchers were facing was the convergence problem of generative models [87]. To avoid this problem, different approaches were adopted by the researchers to make the GAN more stable (e.g., by understanding the behavior of GAN training). In detail, Mescheder et al. [88] explained the analysis of local convergence and stability properties during the training of GAN. This involves an examination of the eigenvalues of the Jacobian matrix associated with the gradient vector field. Specifically, when the equilibrium point is characterized by solely negative real-part eigenvalues in the Jacobian, GAN training demonstrates local convergence, specifically when utilizing relatively small learning rates. However, the situation changes when the eigenvalues of the Jacobian are situated on the imaginary axis. In such cases, the local convergence of GAN training is generally compromised. Also, it is important to note that if the eigenvalues are in proximity and not directly on the imaginary axis, the training algorithm may necessitate exceedingly small learning rates to achieve convergence [88].

The study by Mescheder et al. [88] identified instances of eigenvalues near the imaginary axis in practical scenarios. This observation does not definitively address whether such proximity to the imaginary axis is a prevalent phenomenon. Furthermore, it does not conclusively establish whether these eigenvalues are the fundamental cause

behind the training instabilities that practitioners commonly encounter in their GAN training endeavors. Moreover, Nagarajan et al. [93] contributed a partial response to this query. They demonstrated that in the context of absolutely continuous data and generator distributions, these findings establish that GANs exhibit local convergence for sufficiently small learning rates. However, to emphasize, this assertion relies on the premise of absolute continuity.

Goodfellow et al. [44] presented the basic GAN architecture, other researchers also advanced more variants with some architectural differences. However, due to the potentially limiting overlap between real and generated data distributions, the Jensen-Shannon divergence presented in the objective function can become a constant value. It is this phenomenon that contributes to the challenge of the vanishing gradient, hindering effective training of GANs when employing gradient descent methods. To address the vanishing gradient problem, the Wasserstein GAN (W-GAN) was introduced [6], using the EarthMover distance instead of the Jensen-Shannon divergence to compare real and generated data distributions. W-GAN employs a critic function f with a Lipschitz constraint as its discriminator, significantly improving GAN training stability. However, W-GAN may still face issues like suboptimal sample generation and occasional convergence problems in specific cases. In order to restrict the discriminative capacity of the discriminator, an alternative approach has also been introduced by [105] in the form of Loss-Sensitive GAN (LS-GAN). Both W-GAN and LS-GAN retain the fundamental GAN architecture.

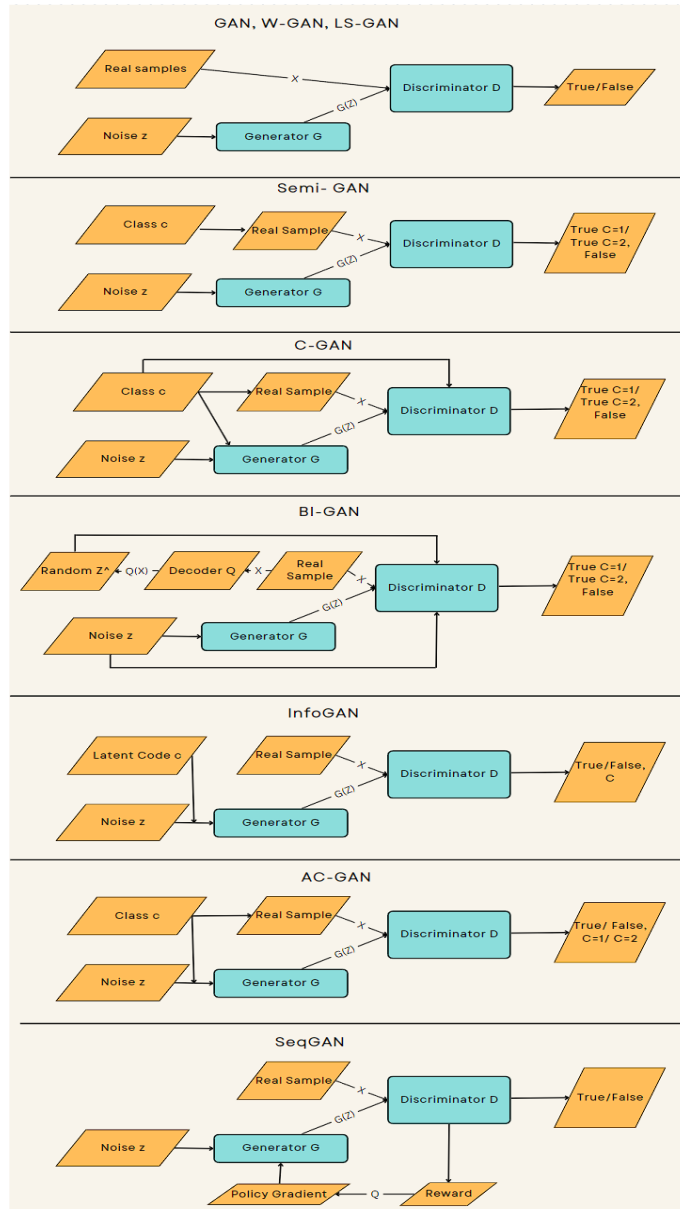


Fig. 2 Basic GANS Variants

Moreover, Qi et al. [97] introduced the Semi-GAN model, which involves the incorporation of real data labels into the discriminator’s training process. Additionally, semi-GAN is an approach involving the integration of auxiliary information ‘y’ into both the generator ‘G,’ the discriminator ‘D,’ and the real data ‘x’ for the discriminator [91]. This auxiliary information can encompass labels or other supplementary data.

In the context of conventional GANs, the primary objective revolves around acquiring a generative model capable of mapping latent variable distributions to intricate real data distributions. Expanding upon this concept, Donahue et al. [33] propose Bidirectional GANs (BiGANs) to facilitate the mapping of real data to the latent variable space, thereby enabling feature learning. BiGANs extend the fundamental GAN structure by incorporating an additional decoder 'Q,' which facilitates the transformation of real data 'x' into the latent space. Consequently, this modification transforms the optimization problem into the form $\min_{G,Q} \max_D f(D, Q, G)$.

Chen et al. [21] introduce InfoGAN to capture mutual information between a subset of latent variables and observed data.

In InfoGAN, the correlation is quantified using $I(c; G(z, c))$, where c is a latent code and $G(z, c)$ is the generated output. The objective function is:

$$\min_G \max_D \{f(D, G) - \lambda I(c; G(z, c))\}$$

Here, $f(D, G)$ includes the adversarial term and penalty term $\lambda I(c; G(z, c))$. The goal is to minimize the generator's loss while maximizing the discriminator's loss with respect to mutual information.

Due to the challenge of computing $p(c|x)$, a lower bound estimation is used through variational information maximization.

Odena et al. [98] introduce the auxiliary classifier GAN (AC-GAN) approach tailored for semi-supervised synthesis. Their formulated objective function comprises two integral components: the logarithmic likelihood related to the accurate data source and the corresponding accurate class. The essence of AC-GAN lies in its capacity to seamlessly integrate label information into the generator and to adapt the discriminator's objective function accordingly. This integration yields noticeable enhancements in the generative and discriminative capabilities of the GAN framework.

In another context, Yu et al. [150] introduced SeqGAN, a pioneering framework for sequence generation using GANs. It extends GANs to handle discrete token sequences, treating the generator as a stochastic policy in reinforcement learning. SeqGAN employs policy gradient-based mechanisms to enhance sequence generation by effectively propagating errors from the discriminator. These advancements build upon the foundational work of GANs [44].

3.2 Earlier Transformer Variants

The concept of Transformers was introduced by Vaswani et al. [134]. It was a revolutionary step in the field of generative AI specifically in natural language processing and generating synthetic content. The basic concept of the Transformers was introduced in [125] by Sutskever et al as a sequence modelling technique. The basic technique of pretraining transformers was introduced and used as a state-of-the-art technique by [106]. This was used in answering different queries and also used as a chatbot to give results that are competitive and accurate. Indeed, these early Transformer Developments paved the way for State-of-the-Art NLP Chatbots.

3.3 Earlier Variational Autoencoder Variants

Variational Autoencoders [72] is one of the oldest techniques of unsupervised learning and generative modeling. The foundational work of Variational Autoencoders was done by Kingma et al in [66]. The Variational Autoencoders model combines the probabilistic modeling with the basics of Autoencoders. This concept not only learns the properties of the latent space but they also learn the probabilistic distribution of it, which gives them the ability to generate new synthetic data samples. Early variants of autoencoders also include denoising autoencoders which use denoising techniques whilst trained locally to get rid of corrupted versions of their inputs [135]. Moreover, [84] used the auto-encoders coupled with the convolutional network to solve the image recognitional problems. Without a doubt, the pioneering work of [66] has paved the way for numerous subsequent developments and applications of VAEs in a wide range of domains, including image generation, natural language processing, and more.

3.4 Earlier Diffusion Model Variants

Diffusion-based models employ a sequential diffusion process to iteratively transform simple data distributions into complex, high-dimensional ones. The Non-Linear independent component estimation (NICE) introduced the concept of invertible transformations as a foundation for generative artificial intelligence [31]. This was followed by Real NVP (Real Non-Volume Preserving), which expanded the capabilities by incorporating neural networks into the transformation process [32]. Additionally, Glow (Generative Latent Optimization) extended these ideas to high-resolution image generation, highlighting the potential of diffusion-based models in computer vision [63]. Furthermore, Diffusion Probabilistic Models (DPMs) leveraged the diffusion process to model the likelihood of data samples, making it an essential contribution to the development of diffusion-based generative models. Continuous-time flows (CTFs) diffusion models ventured into continuous-time modelling using stochastic differential equations [45]. These earlier works have laid the foundation for an exciting and rapidly evolving field of generative modelling using diffusion-based techniques.

4 Advancements in Generative AI and Their Diverse Applications

4.1 Generative AI for Image Translation

Image translation [44] is becoming a rapidly growing technology, particularly within the realm of medical applications. This innovation holds remarkable potential, not only in terms of cost-saving implications related to equipment usage but also in the facilitation of informed medical decisions.

The performance of generative AI models, particularly in the subfield of image translation, is typically assessed using specialized datasets. Among these, two notable datasets stand out: ImageNet [147], ClebA [82], and in the field of medical science: MIMIC [54], BRATS [86], FastMRI [129] and ChestX-ray [139] Each of these datasets is uniquely designed to challenge and evaluate the models' abilities to accurately and

effectively translate images, providing a comprehensive benchmark for their performance capabilities. These datasets are freely accessible to researchers for testing their models, under certain conditions. Users must properly cite the source of the dataset in their work. For datasets containing medical data, researchers are required to sign a Data Use Agreement. This agreement sets forth guidelines on the appropriate usage and security of the data and strictly prohibits any attempts to identify individual patients. This ensures that while fostering innovation and research, the datasets are used ethically and responsibly.

The utilization of AI-driven image translation yields images that are not only more polished and precise but also empower medical professionals with a heightened ability to discern critical information [146]. Yan et al. proposed a GANs-based model that uses the *Swin Transformers* in the Generator. The Swin Transformer represents a notable stride forward in the evolution of architecture. Its most remarkable enhancement entails the replacement of the conventional multiple self-attention (MSA) modules with an innovative shift window-based module while keeping the remaining layers largely unchanged. This transformer-based generator allows for the production of the output content which is the same as source images and the same information required by the target image. They tested the model using the BraTs2018 [86] and FastMRI [118] datasets. The Swin-based Transformers method attains its highest level of performance in the specific task of converting T1 mode to T2 mode images using the clinical brain MRI dataset. Moreover, they conducted evaluations using the unpaired *BraTs2018* dataset (see the results depicted in the Figure 3). These highlight that the innovative MMTrans approach stands out as the leader in terms of translation performance.

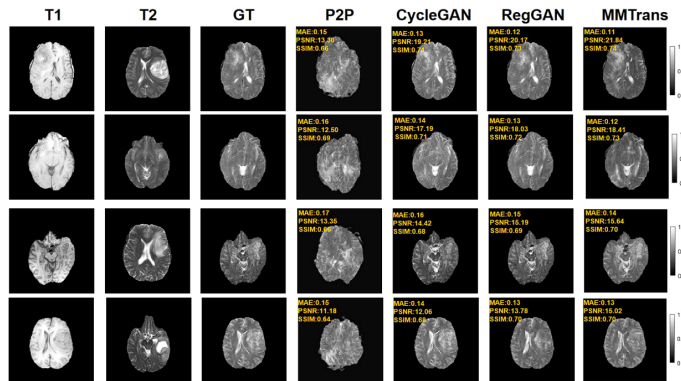


Fig. 3 qualitative outcomes obtained from various translation techniques employed to generate T2 images from T1 images within the unpaired BraTs2018 dataset, Source: [146]

Indeed, Figure 3 shows that the introduced MMTrans method performs better in translating the image when compared to other methods such as Pix2Pix [52], CycleGAN [159] and RegGAN [114]. The Image of MMTrans is closest to the Ground Truth (GT).

Furthermore, Dar et al. [28] used Conditional GANs to solve the problem of Image translation in MRI. This variant of GANs was introduced for image-image translation with the conditional label given to both generator and discriminator to instruct about what they have to forge and to predict real or fake respectively [52]. In fact, there are two types of GAN variants present for the purpose of image translation, Pix2Pix GAN is a conditional GAN [52], where the generator takes both an input image and a target condition as input and then generates an output image that adheres to the specified condition. However, for these types of GANs pixel-aligned images are required which is quite difficult to acquire [146]. Moreover, unpaired GANs do not require corresponding pairs of images for training. Instead, they focus on learning the mapping between two domains by using cycle consistency as a constraint [131].

Indeed, unpaired image translation presents a significant challenge. The objective is to ensure that translating an image from one domain to another should not compromise its fundamental characteristics (e.g., allowing for a seamless reversion back to its initial state)[159]. To address this issue, the Cyclic Generative Adversarial Network (Cyclic GAN) was developed [159]. Torbunov et al. [131] developed a Vision Transformer based GAN (UVCGAN). This works on the principle of cyclic GANs and without sacrificing the image regeneration capabilities gives better results than previous simple cyclic models. The evaluation of image-to-image translation performance commonly employs two widely accepted metrics, namely Frechet Inception Distance (FID) [46] and Kernel Inception Distance (KID) [13]. These metrics quantify the similarity between the translated images and those within the target domain, with a lower score indicative of higher similarity. UVCGAN model’s superior performance is evident across most image-to-image translation tasks, as illustrated in Table 1. Operating similarly to a CycleGAN-like model, their approach consistently produces translated images that exhibit strong correlations with the input images, capturing essential aspects like hair color and facial orientations (as exemplified in Figure 4).

The high-quality result produced by UVCGAN in Figure 4 is paramount for enhancing scientific simulations. It can be observed that translations generated by ACL-GAN and Council-GAN tend to overly emphasize features that aren’t pivotal for achieving the intended translation, such as non-essential attributes like hair color, background color, and length. Even in some cases, Council-GAN changed the background.

Table 1 FID and KID scores of UVCGAN and other models. Lower is better, Source:[131]

Model	Selfie to Anime		Anime to Selfie	
	FID	KID ($\times 100$)	FID	KID ($\times 100$)
ACL-GAN	99.3	3.22 ± 0.26	128.6	3.49 ± 0.33
Council-GAN	91.9	2.74 ± 0.26	126.0	2.57 ± 0.32
CycleGAN	92.1	2.72 ± 0.29	127.5	2.52 ± 0.34
U-GAT-IT	95.8	2.74 ± 0.31	108.8	1.48 ± 0.34
UVCGAN	79.0	1.35 ± 0.20	122.8	2.33 ± 0.38
Model	Male to Female		Female to Male	
	FID	KID ($\times 100$)	FID	KID ($\times 100$)
ACL-GAN	9.4	0.58 ± 0.06	19.1	1.38 ± 0.09
Council-GAN	10.4	0.74 ± 0.08	24.1	1.79 ± 0.10
CycleGAN	15.2	1.29 ± 0.11	22.2	1.74 ± 0.11
U-GAT-IT	24.1	2.20 ± 0.12	15.5	0.94 ± 0.07
UVCGAN	9.6	0.68 ± 0.07	13.9	0.91 ± 0.08
Model	Remove Glasses		Add Glasses	
	FID	KID ($\times 100$)	FID	KID ($\times 100$)
ACL-GAN	16.7	0.70 ± 0.06	20.1	1.35 ± 0.14
Council-GAN	37.2	3.67 ± 0.22	19.5	1.33 ± 0.13
CycleGAN	24.2	1.87 ± 0.17	19.8	1.36 ± 0.12
U-GAT-IT	23.3	1.69 ± 0.14	19.0	1.08 ± 0.10
UVCGAN	14.4	0.68 ± 0.10	13.6	0.60 ± 0.08

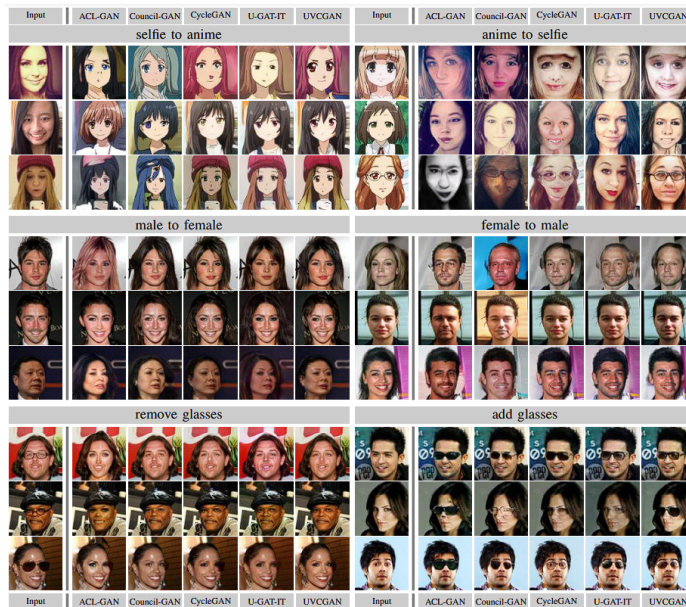


Fig. 4 unpaired UVCGAN vs Others image-to-image translation, Source: [131]

Another application under the umbrella of image translation is the Synthetic Aperture Radar (SAR) image translation [155]. Both the techniques of paired and unpaired GANs are used for this application. Indeed, the SAR-generated images are

not that visually clear but they can be captured at any time so it's a better technique when compared to optical imaging. Particularly, if it is combined with the GANs image translation methods to convert the SAR-generated image into an optical high-resolution clear image [155]. Different GANs methods of unpaired image translation such as CycleGAN [159], NICE GANs [50], Attn-CycleGAN [79] and paired GANs such as Pix2Pix [52], and Bicycle GANs [160] are used for Satellite image translation. In Wei et al. [141] the researchers presented a new technique using the generative AI that is specifically designed for the translation of unpaired SAR images to optical images. In detail, they introduced an approach known as Cross-Fusion Reasoning and Wavelet Decomposition GAN (CFRWD-GAN) [141]. The primary objective of CFRWD-GAN is twofold: to effectively retain structural intricacies and elevate the quality of high-frequency band details. This is achieved through a unique framework that integrates cross-fusion reasoning (CFR) structure, adept at preserving both high-resolution, fine-grained features and low-resolution semantic attributes throughout the entire process of feature reasoning. Additionally, to address speckle noise inherent in SAR images, the method employs discrete wavelet decomposition (WD), enabling the translation of high-frequency components. Through the convergence of these techniques, CFRWD-GAN demonstrates its capability to significantly enhance the translation process for unpaired image-to-image scenarios. The model was evaluated using Root Mean Squared Error (RMSE) [53, 117], structural similarity index (SSIM) [73, 140], peak signal-to-noise ratio (PSNR) [120, 128], learned perceptual image patch similarity (LPIPS) [119, 152] and produced a better result than the other state of art models.

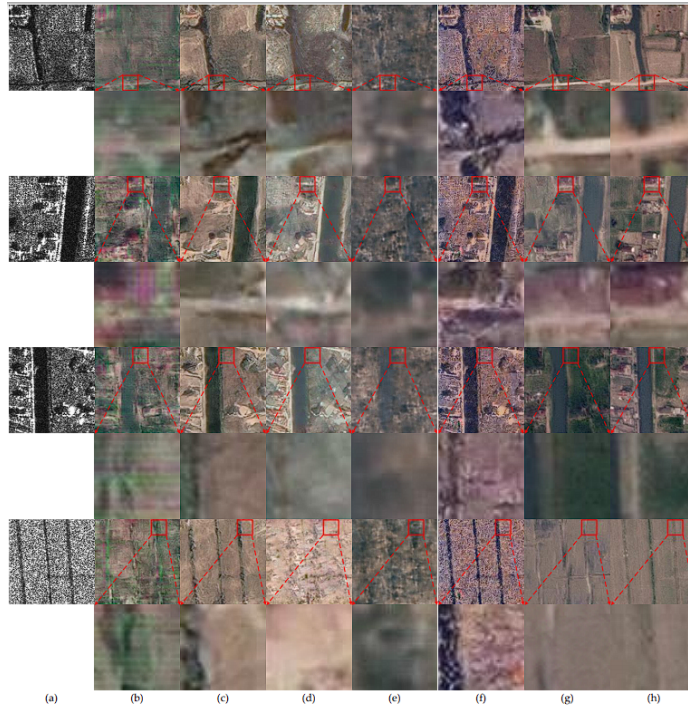


Fig. 5 Highlighted with red boxes and magnified for emphasis, the images are presented in the following order: (a) SAR images, (b) Pix2Pix, (c) CycleGAN, (d) S-cycle-GAN, (e) NICE-GAN, (f) GANILLA, (g) CFRWD-GAN, and (h) ground truth optical images., Source: [141]

The other technique for image generation and translation is Variational Autoencoders developed by Kingma et al. [66]. Furthermore, Zhu et al. [161] compared the generative capabilities of Conditional Variation Autoencoders and also compared it with the other generative techniques for image generation and translation. The main idea of this work was to find the algorithm that performs the best in balancing the diversity and realism in the generated data. The best-performing model in their work was the Bicycle GANs.

Table 2 Generative AI and Its Applications in Image Translation

Domain	Methods	References
Medical-MRI	MM-Transformers, Cyclic-GAN, Pix2Pix GAN, VAE	[1, 5, 17, 28, 30, 144, 146]
Satellite Image Translation	Cyclic-GAN, Pix2Pix GAN, NICE-GAN, Attn-CycleGAN, PSGAN	[80, 101, 155]
Facial Expression Editing	VAE, UPGPT	[22, 39, 149]
Style Transfer	VAE, GANs, DD-GAN	[9, 126, 162]
Text-to-Image Translation	TextControlGAN	[70]
Image Upscaling	GIGA-GAN	[58]

VAE was also used for the molecule generations, generating the 3-Dimensional synthetic molecule structure [151]. Jumper et al. [56] developed an architecture of a generative algorithm specifically for predicting the structure of molecules and proteins. This is the best approach till now in generative AI techniques for generating and predicting molecular architectures.

4.2 Generative AI for Video Synthesis and Generation

Generative AI has transformative applications in the field of video and animation, enabling the creation of visually stunning and dynamic content. The evaluation of video generative models often involves a set of widely recognized datasets: Voxceleb [94], HDTF [154], ClebV [142], Kinetics [59], UCF101 [123], and for specifically audio performance testing VCTK Corpus[132] and LibriSpeech [100]. These datasets are publicly available to researchers, with the stipulation that any use of these resources must include proper citation of the source.

In Hong et al. [49] the researchers introduced a GANs variant called depth-aware GAN. This model provided strong competition to the state of art models and the problem of replacing the face in a video. In detail the dataset on which they tested the built model and compared the results was called Voxceleb [94] and ClebV [142]. The model produced better results than the present state-of-the-art models for achieving talking head video face generation.

Indeed, the replacement of a face in a talking head video is the most predominant application in the video generation problem. Many advancements have been made in generative Artificial intelligence to master this application. Hong et al. [48] state that all this model needs is a target video and a 2-D picture with good pixels and facial features and it will translate the video expression features into the static picture supplied. The *DaGAN++* framework proposed by [48] comprises three key components: (a) an uncertainty-aware face depth learning network that reconstructs detailed 3D facial geometry from self-supervised face videos, without requiring camera parameters or explicit 3D annotations (b) geometry-guided facial keypoint detection, which employs the facial depth network to estimate depth maps, is used alongside RGB images for accurate facial keypoint estimation (c) a geometry-enhanced multi-layer generation process that incorporates learned motion fields, occlusion maps, and facial geometry into each layer of image generation through cross-modal geometry-guided attention. This comprehensive approach enables the synthesis of images enriched with geometry-related attributes derived from facial videos.

It is evident from Figure 6 that *DaGAN++* exhibits superior capabilities. Notably, *DaGAN++* excels in capturing expression-related facial movements within the driving frame, with heightened accuracy observed in regions such as the eyes and mouth. This performance enhancement can be attributed to the precise facial geometry estimation, which greatly contributes to the refinement of expression-related facial motions. Furthermore, Min et al. [90] introduce *StyleTalker*, an innovative audio-driven talking head generative model designed to synthesize a talking person’s video using a single reference image. It features highly accurate lip synchronization, realistic head poses, and natural eye blinks synchronized to the provided audio. To achieve this, they leverage a pre-trained image generator and an image encoder to estimate latent codes for the



Fig. 6 DaGAN++ vs other state of art model on HDTF dataset [154], Source: [48]

talking head video that align faithfully with the given audio input. This achievement stems from the integration of novel components which includes a contrastive lip-sync discriminator. This ensures precise lip synchronization, a conditional sequential variational autoencoder that captures a motion space disentangled from lip movements. In more detail, it enables independent manipulation of motions and lip movements while preserving identity. In addition, it affords an auto-regressive prior enhanced with normalizing flow, facilitating the acquisition of a complex audio-to-motion multi-modal latent space. With these components in place, StyleTalker has the capacity to produce talking head videos, both in a motion-controllable manner when another motion source video is available. Also, entirely driven by audio inputs, wherein it infers real motions from the provided audio.



Fig. 7 StyleTalker vs Other models, Source: [90]

The qualitative evaluation of audio-driven talking head generation performance on VoxCeleb2 dataset in Figure 7 reveals distinct differences. In the first row (marked by a yellow box), frames corresponding to the provided audio are displayed. Conversely, the single image input (highlighted in a red box) represents a reference image of the desired target identity. When observing the generated video frames produced by the StyleTalker in comparison to those generated by other audio-driven generation

models [103, 138, 157], a notable distinction becomes evident. StyleTalker consistently produces talking head videos of exceptional quality, skillfully preserving the distinctive identity of the intended.

Advancing further, Li et al. [76] introduced Multiscale Vision Transformers (MViTv2) as a comprehensive architectural framework suitable for tasks encompassing image and video classification, along with object detection. Within this study, the researchers introduce an enhanced version of MViT, featuring decomposed relative positional embeddings and residual pooling connections. By deploying this upgraded architecture across five different scales, we meticulously assess its performance in scenarios such as ImageNet classification, COCO object detection, and Kinetics video recognition. Remarkably, MViTv2 surpasses previous benchmarks in terms of effectiveness and provided video classification accuracy of 86.1% on the Kinetics-400 dataset as shown in Table 3.

Table 3 Comparative analysis with other Models on the Kinetics-400 [59] dataset, Source: [76]

Model	Top-1	Top-5	FLOPs×views	Param
SlowFast 16×8 +NL [37]	79.8	93.9	234×3×10	59.9
X3D-XL [36]	79.1	93.9	48.4×3×10	11.0
MoViNet-A6 [69]	81.5	95.3	386×1×1	31.4
MViTv1, 16×4 [35]	78.4	93.5	70.3×1×5	36.6
MViTv1, 32×3 [35]	80.2	94.4	170×1×5	36.6
MViTv2-S , 16×4 [76]	81.0	94.6	64×1×5	34.5
MViTv2-B , 32×3 [76]	82.9	95.7	225×1×5	51.2
ViT-B-VTN in 21k [96]	78.6	93.7	4218×1×1	114.0
ViT-B-TimeSformer [12] in 21k	80.7	94.7	2380×3×1	121.4
ViT-L-ViViT [7] in 21k	81.3	94.7	3992×3×4	310.8
Swin-L ⁺ in 21k [83]	84.9	96.7	2107×5×10	200.0
MViTv2-L⁺ , 40×3, in 21k [76]	86.1	97.0	2828×3×5	217.6

In the context of the Kinetics-400 dataset [59], Table 3 presents a comparison between MViTv2 and previous methodologies, encompassing both state-of-the-art Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). Upon training from the ground up, MViTv2-S and MViTv2-B models exhibit top-1 accuracy of 81.0% and 82.9% respectively, surpassing their MViTv1 [35] counterparts by +2.6% and +2.7%. Notably, earlier ViT-based models necessitate substantial pre-training on the IN-21K dataset to achieve peak accuracy on Kinetics-400 as shown in Table 3 (see last 5 rows). In contrast, MViTv2 achieves an exceptional top-1 accuracy of 86.1% when fine-tuning, the MViTv2-L model with a large spatio-temporal input size of 40×3 (time \times space²).

The rest of the reviewed applications of Generative Artificial Intelligence in the field of Video Generation are given in Table 4

Table 4 Generative AI and Its Applications in Video Generation

Domain	Methods	References
Face swapping videos	Depth Awareness GANs, DaGAN++	[48, 49]
Video/Image Classification	MViTv2	[76]
Audio-Based Facial Expression video Translation	Styletalker Sequential VAE	[90]
Simulation in Metaverses	Multi-task DT offloading model	[144]
ECG Synthesis to Improve Deep ECG Classification	SimGAN	[43]
3D human motion prediction	HP-GAN	[11]

4.3 Generative AI for Natural Language Processing

Generative AI models have demonstrated remarkable achievements across a spectrum of Natural Language Processing tasks. These encompass language comprehension, logical reasoning, and text generation.

In the domain of natural language processing, specific datasets have become standard benchmarks for evaluating state-of-the-art models in various tasks. For Named Entity Recognition, the CoNLL-2003 dataset [116] is frequently utilized. In the area of text summarization, two prominent datasets are DUC 2002 [85] and QMSUM [156]. Additionally, for Natural Language Inference (NLI) tasks across multiple languages, the XNLI dataset [24] serves as a crucial resource. All of these datasets are publicly available, providing researchers with essential tools to advance and assess the capabilities of their models in these specific NLP tasks.

Presently, a significant query the AI community poses revolves around the extent and confines of these model’s capabilities [2]. Ahuja et al. [2] raise the question that most of the large language models are made and tested on only the English language. Therefore these researchers took the state-of-the-art models and trained them on other languages on certain available datasets and compared their question-answering and classificational accuracies [2]. Generative AI even has found its way into education[19, 34]. In the midst of this dynamic backdrop that challenges conventional modes of thinking, recent research endeavors investigating the implications of generative AI within the educational landscape yield valuable insights. Notably, these studies shed light on the opportunities and obstacles arising from the integration of generative AI. For instance, in a recent article [130] highlights the need for a new and creative way of teaching that effectively incorporates the progress brought by AI. They mentioned the significance of cultivating an ethical and personalized chatbot solution while augmenting digital proficiency to fully harness the manifold benefits of AI. Furthermore, the researchers advocate for the incorporation of AI literacy as an essential technological skill for navigating the complexities of the 21st century.

In the current landscape, Bozkurt [14] advocates for a significant reevaluation of the roles played by human educators and AI within the educational realm. They assert that the emergence of AI offers a unique juncture to redefine these roles. Especially, as

AI possesses the capacity to assume an increasing array of educational tasks that were traditionally the exclusive domain of human educators. This perspective underscores the importance of adopting a forward-thinking outlook that reconsiders the contributions of both technology and human educators to the educational process. Bozkurt [14] also emphasizes that generative AI’s arrival presents a propitious opportunity to redefine these roles further.

The researchers further delve into the opportunities and challenges ushered in by the advent of generative AI. Generative AI, they elaborate, provides a diverse range of opportunities. These encompass personalized learning, fostering inclusive curriculum provision, and enhancing collaboration and cooperation throughout educational processes. Also, they can cater to automated assessment benefits, ensuring improved accessibility, optimizing efficiency in terms of time and effort, cultivating language skills, and enabling the round-the-clock availability of these technologies.

Generative AI can also enhance synthetic data generation, with the use of Transformers and GANs. It is clear that the understanding of data is better and now the AI is able to generate synthetic data even in the health field. For example, Frid-Adar et al. [40] used GANs to generate synthetic data for the liver lesion classification problem. The variant they used was called DCGAN. In another problem, Wang et al. [137] used the SSIM embedded-cycle GAN to count the number of people in the crowd. The dataset that they used for this problem was fully synthetically generated. It had all the weather conditions covered which made the model perform the best compared to the state-of-the-art models. Another technique of generative AI that is among the state-of-the-art techniques is BERT (Bidirectional Encoder representations from Transformers) which was introduced by the Google AI team [29]. It represents a significant advancement in pre-training techniques for NLP tasks. BERT is based on the transformer architecture and is designed to capture contextual information from both the left and right sides of a word in a sentence, hence the term ‘bidirectional’. In Bert’s pre-training, the model learns to predict missing words in a sentence by training on a large corpus of text. This helps BERT develop a deep understanding of syntax, semantics, and context. After pre-training, the model is fine-tuned on specific downstream tasks, such as sentiment analysis, question answering, and named entity recognition, using task-specific labelled data.

Table 5 shows the BERT model tested with the CoNLL-2003 [116] dataset for the task focusing on named entity recognition. BERTLarge demonstrates strong competitiveness with state-of-the-art techniques. The most successful approach involves concatenating token representations from the uppermost four hidden layers of the pre-trained Transformer. Remarkably, this approach lags by only 0.3 F1 behind the performance achieved by fine-tuning the complete model. This finding underscores the effectiveness of BERT for both fine-tuning and feature-based methodologies.

Another model related to Generative Natural language processing is ELMo [55]. This stands for ‘Embeddings from Language Models’. ELMo utilizes a bidirectional LSTM (Long Short-Term Memory) network [47] for contextual word embeddings. In fact, ELMo embeddings have been shown to be effective in improving the performance of various NLP tasks, including sentiment analysis, question answering, and named entity recognition. The ability to capture context-specific information makes ELMo

Table 5 BERT vs others for Named Entity Recognition task on the CoNLL-2003 [116] dataset (comparison in terms of F1 Score on Validation DataSet (Dev) and Testing Dataset (Test), Source:[29])

System	Dev F1	Test F1
ELMo [55]	95.7	92.2
CVT [23]	-	92.6
CSE [3]	-	93.1
Fine-tuning approach		
BERTLARGE	96.6	92.8
BERTBASE	96.4	92.4
Feature-based approach (BERTBASE)		
Embeddings	91.0	-
Second-to-Last Hidden	95.6	-
Last Hidden	94.9	-
Weighted Sum Last Four Hidden	95.9	-
Concat Last Four Hidden	96.1	-
Weighted Sum All 12 Layers	95.5	-

embeddings particularly useful for tasks where word meanings can vary based on the surrounding context.

Another application of generative AI models used in natural language processing is malware classification. In particular, machine language malware classification is a big concern that can be solved by using generative AI. For example, Kale et al. [57] used the Bert and ELMo to train the embeddings of the models to classify the malware and the results provided remarkable improvements.

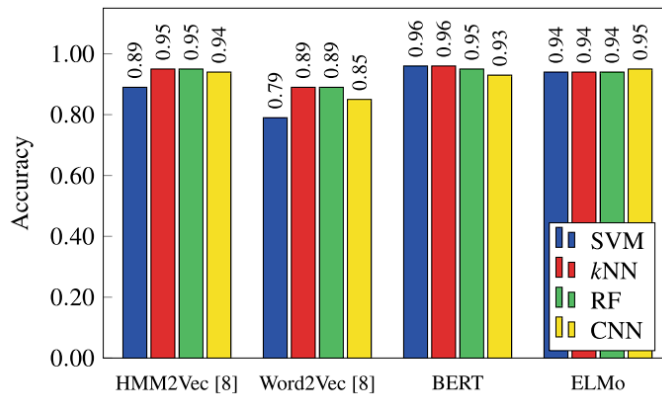


Fig. 8 HMM2Vec, Word2Vec, BERT, and ELMo based classification techniques, Source: [57]

Additionally, Figure 8 provides a summarized depiction of the optimal accuracies achieved by the classification methodologies based on HMM2Vec[20], Word2Vec[89],

BERT[29], and ELMo[102]. The graph illustrates that BERT-SVM and BERT-kNN secured the highest performance with a notable accuracy of 96%. In close pursuit HMM2Vec-kNN, HMM2Vec-RF, BERT-RF, and ELMo-CNN, achieved an accuracy of 95%. Slightly trailing behind, HMM2Vec-CNN, as well as all ELMo-based techniques, demonstrated an accuracy of 94%. Interestingly, the Word2Vec embeddings consistently yielded accuracies below 90% across all four classifiers.

A further application of Generative artificial intelligence is the text-summarization. Joshi et al. [112] introduced *Ranksum* which is an innovative technique designed for extractive text summarization of individual documents. This method hinges on the fusion of four distinct multi-dimensional sentence features, namely topic information, semantic content, significant keywords, and position. By independently acquiring sentence saliency rankings for each feature in an unsupervised manner, Ranksum subsequently amalgamates these scores through weighted fusion, yielding a comprehensive ranking of sentence significance. It is important to note that these scores are generated in a completely unsupervised manner.

Topic ranking is established through the application of probabilistic topic models, while semantic content is captured using sentence embeddings. Sentence embeddings are generated using Siamese networks to craft abstractive sentence representations, followed by a novel strategy to organize them based on their relative importance. To identify significant keywords and their associated sentence rankings within the document, a graph-based approach is employed. Additionally, a mechanism to gauge sentence novelty is formulated, relying on bigrams, trigrams, and sentence embeddings. This eliminates redundant sentences from the summary.

Table 6 Comparative analysis of RankSum with state-of-the-art algorithms conducted on the DUC 2002 [85] and QMSUM [156] dataset, Source: [112, 148]

Method	ROUGE-1	ROUGE-2	ROUGE-L
LEAD	43.6	21.0	40.2
ILP	45.4	21.3	42.8
NN-SE	47.4	23.0	–
SummaRuNNer	47.4	24.0	14.7
Egraph+coh	47.9	23.8	–
Tgraph+coh	48.1	24.3	–
URANK	48.5	21.5	–
SummCoder	51.7	27.5	44.6
HSSAS	52.1	24.5	48.8
CoRank	52.6	25.8	–
Rank-emb	49.9	24.8	45.6
Rank-topic	51.4	25.9	47.2
Rank-keyword	52.0	26.3	48.6
RankSum	53.2	27.9	49.3
PGNet on QMSUM	31.52	8.69	27.63
BART on QMSUM	32.18	8.48	28.56
HMNet on QMSUM	36.06	11.36	31.27
ChatGPT on QMSUM	36.83	12.78	24.23

Table 6 showcases the performance outcomes of the novel RankSum framework in comparison to other state-of-the-art algorithms on the DUC 2002 dataset, evaluated through ROUGE metrics. ROUGE metrics are typically used in the field of machine translation, text summarization, and other tasks where the quality of generated text needs to be evaluated automatically. The metrics involve comparing n-grams (sequences of n words) between the generated and reference texts. Common versions of ROUGE metrics include ROUGE-N (which considers overlapping n-grams), ROUGE-L (which focuses on the longest common subsequence) [78]. Ranksum achieves notable ROUGE-1, ROUGE-2, and ROUGE-L scores of 53.2, 27.9, and 49.3, respectively. Impressively, it outperforms all recent methods examined for this extractive text summarization dataset. Notably, this approach surpasses the highly accurate summarization systems, HSSAS [4] and Co-Rank, with a substantial margin of 0.6, 0.8, and 0.5 for ROUGE-1, ROUGE-2, and ROUGE-L scores respectively [112]. Additionally, the outcomes of PGnet [136], BART[75], HMNet[143], and ChatGPT[15] were assessed using the QMSUM dataset[156]. It is important to note that the current pinnacle of meeting summarization models, HMNet, achieves the most impressive performance in terms of ROUGE-L. This might be attributed to its cross-domain pretraining approach, which imparts HMNet with a heightened familiarity with the style of meeting transcripts [158]. However, it’s worth highlighting that in the case of ROUGE-1 and ROUGE-2 metrics, ChatGPT emerges as the leader. ChatGPT excels due to extensive training on diverse data, enabling superior relationship comprehension in one to two grams, and boosting metric performance.

Model	en	ar	bg	de	el	es	fr	hi	ru	sw	th	tr	ur	vi	zh	avg
Fine-tuned Baselines	80.8	64.3	68.0	70.0	65.3	73.5	73.4	58.9	67.8	49.7	54.1	60.9	57.2	69.3	67.8	65.4
Prompt-Based Baselines	67.5	60.7	46.5	54.0	47.4	61.2	61.4	56.8	53.3	50.4	43.8	42.7	50.0	61.0	56.7	54.2
Open AI Models	76.2	59.0	63.5	67.3	65.1	70.3	67.7	55.5	62.5	56.3	54.0	62.6	49.1	60.9	62.1	62.1
mBERT	80.8	64.3	68.0	70.0	65.3	73.5	73.4	58.9	67.8	49.7	54.1	60.9	57.2	69.3	67.8	65.4
mT5-Base	84.7	73.3	78.6	77.4	77.1	80.3	79.1	70.8	77.1	69.4	73.2	72.8	68.3	74.2	74.1	75.4
XLm-R Large	88.7	77.2	83.0	82.5	80.8	83.7	82.2	75.6	79.1	71.2	77.4	78.0	71.7	79.3	78.2	79.2
TuLRv6 - XXL	93.3	89.0	90.6	90.0	90.2	91.1	90.7	86.2	89.2	85.5	87.5	88.4	82.7	89.0	88.4	88.8
gpt-3.5-turbo	76.2	59.0	63.5	67.3	65.1	70.3	67.7	55.5	62.5	56.3	54.0	62.6	49.1	60.9	62.1	62.1
gpt-3.5-turbo (TT)	76.2	62.7	67.3	69.4	67.2	69.6	69.0	59.9	63.7	55.8	59.6	63.8	54.0	63.9	62.6	64.3
text-davinci-003	79.5	52.2	61.8	65.8	59.7	71.0	65.7	47.6	62.2	50.2	51.1	57.9	50.0	56.4	58.0	59.3
text-davinci-003 (TT)	79.5	65.1	70.8	71.7	69.3	72.2	71.8	63.3	67.3	57.3	62.0	67.6	55.1	66.9	65.8	67.1
gpt-4-32k	84.9	73.1	77.3	78.8	79.0	78.8	79.5	72.0	74.3	70.9	68.8	76.3	68.1	74.3	74.6	75.4

Table 7 Performance comparison among different models on all languages within the XNLI dataset, Source:[2].

Ahuja et al. [2] tested the performance of the state-of-the-art models on multilingual XNLI dataset [24] data. The results are given in Table 7, TuLRv6 - XXL achieves the highest average accuracy across all languages (88.8%). It performs exceptionally well in most languages, with accuracy scores consistently above 85%, XLM-R Large is the model that comes in second place with an average accuracy of 79.2%. While not quite as high as TuLRv6, it still maintains a strong performance across all languages and demonstrates its multilingual capabilities. With an average accuracy of 75.4%, mT5-Base takes the third spot. GPT-4-32k achieves an average accuracy of 75.4%. It exhibits consistent performance across languages, demonstrating its effectiveness in handling multilingual tasks.

Table 8: Generative AI Applications in Natural Language Processing: Major Papers and Descriptions

Paper	Field	Description	Citation
“Attention Is All You Need”	NLP / Machine Translation	Introduces the Transformer model using self-attention mechanisms for various NLP tasks, revolutionizing sequence-to-sequence models	[134]
“BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension”	NLP / Text Generation	Introduces BART, a sequence-to-sequence model pre-trained using denoising autoencoders, capable of various NLP tasks	[75]
“CTRL: A Conditional Transformer Language Model”	Language Generation	Proposes a generative model (CTRL) that can condition its output on specific attributes, enabling fine-grained control over text generation	[61]
“T5: Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer”	NLP / Transfer Learning	Presents T5, a model that casts all NLP tasks as a text-to-text problem, achieving state-of-the-art results across diverse tasks	[111]
“GPT-2: Language Models are Unsupervised Multitask Learners”	Language Generation	Describes the GPT-2 model, a large-scale generative model that demonstrates impressive text generation capabilities across a range of tasks	[110]
“LayoutLM: Pre-training of Text and Layout for Document Image Understanding”	Document Analysis	Presents LayoutLM, a model that pre-trains on document images with associated text, improving document understanding tasks	[145]
“ERNIE: Enhanced Language Representation with Informative Entities”	NLP / Knowledge Enhancement	Introduces ERNIE, a model that enhances language representations by incorporating knowledge from knowledge bases	[153]
“DALL-E: Creating Images from Text”	NLP / Image Generation	Introduces DALL-E, a generative model capable of generating images from textual descriptions	[113]

Continued on next page

Table 8 – continued from previous page

Paper		Field	Description	Citation
“CLIP: Connecting Text and Images for Supervised Learning”	Connecting	NLP / Vision	Proposes CLIP, a model that learns to understand images and text jointly, achieving impressive results in cross-modal tasks	[108]
“WebGPT: Browser-assisted question-answering with human feedback”		NLP, Human feedback	Introduces an approach that leverages a text-based web-browsing environment, enabling the model to access and navigate online resources. The methodology is structured in a manner that aligns with human capabilities, thus facilitating model training through imitation learning.	[95]
“GPT-4 Report”	Technical	NLP/ Text Generation	GPT-4, a groundbreaking advancement, is introduced as a multimodal model with the ability to process both image and text inputs while generating text-based outputs. GPT-4’s accomplishments encompass passing a simulated bar exam with a score that ranks within the top 10% of test takers.	[99]
“Let’s Verify Step by Step”	Step by Step	Mathematical Reasoning	The introduced approach, centered around a process-supervised model, achieves a commendable success rate of 78% when addressing problems sourced from a representative subset of the MATH test set.	[77]

The landscape of Natural Language Processing (NLP) has witnessed remarkable advancements in recent years, driven primarily by the innovative applications of generative AI. The Table 8 highlights a selection of influential papers that showcase the evolution and impact of generative AI techniques within the NLP domain. These advancements have led to groundbreaking developments in various subfields of NLP, transforming the way we process and understand human language.

As evidenced by the papers presented, recent years have seen generative AI techniques reshape NLP in profound ways. These advancements not only enhance the quality and diversity of text generation but also enable more sophisticated control, cross-modal understanding, and knowledge integration. As the field continues to

evolve, it is likely that further innovations in generative AI will continue to drive NLP’s progress, unlocking new frontiers of language understanding and generation across a wide array of applications.

4.4 Generative AI for Knowledge Graph Generation

Researchers and practitioners have leveraged the power of generative AI to enhance the creation and refinement of knowledge graphs—a structured representation of relationships between entities. This section explores the burgeoning landscape of generative AI applications within knowledge graph generation, highlighting pioneering research papers and their contributions to this evolving field. For evaluating models in the field of knowledge graph generation, a variety of datasets are employed, with Dbpedia [74], Cora Dataset [16] and Google’s Knowledge graph [122] being among the most commonly used. However, the scope of available datasets extends beyond just these. These datasets are freely accessible, offering researchers and developers an invaluable resource to test and refine the capabilities of their knowledge graph generation models. A knowledge graph is a structured representation of information that captures relationships between entities and concepts. It goes beyond traditional databases by not only storing data but also organizing it in a way that highlights connections and context. Knowledge graphs, first introduced by Google in 2012 [124], are designed to model real-world relationships, making them a powerful tool for representing and querying complex information. Cai and Wang [18] introduced KBGAN, an innovative adversarial learning framework designed to enhance the performance of various existing knowledge graph embedding models. This approach is not dependent on the specific structures of the generator and discriminator, allowing for the incorporation of a wide range of knowledge graph embedding models as fundamental components. This enables KBGAN to significantly enhance the training dynamics and performance of existing knowledge graph embedding models. Liu et al. [81] introduce K-BERT, a novel approach that empowers language representation with knowledge graphs, enabling the incorporation of commonsense and domain-specific knowledge. The K-BERT methodology comprises two fundamental steps. Initially, knowledge from a knowledge graph (KG) is seamlessly integrated into a sentence, rendering it a knowledge-rich sentence tree. Subsequently, the utilization of soft-position and visible matrix techniques serves to regulate the extent of knowledge integration, thereby preventing any deviation from the original sentence meaning.

Despite the challenges presented by handling heterogeneous entity spans (HES) and keyphrases not seen in training (KN), the investigation yields promising outcomes across a spectrum of twelve open-domain and specific-domain natural language processing (NLP) tasks. Empirical evidence underscores the considerable efficacy of knowledge graphs, particularly in tasks that are driven by domain-specific knowledge. Moreover, K-BERT’s compatibility with the model parameters of BERT offers a seamless integration of knowledge enhancement within a well-established framework.

Link prediction is a fundamental task involving the prediction of missing facts within a knowledge graph using available information. In this context, Balazevic et al. [10] introduce TuckER, a linear model that employs Tucker decomposition of the binary tensor representation of knowledge graph triples. Despite its straightforward

nature, TuckER demonstrates remarkable efficacy. It surpasses previous state-of-the-art models on widely recognized link prediction datasets, solidifying its position as a potent baseline for more sophisticated models in this domain. Moreover, Zeb et al. [121] have introduced ComplexGCN, an innovative graph convolutional network that leverages standard graph convolutional architecture(GCN) [67] to learn complex embeddings. Within the ComplexGCN framework, both node and relation features are projected into complex space through the use of learnable weights associated with neighboring nodes at each convolutional layer. To maintain the integrity of initial embedding information in the final node embeddings, a residual connection between the input and output of the convolutional stack is implemented. These researchers [121]

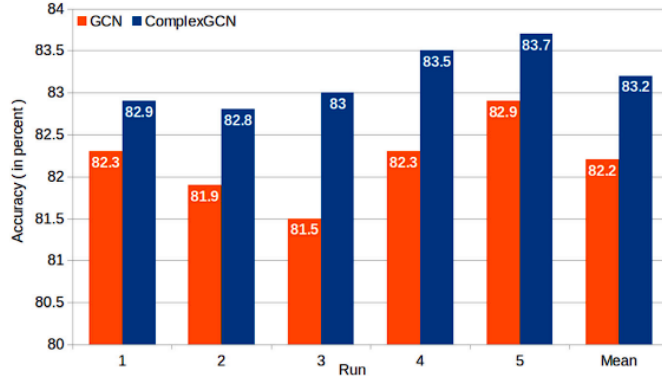


Fig. 9 GCN vs CGCN on node completion task- Cora Dataset, Source: [121]

conducted an evaluation of ComplexGCN’s performance on the node classification task using the Cora dataset. This consisted of 2708 nodes categorized into 7 classes and 5429 edges representing citation links between documents. Both the standard GCN and ComplexGCN were trained on the Cora dataset for 200 epochs, with the objective of minimizing the cross-entropy loss. The training process was repeated 5 times for each model, and the results were averaged and reported in Figure 9. In terms of accuracy percentage, ComplexGCN exhibited improved performance compared to the standard GCN, achieving a 1% increase in mean accuracy.

The presented selection of papers highlights the innovative strides researchers have made in leveraging generative AI to enhance knowledge graph generation. From employing adversarial learning for improved embeddings to bridging the gap between unstructured text and structured knowledge, these papers showcase the multifaceted nature of the advancements.

4.5 Interdisciplinary Applications of Generative AI

The groundbreaking capabilities of generating synthetic data and content generation have given Generative AI the ability to be applicable in interdisciplinary fields. Further recent applications are discussed in this section. The synthetic data generation capability of generative AI is useful in mechanical fault detections. Gao et al. [42]

highlight a fault detection method that combined Finite Element Method (FEM)[26] simulations and Generative Adversarial Network (GAN) [44] to address two primary challenges. Firstly, it aimed to fill the gaps in fault samples by leveraging FEM simulations. Secondly, it sought to enhance fault detection accuracy by utilizing GAN to generate a substantial number of synthetic fault samples. Initially, FEM was employed to generate simulation signals of specific lengths to complete missing fault samples. Subsequently, GAN was utilized to create additional fault samples based on FEM simulations, resulting in a more comprehensive dataset. Finally, classifiers such as Support Vector Machines (SVM)[25], Extreme Learning Machines (ELM)[51], Decision Trees (DTree)[107], and others were employed to detect faults in cases where the faults were previously unknown.

Table 9 Comparison of Classification accuracy with and without GANs sampling, Source: [42]

Description	FEM & AI	FEM, GAN & AI
No of fault Samples	360	360
No of Synthetic Fault Samples	–	3240
Accuracy SVM (%)	84.44	86.11
Accuracy ELM (%)	89.72	91.67
Accuracy Decision Tree (%)	91.11	97.5
Average Accuracy (%)	88.42	91.76

In Table 9 the result clearly identifies the use of Generative AI to generate the synthetic samples which helps the machine learning models to achieve greater accuracy. In Feng et al. [38] a traffic generation model is developed using the Generative AI, named TrafficGen. The model outperforms the previous state-of-art techniques such as SeneGen [127] in generating the synthetic traffic scenarios. The introduced models were also able to generate the trajectory of the generated traffic and synthetic snapshots. This enabled the creation of numerous fresh traffic scenarios and the enhancement of the ones that already exist.

Music generation is also gaining popularity in the field of Generative AI, different methods such as Variational Auto Encoders, Transformers and Recurrent Neural Networks are being used in generating synthetic music, and different new approaches of music generation [60]. Another application is the Handwriting generation [41]. This proposed HiGAN+ which presented the capability to generate a wide range of authentic handwritten texts while being guided by arbitrary textual content and distinct calligraphic styles. These styles are separated from reference images or randomly drawn from a prior normal distribution. Traditional style transfer methods, which rely on pixel-level mappings, may not be suitable for HiGAN+, hence they introduce the contextual loss to notably enhance the stylistic consistency of generated images. The model performed very well in generating readable handwriting samples. In the field of software engineering, generative AI is introduced to help write better code and solve errors in the code, debugging and even writing the documentation of the work (e.g., Copilot) [92].

Emerging technologies in the realm of generative AI aim to simplify human life, yet they also underscore the imperative for responsible AI development. These innovations should prioritize ethical considerations, ensuring that their generated content aligns with societal values. The Contemporary efforts in the generative AI method development are increasingly conscious of these ethical dimensions. In a recent review, Pudari and Ernst [104] delved into Copilot and emphasized that generative AI, while valuable, won't supplant humans in the field of software engineering. This is because it struggles to grasp intricate software design principles and identify coding issues, known as 'code smells'. Instead, its role primarily revolves around aiding developers in crafting more efficient code. Researchers [8] highlight responsible AI as a comprehensive concept, mandating systematic adoption of AI principles. Besides explainability, it emphasizes fairness, accountability, and privacy in real-world AI model implementations, especially in scenarios involving sensitive information and regulatory demands for data privacy.

Apart from these discussed papers, a multitude of groundbreaking advancements are unfolding within the domain of generative AI. It is evident that in this rapidly advancing research, the previously discussed papers represent just a prominent subset of the recent developments in this field.

5 Challenges and opportunities of Generative AI

There are various domains for which we can discuss both the challenges and opportunities of Generative AI. Let's start with challenges and their proposed solutions followed by opportunities-

5.1 Challenges and their proposed solutions

Ethical Concerns:

Challenge: GenAI can be used for malicious purposes, such as the creation of deep-fakes for identity theft or misinformation.

Solution: Establishing ethical governance structures, guidelines, and regulations to guide the responsible development and deployment of GenAI.

Security Concerns:

Challenge: There might be vulnerabilities in generative models that could be exploited for adversarial attacks.

Solution: The development of security measures to protect generative models from manipulation and continuous research into adversarial robustness.

Bias and Fairness:

Challenge: Generative models may amplify and perpetuate biases in the training data, leading to discriminatory and unfair outputs.

Solution: Extensive research and implementation of methods to detect and mitigate bias in training data, as well as encouraging inclusivity and diversity in datasets.

Data Privacy:

Challenge: Generative models trained on large datasets may inadvertently remember sensitive information, posing privacy risks.

Solution: Adherence to data protection regulations and implementation of privacy-preserving approaches to protect personal privacy.

Interpretability:

Challenge: Mostly, it is difficult to understand the decision-making process of generative algorithms due to their ‘black boxes’ nature.

Solution: Research and development of explainable AI approaches to improve interpretability and transparency and, permitting users to understand algorithm outputs.

5.2 Opportunities

Human-AI Collaboration: Collaborative work between GenAI and humans can lead to innovative solutions in design, problem-solving, and creativity.

Creative Expression: Generative AI facilitates innovative creative expressions, such as music, generative art, and literature.

Content Generation: Applications in content creation, for example- image synthesis, text generation, and video creation enhance productivity in numerous industries.

Education and Training: Generative models can be employed for simulating scenarios for training, creating interactive educational materials, and enhancing learning experiences.

Personalization: Generative algorithms can be used for personalized recommendations in entertainment, e-commerce, and other user-oriented domains.

Innovative Design: GenAI can assist in creating optimized and innovative designs in industries like architecture and product design.

Scientific Discovery: GenAI contributes to scientific investigation by simulating complex systems, predicting outcomes, and generating hypotheses.

Healthcare Applications: GenAI advances personalized medicine, drug discovery, medical imaging, and healthcare system.

Balancing the potential benefits of generative AI with the demand for responsible development and deployment is essential. Ethical considerations, transparency, and ongoing research will play major roles in maximizing the positive impact of generative AI while minimizing risks.

6 Conclusion and Future Direction

This paper offers a comprehensive systematic literature review of recent advancements in the field of generative AI. Specifically, it thoroughly explores key algorithms within the realm of Generative AI, including Diffusion Models, Transformer-based models, Generative Adversarial Networks, Variational Autoencoders, and their advancements tailored to specific applications.

Within the paper, we discuss advanced methodologies developed by various researchers, representing the current state-of-the-art achievements in the field of generative AI. A primary focus of generative AI's impact is evident in the domains of NLP and Video Translation, where advanced models have emerged with the capacity to tackle a wide array of human-centric challenges. These include tasks like question answering, code generation, language translation, image transformation, and more interdisciplinary applications. The paper highlights the recent achievements made

in these areas, shedding light on the cutting-edge advancements achieved through generative AI techniques.

Moreover, it seems evident that the future direction of generative AI will be a transformative journey. One critical avenue of exploration involves the continuous evolution of AI architectures, aiming to create models that surpass current machine and human capabilities. Additionally, the ethical dimension of AI is set to gain even more prominence, with research and development focusing on ensuring responsible AI generation, minimizing biases, and aligning with evolving ethical standards. Indeed, interdisciplinary collaborations will flourish, as generative AI is applied to complex challenges in fields like healthcare, climate science, and education, amplifying its real-world impact.

No doubt, the synergy between humans and AI will deepen, emphasizing AI's role as a collaborative partner across various domains. Advancements in NLP will persist, with an emphasis on question-answering, multilingual translation, and code generation. The domain of image, video, and multimedia processing will witness expansion, with generative AI contributing to content creation, enhancement, and interpretation. As we journey into this new and exciting future, it is also clear that we need to remain committed to responsible AI development and ethical considerations in parallel to developing these more advanced generative AI methods.

Statements and Declarations

Funding

This research received partial financial support from the Wales Innovation Network and Global Wales Small Grant Fund, Grant Number: GW-230433/414 (1.3.1b).

Conflict of interest/Competing interests

All authors declare that they have no conflict of interest.

Ethics approval

Not Required.

Data availability statement

This research paper is a systematic review that does not utilize any specific dataset for analysis. Instead, it focuses on collating and evaluating existing literature to provide a comprehensive overview of the field.

References

- [1] Ahmad, B., Sun, J., You, Q., Palade, V., and Mao, Z. (2022). Brain tumor classification using a combination of variational autoencoders and generative adversarial networks. *Biomedicines*, 10(2):223.

- [2] Ahuja, K., Diddee, H., Hada, R., Ochieng, M., Ramesh, K., Jain, P., Nambi, A., Ganu, T., Segal, S., Axmed, M., Bali, K., and Sitaram, S. (2023). Mega: Multilingual evaluation of generative ai.
- [3] Akbik, A., Blythe, D., and Vollgraf, R. (2018). Contextual string embeddings for sequence labeling. In *Proceedings of the 27th international conference on computational linguistics*, pages 1638–1649.
- [4] Al-Sabahi, K., Zuping, Z., and Nadher, M. (2018). A hierarchical structured self-attentive model for extractive document summarization (hssas). *IEEE Access*, 6:24205–24212.
- [5] Ali, H., Biswas, M. R., Mohsen, F., Shah, U., Alamgir, A., Mousa, O., and Shah, Z. (2022). The role of generative adversarial networks in brain mri: a scoping review. *Insights into imaging*, 13(1):98.
- [6] Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein gan.
- [7] Arnab, A., Dehghani, M., Heigold, G., Sun, C., Lučić, M., and Schmid, C. (2021). Vivit: A video vision transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6836–6846.
- [8] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115.
- [9] Atapour-Abarghouei, A. and Breckon, T. P. (2018). Real-time monocular depth estimation using synthetic data with domain adaptation via image style transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [10] Balazevic, I., Allen, C., and Hospedales, T. (2019). Tucker: Tensor factorization for knowledge graph completion. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics.
- [11] Barsoum, E., Kender, J., and Liu, Z. (2018). Hp-gan: Probabilistic 3d human motion prediction via gan. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- [12] Bertasius, G., Wang, H., and Torresani, L. (2021). Is space-time attention all you need for video understanding? In *ICML*, volume 2, page 4.
- [13] Bińkowski, M., Sutherland, D. J., Arbel, M., and Gretton, A. (2018). Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*.

- [14] Bozkurt, A. (2023). Generative artificial intelligence (ai) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, 18(1).
- [15] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Nee-lakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.
- [16] Cabanes, C., Grouazel, A., von Schuckmann, K., Hamon, M., Turpin, V., Coatanoan, C., Guinehut, S., Boone, C., Ferry, N., Reverdin, G., et al. (2012). The cora dataset: validation and diagnostics of ocean temperature and salinity in situ measurements. *Ocean Science Discussions*, 9(2):1273–1312.
- [17] Cabreza, J. N., Solano, G. A., Ojeda, S. A., and Munar, V. (2022). Anomaly detection for alzheimer’s disease in brain mris via unsupervised generative adversarial learning. In *2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, pages 1–5.
- [18] Cai, L. and Wang, W. Y. (2018). Kbgan: Adversarial learning for knowledge graph embeddings.
- [19] Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., and Sun, L. (2023). A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt.
- [20] Chandak, A., Lee, W., and Stamp, M. (2021). A comparison of word2vec, hmm2vec, and pca2vec for malware classification.
- [21] Chen, X., Duan, Y., Houthoofd, R., Schulman, J., Sutskever, I., and Abbeel, P. (2016). Infogan: Interpretable representation learning by information maximizing generative adversarial nets.
- [22] Cheong, S. Y., Mustafa, A., and Gilbert, A. (2023). Upgpt: Universal diffusion model for person image generation, editing and pose transfer.
- [23] Clark, K., Luong, M.-T., Manning, C. D., and Le, Q. V. (2018). Semi-supervised sequence modeling with cross-view training. *arXiv preprint arXiv:1809.08370*.
- [24] Conneau, A., Lample, G., Rinott, R., Williams, A., Bowman, S. R., Schwenk, H., and Stoyanov, V. (2018). Xnli: Evaluating cross-lingual sentence representations. *arXiv preprint arXiv:1809.05053*.
- [25] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20:273–297.
- [26] Courant, R. (1943). Variational methods for the solution of problems of equilibrium and vibrations.

- [27] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65.
- [28] Dar, S. U. H., Yurt, M., Karacan, L., Erdem, A., Erdem, E., and Çukur, T. (2018). Image synthesis in multi-contrast mri with conditional generative adversarial networks.
- [29] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [30] Dimitriadis, A., Trivizakis, E., Papanikolaou, N., Tsiknakis, M., and Marias, K. (2022). Enhancing cancer differentiation with synthetic mri examinations via generative models: a systematic review. *Insights into Imaging*, 13(1):188.
- [31] Dinh, L., Krueger, D., and Bengio, Y. (2015). Nice: Non-linear independent components estimation.
- [32] Dinh, L., Sohl-Dickstein, J., and Bengio, S. (2017). Density estimation using real nvp.
- [33] Donahue, J., Krähenbühl, P., and Darrell, T. (2016). Adversarial feature learning. *arXiv preprint arXiv:1605.09782*.
- [34] Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., Carter, L., Chowdhury, S., Crick, T., Cunningham, S. W., Davies, G. H., Davison, R. M., Dé, R., Dennehy, D., Duan, Y., Dubey, R., Dwivedi, R., Edwards, J. S., Flavián, C., Gauld, R., Grover, V., Hu, M.-C., Janssen, M., Jones, P., Junglas, I., Khorana, S., Kraus, S., Larsen, K. R., Latreille, P., Laumer, S., Malik, F. T., Mardani, A., Mariani, M., Mithas, S., Mogaji, E., Nord, J. H., O’Connor, S., Okumus, F., Pagani, M., Pandey, N., Papagiannidis, S., Pappas, I. O., Pathak, N., Pries-Heje, J., Raman, R., Rana, N. P., Rehm, S.-V., Ribeiro-Navarrete, S., Richter, A., Rowe, F., Sarker, S., Stahl, B. C., Tiwari, M. K., van der Aalst, W., Venkatesh, V., Viglia, G., Wade, M., Walton, P., Wirtz, J., and Wright, R. (2023). Opinion paper: “so what if chatgpt wrote it?” multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy. *International Journal of Information Management*, 71:102642.
- [35] Fan, H., Xiong, B., Mangalam, K., Li, Y., Yan, Z., and Malik, J. (2021). Christoph feichtenhofer. multiscale vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6824–6835.

- [36] Feichtenhofer, C. (2020). X3d: Expanding architectures for efficient video recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 203–213.
- [37] Feichtenhofer, C., Fan, H., Malik, J., and He, K. (2019). Slowfast networks for video recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6202–6211.
- [38] Feng, L., Li, Q., Peng, Z., Tan, S., and Zhou, B. (2023). Trafficgen: Learning to generate diverse and realistic traffic scenarios. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3567–3575.
- [39] Fontanini, T., Ferrari, C., Bertozzi, M., and Prati, A. (2023). Automatic generation of semantic parts for face image synthesis.
- [40] Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. (2018). Synthetic data augmentation using gan for improved liver lesion classification.
- [41] Gan, J., Wang, W., Leng, J., and Gao, X. (2022). Higan+: Handwriting imitation gan with disentangled representations. *ACM Trans. Graph.*, 42(1).
- [42] Gao, Y., Liu, X., and Xiang, J. (2020). Fem simulation-based generative adversarial networks to detect bearing faults. *IEEE Transactions on Industrial Informatics*, 16(7):4961–4971.
- [43] Golany, T., Radinsky, K., and Freedman, D. (2020). SimGANs: Simulator-based generative adversarial networks for ECG synthesis to improve deep ECG classification. In III, H. D. and Singh, A., editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3597–3606. PMLR.
- [44] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial networks.
- [45] Grathwohl, W., Chen, R. T., Bettencourt, J., Sutskever, I., and Duvenaud, D. (2018). Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*.
- [46] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- [47] Hochreiter, S. and Schmidhuber, J. (1996). Lstm can solve hard long time lag problems. *Advances in neural information processing systems*, 9.

- [48] Hong, F.-T., Shen, L., and Xu, D. (2023). Dagan++: Depth-aware generative adversarial network for talking head video generation.
- [49] Hong, F.-T., Zhang, L., Shen, L., and Xu, D. (2022). Depth-aware generative adversarial network for talking head video generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3397–3406.
- [50] Hoyez, H., Schockaert, C., Rambach, J., Mirbach, B., and Stricker, D. (2022). Unsupervised image-to-image translation: A review. *Sensors*, 22(21).
- [51] Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2004). Extreme learning machine: a new learning scheme of feedforward neural networks. In *2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541)*, volume 2, pages 985–990. Ieee.
- [52] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2018). Image-to-image translation with conditional adversarial networks.
- [53] Jain, V., Sengar, S. S., and Ronickom, J. F. A. (2023). Age-specific diagnostic classification of asd using deep learning approaches. *Studies in Health Technology and Informatics*, 309:267–271.
- [54] Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L.-w. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., and Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1):1–9.
- [55] Joshi, V., Peters, M., and Hopkins, M. (2018). Extending a parser to distant domains using a few dozen partially annotated examples.
- [56] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589.
- [57] Kale, A. S., Pandya, V., Di Troia, F., and Stamp, M. (2023). Malware classification with word2vec, hmm2vec, bert, and elmo. *Journal of Computer Virology and Hacking Techniques*, 19(1):1–16.
- [58] Kang, M., Zhu, J.-Y., Zhang, R., Park, J., Shechtman, E., Paris, S., and Park, T. (2023). Scaling up gans for text-to-image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10124–10134.
- [59] Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., Viola, F., Green, T., Back, T., Natsev, P., Suleyman, M., and Zisserman, A. (2017). The kinetics human action video dataset.

- [60] Keerti, G., Vaishnavi, A., Mukherjee, P., Vidya, A. S., Sreenithya, G. S., and Nayab, D. (2022). Attentional networks for music generation. *Multimedia Tools and Applications*, 81(4):5179–5189.
- [61] Keskar, N. S., McCann, B., Varshney, L. R., Xiong, C., and Socher, R. (2019). Ctrl: A conditional transformer language model for controllable generation.
- [62] Khamparia, A., Gupta, D., Rodrigues, J. J., and de Albuquerque, V. H. C. (2021). Dcavn: Cervical cancer prediction and classification using deep convolutional and variational autoencoder network. *Multimedia Tools and Applications*, 80:30399–30415.
- [63] Kingma, D. P. and Dhariwal, P. (2018). Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems*, 31.
- [64] Kingma, D. P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I., and Welling, M. (2016). Improved variational inference with inverse autoregressive flow. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.
- [65] Kingma, D. P. and Welling, M. (2013a). Auto-encoding variational bayes.
- [66] Kingma, D. P. and Welling, M. (2013b). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.
- [67] Kipf, T. N. and Welling, M. (2017). Semi-supervised classification with graph convolutional networks.
- [68] Kollem, S., Reddy, K. R., and Rao, D. S. (2023). A novel diffusivity function-based image denoising for mri medical images. *Multimedia Tools and Applications*, 82(21):32057–32089.
- [69] Kondratyuk, D., Yuan, L., Li, Y., Zhang, L., Tan, M., Brown, M., and Gong, B. (2021). Movinets: Mobile video networks for efficient video recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16020–16030.
- [70] Ku, H. and Lee, M. (2023). Textcontrolgan: Text-to-image synthesis with controllable generative adversarial networks. *Applied Sciences*, 13(8):5098.
- [71] Kumar, L. and Singh, D. K. (2023). A comprehensive survey on generative adversarial networks used for synthesizing multimedia content. *Multimedia Tools and Applications*, 82(26):40585–40624.
- [72] Kumar, S., Mallik, A., and Sengar, S. S. (2023). Community detection in complex networks using stacked autoencoders and crow search algorithm. *The Journal of Supercomputing*, 79(3):3329–3356.

- [73] Lakshmi, P. B., Reddy, V. D., Ghosh, S., and Sengar, S. S. (2023). Classification of autism spectrum disorder based on brain image data using deep neural networks. In *International Conference on Frontiers of Intelligent Computing: Theory and Applications*, pages 209–218. Springer.
- [74] Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S., et al. (2015). Dbpedia—a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic web*, 6(2):167–195.
- [75] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., and Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension.
- [76] Li, Y., Wu, C.-Y., Fan, H., Mangalam, K., Xiong, B., Malik, J., and Feichtenhofer, C. (2022). Mvitv2: Improved multiscale vision transformers for classification and detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4804–4814.
- [77] Lightman, H., Kosaraju, V., Burda, Y., Edwards, H., Baker, B., Lee, T., Leike, J., Schulman, J., Sutskever, I., and Cobbe, K. (2023). Let’s verify step by step.
- [78] Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.
- [79] Lin, Y., Wang, Y., Li, Y., Gao, Y., Wang, Z., and Khan, L. (2021). Attention-based spatial guidance for image-to-image translation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 816–825.
- [80] Liu, Q., Zhou, H., Xu, Q., Liu, X., and Wang, Y. (2020). Psgan: A generative adversarial network for remote sensing image pan-sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12):10227–10242.
- [81] Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., and Wang, P. (2019). K-bert: Enabling language representation with knowledge graph.
- [82] Liu, Z., Luo, P., Wang, X., and Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*.
- [83] Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., and Hu, H. (2022). Video swin transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3202–3211.
- [84] Masci, J., Meier, U., Cireşan, D., and Schmidhuber, J. (2011). Stacked convolutional auto-encoders for hierarchical feature extraction. In *Artificial Neural*

Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14–17, 2011, Proceedings, Part I 21, pages 52–59. Springer.

- [85] McKeown, K., Barzilay, R., Blair-Goldensohn, S., Evans, D., Hatzivassiloglou, V., Klavans, J., Nenkova, A., Schiffman, B., and Sigelman, S. (2002). The columbia multi-document summarizer for duc 2002. In *Workshop on Automatic Summarization*, pages 1–8.
- [86] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al. (2014). The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024.
- [87] Mescheder, L., Geiger, A., and Nowozin, S. (2018a). Which training methods for gans do actually converge?
- [88] Mescheder, L., Nowozin, S., and Geiger, A. (2018b). The numerics of gans.
- [89] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- [90] Min, D., Song, M., and Hwang, S. J. (2022). Styletalker: One-shot style-based audio-driven talking head video generation.
- [91] Mirza, M. and Osindero, S. (2014). Conditional generative adversarial nets.
- [92] Moradi Dakhel, A., Majdinasab, V., Nikanjam, A., Khomh, F., Desmarais, M. C., and Jiang, Z. M. J. (2023). Github copilot ai pair programmer: Asset or liability? *Journal of Systems and Software*, 203:111734.
- [93] Nagarajan, V. and Kolter, J. Z. (2017). Gradient descent gan optimization is locally stable. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- [94] Nagrani, A., Chung, J. S., and Zisserman, A. (2017). VoxCeleb: A large-scale speaker identification dataset. In *Interspeech 2017*. ISCA.
- [95] Nakano, R., Hilton, J., Balaji, S., Wu, J., Ouyang, L., Kim, C., Hesse, C., Jain, S., Kosaraju, V., Saunders, W., Jiang, X., Cobbe, K., Eloundou, T., Krueger, G., Button, K., Knight, M., Chess, B., and Schulman, J. (2022). Webgpt: Browser-assisted question-answering with human feedback.
- [96] Neimark, D., Bar, O., Zohar, M., and Asselmann, D. (2021). Video transformer network.
- [97] Odena, A. (2016). Semi-supervised learning with generative adversarial networks.

- [98] Odena, A., Olah, C., and Shlens, J. (2017). Conditional image synthesis with auxiliary classifier gans.
- [99] OpenAI (2023). Gpt-4 technical report.
- [100] Panayotov, V., Chen, G., Povey, D., and Khudanpur, S. (2015). Librispeech: An asr corpus based on public domain audio books. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5206–5210.
- [101] Paola, Z. L., Jesús, L. S., Christian, A. H., and Sonia, R. U. (2023). Correction of banding errors in satellite images with generative adversarial networks (gan). *IEEE Access*.
- [102] Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations.
- [103] Prajwal, K., Mukhopadhyay, R., Namboodiri, V. P., and Jawahar, C. (2020). A lip sync expert is all you need for speech to lip generation in the wild. In *Proceedings of the 28th ACM international conference on multimedia*, pages 484–492.
- [104] Pudari, R. and Ernst, N. A. (2023). From copilot to pilot: Towards ai supported software development.
- [105] Qi, G.-J. (2018). Loss-sensitive generative adversarial networks on lipschitz densities.
- [106] Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., and Huang, X. (2020). Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 63(10):1872–1897.
- [107] Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1:81–106.
- [108] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. (2021). Learning transferable visual models from natural language supervision.
- [109] Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al. (2018). Improving language understanding by generative pre-training.
- [110] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.
- [111] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551.

- [112] Rani, R. and Lobiyal, D. (2021). An extractive text summarization approach using tagged-lda based topic modeling. *Multimedia tools and applications*, 80:3275–3305.
- [113] Reddy, M. D. M., Basha, M. S. M., Hari, M. M. C., and Penchalaiah, M. N. (2021). Dall-e: Creating images from text. *UGC Care Group I Journal*, 8(14):71–75.
- [114] Rezagholiradeh, M. and Haidar, M. A. (2018). Reg-gan: Semi-supervised learning based on generative adversarial networks for regression. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 2806–2810. IEEE.
- [115] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., and Chen, X. (2016). Improved techniques for training gans. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.
- [116] Sang, E. F. and De Meulder, F. (2003). Introduction to the conll-2003 shared task: Language-independent named entity recognition. *arXiv preprint cs/0306050*.
- [117] Sengar, S. S. and Kumar, S. (2022). Content-based secure image retrieval in an untrusted third-party environment. In *International Conference on Frontiers of Intelligent Computing: Theory and Applications*, pages 287–297. Springer.
- [118] Sengar, S. S., Meulengracht, C., Boesen, M. P., Overgaard, A. F., Gudbergensen, H., Nybing, J. D., Perslev, M., and Dam, E. B. (2023). Multi-planar 3d knee mri segmentation via unet inspired architectures. *International Journal of Imaging Systems and Technology*, 33(3):985–998.
- [119] Sengar, S. S. and Mukhopadhyay, S. (2016). Moving object tracking using laplacian-dct based perceptual hash. In *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pages 2345–2349. IEEE.
- [120] Sengar, S. S. and Mukhopadhyay, S. (2020). Motion segmentation-based surveillance video compression using adaptive particle swarm optimization. *Neural Computing and Applications*, 32(15):11443–11457.
- [121] Shi, X., Lv, F., Seng, D., Zhang, J., Chen, J., and Xing, B. (2021). Visualizing and understanding graph convolutional network. *Multimedia Tools and Applications*, 80:8355–8375.
- [122] Singhal, A. (2012). Introducing the knowledge graph: Things, not strings,.
- [123] Soomro, K., Zamir, A. R., and Shah, M. (2012). Ucf101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*.

- [124] Steiner, T., Verborgh, R., Troncy, R., Gabarro, J., and Van de Walle, R. (2012). Adding realtime coverage to the google knowledge graph. In *11th International Semantic Web Conference (ISWC 2012)*, volume 914, pages 65–68. Citeseer.
- [125] Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.
- [126] Tahir, R., Cheng, K., Memon, B. A., and Liu, Q. (2022). A diverse domain generative adversarial network for style transfer on face photographs.
- [127] Tan, S., Wong, K., Wang, S., Manivasagam, S., Ren, M., and Urtasun, R. (2021). Scenegen: Learning to generate realistic traffic scenes. In *Proceedings - 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021*, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 892–901. IEEE Computer Society. Funding Information: Work done at Uber ATG. Publisher Copyright: © 2021 IEEE; 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021 ; Conference date: 19-06-2021 Through 25-06-2021.
- [128] Tanchenko, A. (2014). Visual-psnr measure of image quality. *Journal of Visual Communication and Image Representation*, 25(5):874–878.
- [129] Tibrewala, R., Dutt, T., Tong, A., Ginocchio, L., Keerthivasan, M. B., Baete, S. H., Chopra, S., Lui, Y. W., Sodickson, D. K., Chandarana, H., and Johnson, P. M. (2023). Fastmri prostate: A publicly available, biparametric mri dataset to advance machine learning for prostate cancer imaging.
- [130] Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., and Agyemang, B. (2023). What if the devil is my guardian angel: Chatgpt as a case study of using chatbots in education. *Smart Learning Environments*, 10(1):15.
- [131] Torbunov, D., Huang, Y., Yu, H., Huang, J., Yoo, S., Lin, M., Viren, B., and Ren, Y. (2023). Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 702–712.
- [132] van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio.
- [133] Vasanthi, P. and Mohan, L. (2023). Multi-head-self-attention based yolov5x-transformer for multi-scale object detection. *Multimedia Tools and Applications*, pages 1–27.

- [134] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- [135] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., and Bottou, L. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(12).
- [136] Wang, P., Zhang, C., Qi, F., Liu, S., Zhang, X., Lyu, P., Han, J., Liu, J., Ding, E., and Shi, G. (2021a). Pgnnet: Real-time arbitrarily-shaped text spotting with point gathering network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 2782–2790.
- [137] Wang, Q., Gao, J., Lin, W., and Yuan, Y. (2019). Learning from synthetic data for crowd counting in the wild.
- [138] Wang, S., Li, L., Ding, Y., Fan, C., and Yu, X. (2021b). Audio2head: Audio-driven one-shot talking-head generation with natural head motion. *arXiv preprint arXiv:2107.09293*.
- [139] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., and Summers, R. M. (2017). Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106.
- [140] Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.
- [141] Wei, J., Zou, H., Sun, L., Cao, X., He, S., Liu, S., and Zhang, Y. (2023). Cfrwd-gan for sar-to-optical image translation. *Remote Sensing*, 15(10):2547.
- [142] Wu, W., Zhang, Y., Li, C., Qian, C., and Loy, C. C. (2018). Reenactgan: Learning to reenact faces via boundary transfer.
- [143] Xiao, S., Duan, L., Xie, G., Li, R., Chen, Z., Deng, G., and Nummenmaa, J. (2021). Hmnet: Hybrid matching network for few-shot link prediction. In *International Conference on Database Systems for Advanced Applications*, pages 307–322. Springer.
- [144] Xu, I. R., Van Booven, D. J., Goberdhan, S., Breto, A., Porto, J., Alhuseini, M., Algohary, A., Stoyanova, R., Punnen, S., Mahne, A., et al. (2023). Generative adversarial networks can create high quality artificial prostate cancer magnetic

- resonance images. *Journal of Personalized Medicine*, 13(3):547.
- [145] Xu, Y., Li, M., Cui, L., Huang, S., Wei, F., and Zhou, M. (2020). Layoutlm: Pre-training of text and layout for document image understanding. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1192–1200.
- [146] Yan, S., Wang, C., Chen, W., and Lyu, J. (2022). Swin transformer-based gan for multi-modal medical image translation. *Frontiers in Oncology*, 12:942511.
- [147] Yang, K., Yau, J., Fei-Fei, L., Deng, J., and Russakovsky, O. (2022). A study of face obfuscation in imagenet. In *International Conference on Machine Learning (ICML)*.
- [148] Yang, X., Li, Y., Zhang, X., Chen, H., and Cheng, W. (2023). Exploring the limits of chatgpt for query or aspect-based text summarization.
- [149] Yeh, R., Liu, Z., Goldman, D. B., and Agarwala, A. (2016). Semantic facial expression editing using autoencoded flow.
- [150] Yu, L., Zhang, W., Wang, J., and Yu, Y. (2017). Seqgan: Sequence generative adversarial nets with policy gradient.
- [151] Zeng, X., Wang, F., Luo, Y., Kang, S.-g., Tang, J., Lightstone, F. C., Fang, E. F., Cornell, W., Nussinov, R., and Cheng, F. (2022). Deep generative molecular design reshapes drug discovery. *Cell Reports Medicine*.
- [152] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595.
- [153] Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., and Liu, Q. (2019). Ernie: Enhanced language representation with informative entities. *arXiv preprint arXiv:1905.07129*.
- [154] Zhang, Z., Li, L., Ding, Y., and Fan, C. (2021). Flow-guided one-shot talking face generation with a high-resolution audio-visual dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3661–3670.
- [155] Zhao, Y., Celik, T., Liu, N., and Li, H.-C. (2022). A comparative analysis of gan-based methods for sar-to-optical image translation. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5.
- [156] Zhong, M., Yin, D., Yu, T., Zaidi, A., Mutuma, M., Jha, R., Awadallah, A. H., Celikyilmaz, A., Liu, Y., Qiu, X., and Radev, D. (2021). QMSum: A new benchmark for query-based multi-domain meeting summarization. In *Proceedings of the 2021*

Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 5905–5921, Online. Association for Computational Linguistics.

- [157] Zhou, Y., Han, X., Shechtman, E., Echevarria, J., Kalogerakis, E., and Li, D. (2020). Makelttalk: speaker-aware talking-head animation. *ACM Transactions On Graphics (TOG)*, 39(6):1–15.
- [158] Zhu, C., Xu, R., Zeng, M., and Huang, X. (2020). A hierarchical network for abstractive meeting summarization with cross-domain pretraining. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 194–203, Online. Association for Computational Linguistics.
- [159] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017a). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251.
- [160] Zhu, J.-Y., Zhang, R., Pathak, D., Darrell, T., Efros, A. A., Wang, O., and Shechtman, E. (2017b). Toward multimodal image-to-image translation. *Advances in neural information processing systems*, 30.
- [161] Zhu, J.-Y., Zhang, R., Pathak, D., Darrell, T., Efros, A. A., Wang, O., and Shechtman, E. (2017c). Toward multimodal image-to-image translation. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- [162] Zuo, Z., Zhao, L., Lian, S., Chen, H., Wang, Z., Li, A., Xing, W., and Lu, D. (2022). Style fader generative adversarial networks for style degree controllable artistic style transfer. In *Proc. Int. Joint Conf. on Artif. Intell. (IJCAI)*, pages 5002–5009.