Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 24 May 2026
Revised 28 June 2026
Accepted 16 July 2026

Natural Resources for Human Health 2026, Vol. 6 (7s): 434–447

KEYWORDS: Generative Models, Personalized Content Generation, Generative adversarial Networks (GANs), Variational Autoencoders (VAEs), and large language models (LLMs), Finance sector

Generative Models AI Models for Personalized Content Generation in Financial Industry

Dr.M. Sakthivanitha[1], Dr.B. Mahalakshmi[2], M. VedhaPriya[3], Dr. S. Silvia Priscila[4], Dr.Shiammala P N[5], Dr.R. Bagavathi Lakashmi[6]

[1]Associate Professor, Department of Computer Science, Hindusthan College of Arts & Science, Coimbatore sakthivanithamsc@gmail.com
[2]Assistant Professor,Department of Computer Science, Faculty of Liberal Arts and Business Studies, SRM Institute of Science and Technology, Kattankulathur. Mahalakb1@srmist.edu.in
[3]Assistant Professor, Department of Computer Applications, Vels Institute of Science Technology and Advanced Studies (VISTAS), Pallavaram,Chennai.
[4] Department of Computer Science,Bharath Institute of Higher Education and Research, Bharath University, Chennai,Tamil Nadu, India.
[5]Assistant Professor, Department of Computer Applications, Vels Institute of Science Technology and Advanced Studies (VISTAS), Pallavaram,Chennai. shiammala@gmail.com
[6]Associate Professor, Department of Computer Applications, Vels Institute of Science Technology and Advanced Studies (VISTAS), Pallavaram, Chennai.

ABSTRACT: This study provides a comprehensive review of state-of-the-art generative artificial intelligence (AI) models and their applications in the personalized generation of content, especially in the financial field. The study focuses on some key generative methods such as Generative Adversarial Networks (GANs), Transformer-based models, Diffusion models, Variational Autoencoders (VAEs), and their architecture, operation, advantages and limitations. A comparative analysis is given to help or understand their suitability in different use cases in terms of quality, complexity, diversity, and computational requirements. Furthermore, standard evaluation metrics for text, image, and audio generation are discussed in the paper, e.g. Perplexity, BLEU, ROUGE, METEOR, Frechet Inception Distance (FID), and Inception Score (IS), as well as personalization-specific measures. Key challenges such as data privacy, ethical bias, scalability, and computational complexity are critically analyzed in order to gain a range various practical constraints of implementing generative AI systems. The review concludes by stating that while GANs and Diffusion models are good at creating high-quality content, Transformers are great for tasks that involve sequential and text-based information while VAEs are stable and interpretable representations. While there are currently limitations, generative AI has a considerable opportunity to make finance personalized services, automation, and decision-making much more effective. Future developments in model efficiency, robustness and ethical practices in AI are expected to take it even further, in terms of its real world applicability.

1 INTRODUCTION

Financial services marketers are experiencing new challenges. They must not only keep up with emerging marketing practices in the financial services business, but also with modern finance and marketing technologies in order to meet customer demands and increase their market position. Generative AI in financial services is expected to increase worker productivity by 0.1% to

0.6% each year till 2040. This breakthrough technology is estimated to generate an additional $200 billion to $340 billion to the industry. Given the present emphasis on lowering expenses in bank strategic planning, using generative AI in the financial sector to boost efficiency is becoming increasingly important (Julia & Jersy, 2024). In a constantly transforming digital world, the introduction of generative AI has brought in a new era of personalised client relationships. As organizations attempt to fulfill unique client demands across many platforms and touchpoints, generative AI has demonstrated to be an effective tool for enabling customisation on a much bigger scale.

1.1 Overview of Generative Models in Finance

AI Modeling in finance is not an easy task as the data often contains complex statistical features and the inner workings are not well understood. DL algorithms have started to break ground in the domain of data-driven modelling, but a lack of rich enough information to train algorithms is currently limiting various potential applications. Recently, Generative artificial intelligence (Generative AI) has become a success factor in many industries, including the financial industry. Identification of how the industry views Generative AI is important to its effective integration(Zacharias et al., 2024).

GANs are a class of neural network architectures that have gained significant success in image synthesis and are increasingly being used for the generation of financial time series data (Eckerli et al., 2021). Through an adversarial training process of continually updating models, GANs are capable of generating highly diverse synthetic data preservation of the underlying statistical characteristics of a real financial model. This ability makes them useful in downstream applications such as credit scoring, anti-money laundering (AML) and financial risk assessment (Selvaraj et al., 2022).

Generative AI, in general, learned in complex ways patterns and structures from extensive data sets and creates new content, both text and images, audio and code. Unlike the analytical AI, which is more focused on prediction and classification, generative AI aims at content creation, reasoning, and understanding the context for broader applicability in different domains (Qian et al., 2024). The evolution of generative AI in the field of artificial intelligence can be illustrated in Figure 1.

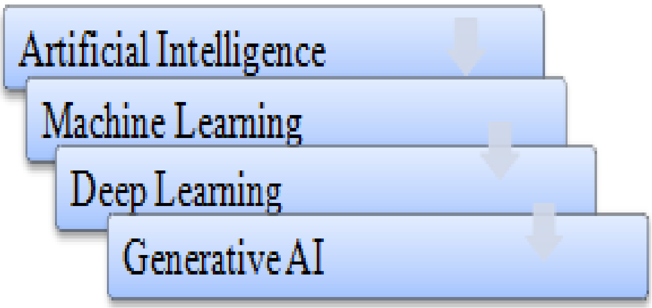

Figure 1. AI to Generative AI

1.2 Personalization in Financial Content Generation

Personalization in financial services is a technique in which an offer of a product or service is made on the basis of customer's specific needs and historical data. It has the potential of increasing income and performance apart from gaining trust. Generative AI makes use of good machine learning algorithms to understand large datasets and detect patterns, trends and preferences. This feature makes it possible to do new things with the purpose to represent ingenuity of its author and individualization according to specific user profiles. Generative AI may be deployed to dynamically tailor content to be more relevant and interesting, whether that is through personalised email communications or through the suggestion of items based on what the user's habits have been (Patil et al., 2024).

Generative AI is the use of ML, neural networks, and other techniques to create creative content (such as text, pictures, music) based on patterns and data from the data of training. Generative AI currently has a huge diversity of applications from personalised content to more effective business processes (Ooi et al., 2023). Generative AI can be used to help with hyper-personalization of information, by combining the potential customer's surfing history and previous purchases with other accessible digital footprints. These adaptive real-time offers have the potential to boost promotion offer conversion at the cost of click-through ratios. For instance, consumers are provided with personalized adverts and offers based on the perspective, like, remark and other interaction factors in real-time.

This section provides the fundamentals of generative AI models and the recent studies in the literature for the application of Generative AI in financial industry. Various studies were found leveraging the large language, transformer and diffusion models. This section also provides the key concepts of reinforcement learning(RL) with human feedback and the synergies between Generative Models and RL and ends up with the applications of Generative AI models in finance sector.

2. FUNDAMENTALS OF GENERATIVE AI MODELS

Generative AI encompasses unsupervised and semi-supervised machine learning methods that allow machines to generate fresh information from existing content such as text, audio and video files, pictures, and code. The fundamental goal is to create fully unique artifacts that seem like actual thing. The figure 2 depicts the supervised and unsupervised learning algorithms in Generative AI models

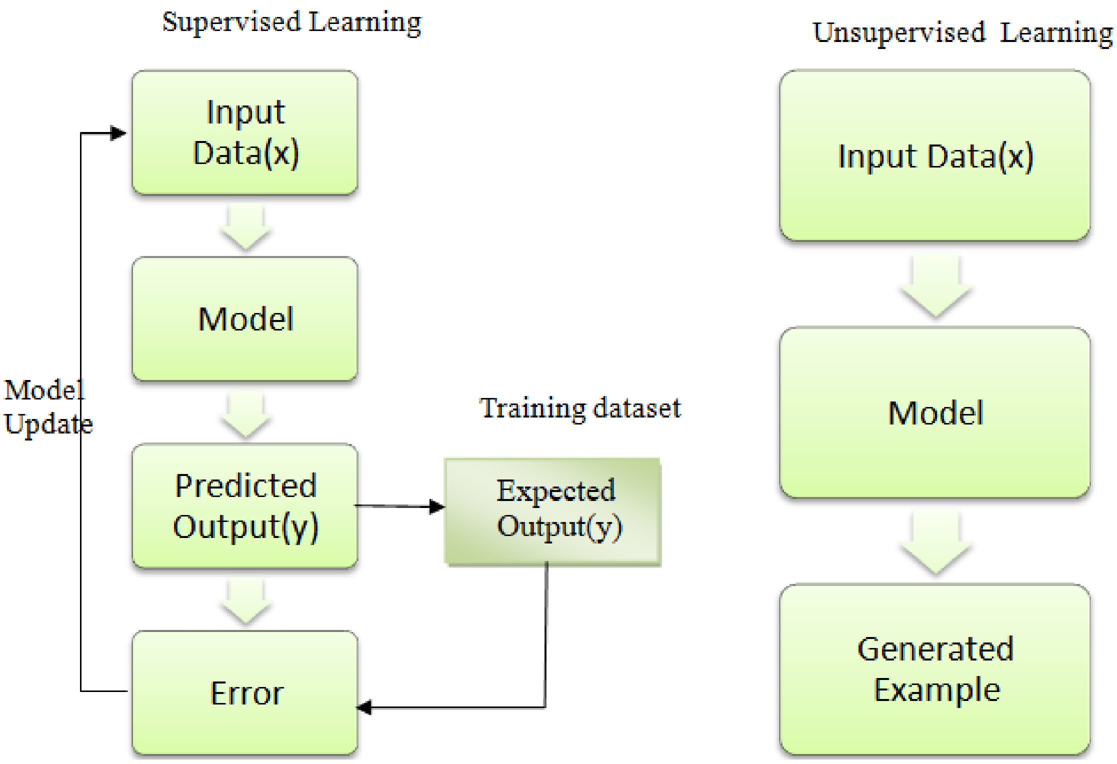

Figure 2 . Supervised and unsupervised algorithms in Generative AI Models

. Generative AI, as represented by models such as ChatGPT, DALL-E, and Midjourney, has grown fast, owing to advances in DL techniques and the availability of large datasets. LLMs have also transformed NLP tasks, using large text samples to synthesize human-like prose. Linkon et al.(2024) investigates current advances, including the introduction of sophisticated models like GPT-4 and PaLM2 and the rise of specialized LLMs such small LLMs (sLLMs), which try to overcome hardware limits and cost considerations. Generative AI refers to all AI technologies used to produce content, including LLMs processing and generating textual data and LMMs processing and generating various data modalities, also known as multimodal AI models. AI chatbots and multimodal AI models provide several benefits, such as rapidity, preciseness, imaginative thinking, productivity, and adaptability; thus, they can be used in nearly every sector, such as a writer, diligent aid, trustworthy companions client assistance, and commercial operations coordinator(Naik et al., 2024).

The diffusion model (DM) is a probabilistic generative model that operates in two steps. First, the forward diffusion process inserts Gaussian noise into the training data. The reverse diffusion process, referred to as denoising, progressively reverses the diffusion step by step to obtain fresh sample data.(Yang et al.,2023). These frameworks effectively tackle the difficulties faced in aligning the posterior distributions within VAEs, reduce the inherent unpredictability in the adversarial targets of GANs by offering an improved training objective, and address the technical burdens posed by Markov Chain methods(Cao et al., 2024). Figure 3 presents the various Generative AI models

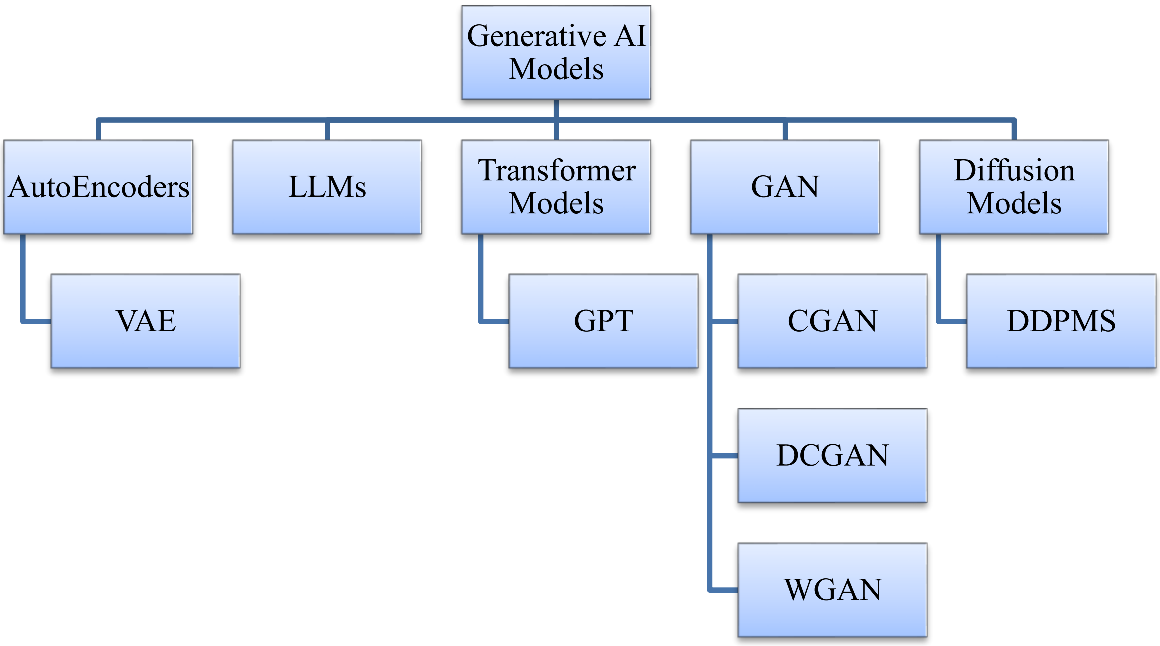

Figure 3 Various Generative AI models

The launching of ChatGPT in 2022 introduced generative AI to the attention of everyone in the world, which is considered a big turning point in the development of GAI. While the project GAI had been evolving since several years, the fact is that this milestone enhanced the research interest and technological innovation about the AI landscape. As a result, many advanced tools have emerged including Bard, Stable Diffusion, DALL*E, Make-A-Video, Runway ML, and Jukebox.

These systems show an astonishing versatility, and can be used for a large variety of applications such as text generation, music creation, image synthesis, video generation, code generation, and even scientific exploration. Their capabilities are driven by advanced underlying models include transformer-architecture (e.g., GPT-3 and GPT-4), diffusion-model, variational autoencoders and generative adversarial network (Bengesi et al., 2024).

LLMs have shown incredible abilities in a wide range of NLP applications, attracting interest from a number of areas, including financial services. Although thorough research has been made on such a general-domain LLM and its massive potential in the finance domain, the relations of research on Financial LLM (FinLLM) is limited at this point. Lee et al.(2024) gives an in-depth description of the FinLLMs including background, approach, effectiveness, opportunities and challenges.

Gemini is a subdivision of Google AI, and hence can access massive databases and computing resources, which would provide it a competitive edge when it comes to intricate financial analysis. This advantage however, adds to the cost of computing. On the contrary, ChatGPT by OpenAI promotes the speed and efficiency of resources, and should be applied in the application where text processing is required quickly. Rane et al. (2024) discusses variations in the features of models, architectural models and methods for allocating resources between Gemini and ChatGPT. It goes deeper on their characteristic differences including the ability to naturally converse as evidenced by ChatGPT and multi-dimensionality and precision of Gemini.

Krause et al.(2023) covers the topic of the use of LLMs and generative AI in banking. It is focused on different types of AI models. Such models have displayed a tremendous potential in financial analysis, chores being automated and useful insights provided. They use their language generating skills to speed up financial evaluation processes, improve decision-making and create NLP in finance. The examination reveals the strengths and weaknesses of each model, when it comes to the importance of good prompting and verification of the results generated by an AI model.

Chen et al. (2023) discuss recent developments in generative AI in the business and finance. special focus on real-world applications of generative AI. The study shows how generative AI will be able to change how data is analyzed, not only in industry but also in academia. In particular, the application of ChatGPT is for analyzing financial statements and sentiment towards corporate environmental policies. The results suggest that such sentiment scores can be used as indicators for the ability of a firm to manage risks and even predict stock performance.

Natural Language Processing (NLP) is an important aspect of financial technology, and it is used for tasks such as sentiment analysis, named entity recognition, and question answering. While LLMs have demonstrated excellent performance in these applications, domain-specific financial LLMs have been relatively small. Addressing this gap, a new model called BloombergGPT was developed by Wu et al. (2023), which is a 50-billion-parameter model trained on a massive corpus of data on the topics of finance and general data and is therefore one of the largest models across a particular domain.

Similarly, Huang et al. (2023) created FinBERT, a language model with a finance-specific design in mind that was created to better understand the contextual meaning from financial texts. FinBERT is able to show better performance in sentiment classification, especially identifying the correct sentiments of domain-specific language, even with limited training data, compared to general models such as BERT.

Arched et al.(2024) attempts to discriminate between human-generated financial content and AI-generated content, focusing. It creates a dataset with both real human text and AI generated text. in this study the BERT model is shown to be able to distinguish between human and AI generated financial material. It highlights the need of testing different approaches of AI in the detection of fake material especially in the banking business of Pakistan. Marchena et al.(2024) provides two important insights for businesses and allows to better understand transformer-based models (TBMs), which are among the most commonly used generative AI models and have attracted much interest for their capability of processing and generating complex data.

2.1 Applications of Personalized Generative Models in Finance

Finance is in a constant mode of development and in the field, AI is playing an important role. The most promising AI technologies that have been applied in banking industry is LLMs and generative AI (Krause et all, 2024).

Generative AI models are quickly being applied to the automation and customization of generation of financial reports as well as adapting the generated outputs according to the requirements of different stakeholders. These models are used to provide customized, precise and informative financial data along with the help of advanced artificial intelligence techniques like DL and natural language generation (NLG). It uses ML to provide the detection of trends, anomalies and key performance indicators before the conversion of raw data to written information, infographics and actionable intelligence. Generative AI models for financial statements production are a good tool to provide customized accounts which are effective and informative to the stakeholders. It not only save time, costs, but also enable the firms to have a competitive edge by virtue of having better, a data driven interaction (Rane et al,2024).

Generative AI models are making a transformational impact on financial risk analysis and scenario evaluation because of its capacity to model complex distribution of data and model different financial scenarios. GANs may have the ability to simulate market behaviors such as stock price fluctuations and economic downturns. Diffusion models are employed for simulating slow movements in market circumstances during scenario generation. Transformer models are used to assess time-series financial data and create personalized forecasts of trends and interpretive risk evaluation reports using a mix of structured and unstructured information (Wang et al., 2023)

Generative AI models are transforming the method of delivery of personalised financial guidance by utilizing superior processing of information and pattern generating skills. These models deliver specialized perspectives, simulate finance circumstances, and build interactive advising systems to improve user experiences. It may create personalized investment strategies, savings programs, and risk evaluations and use conversational AI to provide real-time financial advice via chatbots or AI-powered assistants. Financial institutions are increasingly using generative AI models to improve identification of fraud and abnormality reporting. It identifies tiny irregularities and creates fake fraud situations to teach detection systems.

3. STATE-OF-THE-ART METHODS AND COMPARATIVE ANALYSIS

This section provides the overview of existing approaches and provides the basic working principle of some of the state –of-art methods like Generative Adversarial Networks; Transformer based Models, Diffusion Models, Variational Encoders(VAE's) and the comparative analysis of those methods.

3.1 Generative Adversarial Networks

The "adversarial" name arises from the fact that GAN is a ML architecture that pits two neural networks - the discriminator and the generator against each other. The generator network creates fake data by taking noise as inputs. To differentiate between legitimate and fraudulent inputs, the discriminator network uses an activation function that has a sigmoid and binary cross entropy loss. The generator only learns through interaction with the discriminator, who has access to both synthetic and real photos as it doesn't have direct access to real images.Backpropagation is used to improve the training process after categorization. Until the gap between the real and fake data samples is negligible, this process is repeated (Aggarwal et al., 2021). The conflict between them is a zero sum game in which the gain of one actor is the loss of another. In the case of a minimax game (zero-sum game), the discriminator (D) and generator (G) are trained together. Equation 8 shows that in the game, G is trying to maximize the likelihood of D misclassifying the output by D as data, while D is trying to minimize the chance that it misclassifies the output of G as data.

$$min_G max_D(D,G) = E_{x \sim p_{data}}(x)\{\log D(x)\} + E_{z \sim} p_z(x)[\log(1 - D(G(z))] \quad (1)$$

where $\boldsymbol{E}$ is the Expected Value, $\boldsymbol{P_{data(x)}(x)}$ is Real data distribution and $\boldsymbol{pz}$(x) implies Noise data distribution. The figure 4 shows the GAN architecture with generator and discriminator models.

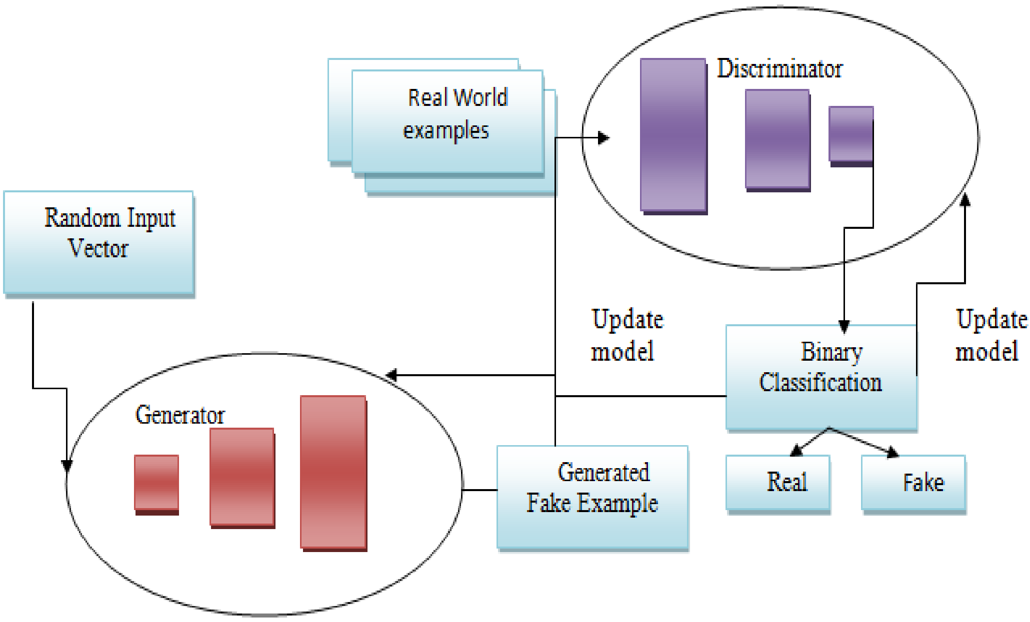

Figure 4 .GAN Architecture

3.2 Transformer based Models

The transformer architecture shown in figure 5 is a strong ML framework for natural language processing (NLP) problems. It learns to identify trends in sequential information such as written material or conversations. The model can anticipate the next element in the series based on its context, such as the word that follows in a phrase. It is ideal for translations and text creation.

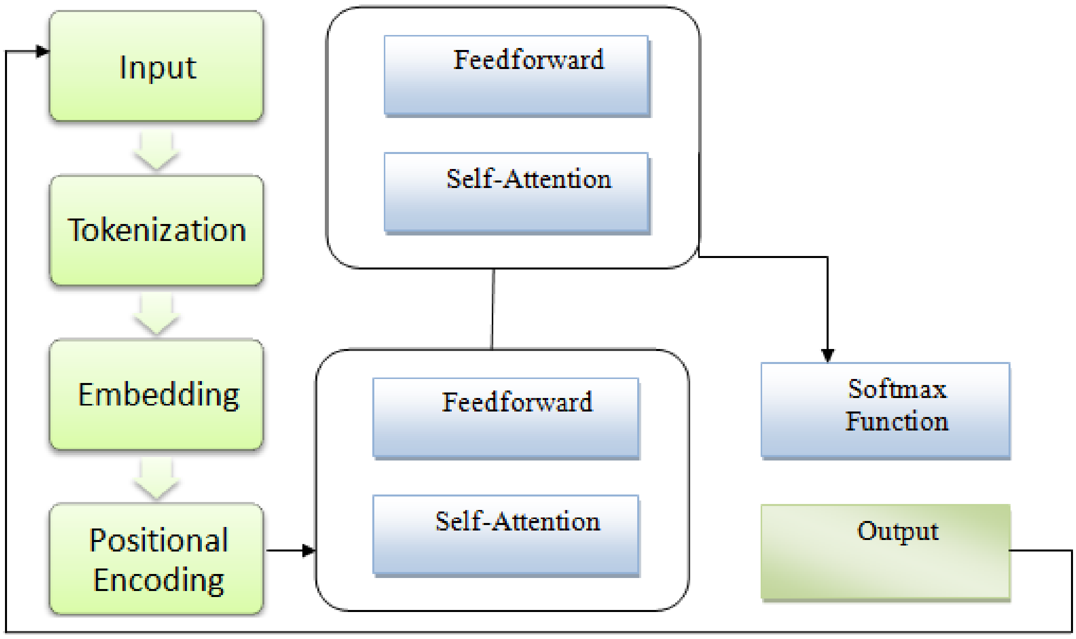

Figure 5 Transformer based models

It is made up of three major concepts: self-attention to assess input sequences based on their importance, reducing long-range dependencies; multi-head attention to learn various aspects of the input sequence; and word embedding, which converts inputs into vectors.

Encoder and Decoder: The encoder-decoder architecture where completely linked layers are used for both the encoder and decoder including stacked self-attention and point-wise layers forms the base of transformer architecture. The encoder is composed of a fully connected feed-forward network, a multi-head self attention mechanism and six identical layers with two sub layers each. With the exception of an added sublayer that deals with multi-head attention, a decoder is similar to an encoder. Sublayers have residual connections from both encoders and decoders before applying normalization (Zong et al., 2022).

Self Attention. It is a description of how to focus on the most important part of a sentence or information to better understand the context of a word. It consists of the transformation of a set of key-value pairs and a vector of queries into an output vector. Scaled Dot-Product Attention, which is comprised of queries, key dimensions dk, and dimension dv values computed using the following formula, is the definition of self-attention.In 2023, Bengesi et al.

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

Multi-Head Attention: By incorporating knowledge of the same attention pooling via different representation subspaces of queries, keys and values, multi-head attention mechanism asserts that self attention might be carried out in a recurrent manner and in parallel mode. Lastly a linear translation forth to the dimension anticipated is made by concatenating the separate attention outputs.

$$Multihead(Q,K,V) = concat(head_1 \dots\dots head_h)W^o \quad (3)$$

Where $head_i = Attention(QW_i^Q\ , K\ W_i^k\ , V\ W_i^v$

3.3 Diffusion Models

The diffusion model is a generative model that generates new data, such as photographs or sounds, by imitating the data it was trained with. These models are mostly connected with image production and other image processing activities, such as inpainting and super-resolution (Cao et al., 2024). DM has several primary compositions: denoising diffusion probabilistic models (DDPMs), stochastic differential equations (SDEs), and score-based generative models (SGMs).

DDPMs uses DM processes, whereby the forward technique introduces noises ε~N(0,1) into data distribution sample X0~p(X0) to create a noisy data distribution or prior distribution, p(Xt|Xt−1) using Markov chain as shown by equation 4, where index 0 represents the original data and p(X0) indicates the probabilistic density of the data.

$$p(X_t \mid X_{t-1}) = N(X_t\,;\sqrt{1-\beta_t\,X_{t-1},\beta_t\,I)} \qquad (4)$$

In equation 4 , I refer to the identity matrix, and βt represents the variance schedule over different diffusion steps (t). Use a single step to generate the noisy data Xt and the sample distribution illustrated in equation 5

$$P(x_t + x_0) = N(X_t\,;\sqrt{\bar{\alpha}_t\,X_0}\,,\sqrt{1-\bar{\alpha}_t\,I}) \qquad (5)$$

From equation 5, obtain the definition of Xt as indicated in equation 6.

$$X_t = \sqrt{\bar{\alpha}_t\,X_0} + \sqrt{1-\bar{\alpha}_t\,I} \qquad (6)$$

In contrast, the reverse procedure denoises noisy data beginning from a random sample Xt~N(0,I) to generate the original distribution specified in equation 7

$$P_\theta\left(X_{t-1} \mid X_t\right) = N\left(X_{t-1};\mu_\theta\left(X_t,t\right),\sum\nolimits_\theta\left(X_t,t\right)\right) \qquad (7)$$

3.4 Variational Encoders(VAE's)

Variational inference, a statistical method for approximating complicated distributions, was introduced to AE, which led to the development of VAE. An encoder and a decoder are the two halves of VAE, which is an unsupervised neural network. In the process of training phase, the encoder learns to reduce the input data into a less-complex representation (called latent space, which is lower-dimensional than the original data) that retains only the most important elements of the original input data. In a VAE, the modeling of each data point is not a fixed value, but a probability distribution, which adds the "variational" characteristic of the approach. During training, the encoder is used to map the input data into a latent space which is represented as a distribution and latent vectors are sampled from it. Then using the samples taken from the decoder vectors are passed to recreate the original input. The difference between the original and reconstructed data is called reconstruction error and backpropagated to optimize the network. Additionally, the techniques of regularization are included in the training process, which helps to control the overfitting and guarantee meaningful latent representations (Kingma et al., 2023). An overview of the architecture of the VAE is shown in Figure 6.

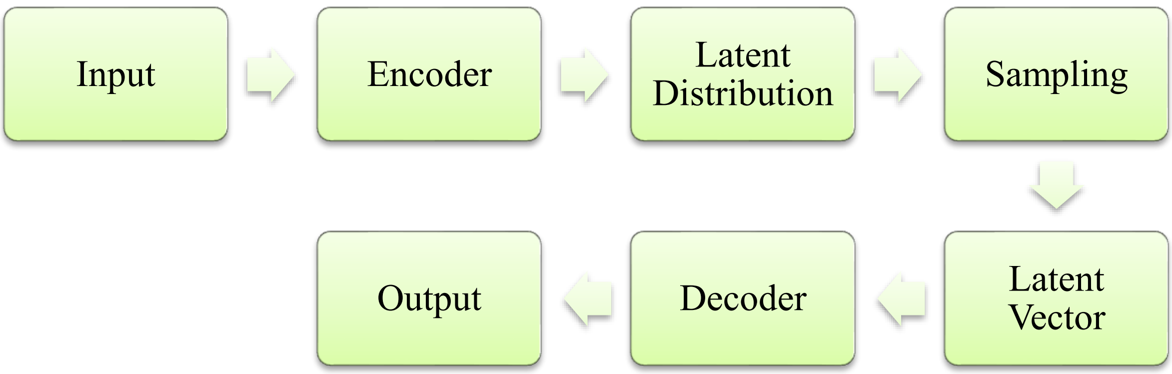

Figure 6 VAE Architecture

VAE is a probabilistic decoder that consists of a latent representation z from the prior distribution p(z) and data x from the conditional likelihood distribution p(x| z). The equation is given by $P(x,z) = P(x\,|z)\,p(z)$

The model's inference is tested by estimating the posterior of the latent vector using the Bayes theorem as indicated in equation.

$$p(z\,|x) = \frac{p(x \mid z)\,p(z)}{p(x)}$$

3.5 Comparison of existing Generative AI Models

The table 1 provides the detailed comparison of GAN, Transformer models, diffusion models and VAE

Table 1. Comparison of various Generative Models

Feature /Model	GAN	Transformer based Model	Diffusion Model	VAE
Architecture	Generator and Discriminator	Self attention	Denoising process	Encoder-Decoder
Use	Image generation	Text generation	Image Analysis	Latent space and image generation
Quality	High	High	Very high	Good
Complexity	Complex	Highly complex	High	Moderate
Diversity	Moderate	High	High	Good
Advantages	Generate realistic images	Performs well for sequential data	Stat of art quality	Stable training
Disadvantages	Unstable training, requires lot of tuning	Computationally intensive	Very slow, cost and resource intensive	Limited generative tasks

Choosing the proper generative model is dependent on the unique use case and limitations. GANs thrive at creating high-quality photos with complex textures, but they require consistent training. Transformers based models are good for handling vast volumes of sequential data and producing high-quality text, but they may be computationally intensive. Diffusion models now provide the greatest quality picture production, albeit at the expense of computing efficiency. VAEs are stable and interpretable, making them appropriate for learning significant representations, albeit they may trade crispness in produced outputs.

4. EVALUATION METRICS AND BENCHMARKS

The performance of generative AI models must be evaluated to verify that the generated material is of high quality, relevant, and effective. The selection of metrics relies on the kind of model (e.g., text, picture, audio creation) and the intended application.

Text Generation Metrics:

Perplexity: It measures the model's ability to anticipate a word sequence. Lower the perplexity, higher the performance. Perplexity is mathematically defined as:

$$Perplexity = 2^{-\frac{-1}{N}\sum_{i=1}^{N} log_2 P(w_i | w_1, w_2 ..... w_{i-1})} \qquad (8)$$

In equation (8) N indicates the total number of items in the set., $P(w_i | w_1, w_2 \ldots\ldots w_{i-1})$ denotes the probability of the ith token with w_i(Krasue et al., 2024)

Bilingual Evaluation Understudy (BLEU): It uses n-gram overlaps to compare produced and reference text. It is commonly used in machine translation, but also useful in other text-generation applications. BLEU score is defined as

$$BLEU = BP \cdot exp\left(\sum_{n=1}^{N} w_n \log p_n\right) \qquad (9)$$

In equation (9) BP is the Brevity penalty and it is defend as follows

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \le r \end{cases} \quad (10)$$

and P_n in equation 9 is the modified precision score and it is defined as

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n\text{-gram} \in C} Count_{clip}(n\text{-gram})}{\sum_{C' \in \{Candidates\}} \sum_{n\text{-gram}' \in C'} Count\ (n\text{-gram}')} \quad (11)$$

Where $count_{clip}$ in equation (11) implies counting the number of n-grams in the candidate translation that match those in the reference, but it is restricted to the maximum number of n-grams in the reference(Soni et al., 2024).

ROUGE (Recall-Orientated Understudy for Gisting Evaluation): It measures overlap between produced and reference text, with a focus on recall and it is useful for summarizing activities. The score ranges from 0 to 1, with a value near to zero suggesting little similarity between the candidate and the references and a score close to one indicating high similarity. It is calculated as follows(Nguygen et al., 2024)

$$ROGUE = \frac{Number\ of\ ngrams\ found\ in\ the\ model\ and\ reference}{Number\ of\ ngrams\ in\ the\ reference} \quad (12)$$

METEOR (Metric for the Evaluation of Translation with Explicit Ordering): To check that the Words are semantically matched with a focus on accuracy, recall and word order. The METEOR score ranges from 0 to 1; the higher the number the better the quality of translation. The METEOR score method is used to evaluate the translated text to the human reference translation by breaking it down into chunks and measuring the similarity between each chunk using metrics such as unigram accuracy, recall, F-score, bigram overlap, and precise word matches Finally, an average of these metrics taken into consideration provides the total METEOR score(MS).

$$MS = \text{Weighted\_Average(unigram accuracy, recall,} \\ -\text{score, bigram overlap, precise word matches)} \quad (13)$$

Content Diversity: Examines the diversity of the created material, which is usually measured as Type-token ratio(TTR) or entropy .The type token ration is calculated as follows.

$$TTR = \frac{Total\ Nmber\ of\ words(tokens)}{Number\ of\ unique\ words} \quad (14)$$

Higher TTR value indicates more diverse vocabulary. The entropy evaluates the unpredictable nature of token distributions.

$$Entropy = -\sum_i P(w_i) \log P(w_i) \quad (15)$$

In equation (15), the probability of the token w_i in the text is denoted by P(w_i), Higher entropy indicates higher lexical diversity(Doshi et al.,2023)

Image Generation -Quantitative Metrics

Fréchet Inception Distance (FID) compares generated and actual pictures using feature representations from a pre-trained Inception networ Lower FID means greater quality(O'relliy et al., 2021). It is calculated as follows.

- Compute the mean vector (μ) and covariance matrix (Σ) for both real and generated images
- Calculate the FID score as follows

$$FID = |\mu_r - \mu_g|^2 + \text{Tr}\ (\textstyle\sum_r + \sum_g - 2(\sum_r \sum_g))^{1/2}\ ) \quad (16)$$

In equation (16) , Tr denotes the Trace of the matrix and μ_r *and* μ_g are the mean of real and generated images. $\sum_r$ *and* $\sum_g$. are the co-variance of real and generated images.

The Inception Score (IS) assesses the quality and diversity of produced pictures. The formula for IS is as follows.

$$IS = \exp(E_x\ [\ D_{KL}\ (P(y\ |x)\ |p(y)] \qquad (17)$$

In equation (7) , p(y|x) is the conditional probability distribution of the classes for image xxx. and p(y) is the marginal distribution of the classes across all images. D_{KL} is the Kullback-Leibler divergence.

Precision and recall for generative models : Precision measures the fidelity of created pictures and recall measures the diversity of produced pictures

The Structural similarities Index (SSIM) measures the structural and texture similarities between produced and reference pictures.

Audio and speech generation - quantitative metrics

- Signal-to-Noise Ratio (SNR): Determines the clarity of produced audio.
- Mel Cepstral Distortion (MCD) measures the spectral difference between produced and target audio.
- Perceptual Evaluation of Speech Quality (PESQ): Evaluates speech quality in produced audio.
- The Log-Likelihood Ratio (LLR) measures how well audio attributes fit with target distributions.

To capture technical performance and real-world effect, generative AI models must be evaluated using a combination of automated and human-centric measures. The metrics are inline with the unique aims of the application, so that the model cannot only generate good quality results but can satisfy the demands of the users successfully.

Personalization-Specific Metrics

- Accuracy: Determines how closely the generated output matches a user's profile, preferences, or past data.
- Context Relevance: Determines how effectively the generated material fits inside a certain context, such as a user query, circumstance, or setting.
- Preference Matching: Measures how closely generated outputs mirror the user's expressed preferences.
- User-Centric Novelty: Determines if the model introduces previously unknown outputs that remain consistent with the user's choices.
- Clickthrough Rate (CTR): Measures how frequently users interact with produced results.

4.1 Challenges and Limitations

Despite their advantages, these models still have issues with accuracy and dependability which brings up a question of confidence and morality in these models and AI in general. To make matters more ambiguous, there is little to no direction on how these models should and should be used. As LLMs become an even more integrated part of society, the threats that they create may outweigh the benefits they offer(Lo et al., 2024).

Data Privacy Challenges: Generative AI models often require access to large datasets that contain personal data in order to train the model and generate content. If proper management is not provided to this information, the illicit access and the attack on information may be caused. It is hard to guarantee that data being used for training and inference is anonymised to secure user identities. Differential privacy techniques are employed to anonymize data but these techniques have an effect on the model performance and usability critical judgment.(Arslan et al., 2024).

Ethical and Bias Issues: Generative AI algorithms that are inputted with historical or biased data can perpetuate negative preconception or discriminatory behavior Regulatory organizations in some areas are beginning to address algorithmic bias in their compliance of AI. Personalized content must be generated in a way that will not jeopardize the safety or dignity of users. This is especially important in industries like healthcare and finance, where, material may have an impact on important decisions.

Scalability & Computational Complexity: : In order to deliver quality outputs, generative AI models, and specifically those based on DL, require humongous quantities of data in order to train on. The collection, storage and interpretation of these huge data sets can be quite resource intensive. Training generative AI model from scratch might take weeks to months despite availability of high-end hardware because of the need to run several iterations for big data sets.Tuning hyperparameters to maximize training is computationally expensive and require multiple trials for each of which might require large resources(Desai et al., 2024)

5. CONCLUSION

This review provides insights into the fundamentals of generative AI models and it application in finance .It also discuss the existing state-of-art methods with its features, advantages and drawbacks. Generative AI models have greatly revolutionized the process of personalized content generation and have particularly impacted the financial industry, making it possible to provide automated, data-driven, and user-centric solutions. This study has gone over the major state-of-the-art models such as GANs, Transformers, Diffusion Models, and VAEs, their architectures, capabilities, and limitations. GANs and Diffusion models have better quality in content generation, while Transformers are better at processing sequential data and VAEs have stable and interpretable representations. Despite their benefits, challenges, data privacy, ethical bias, high computation cost, limitations on scalability make their broad implementation to real world applications.Future studies should aim at the development of efficient and lightweight generative models with low computational overhead and high output quality. Integrating privacy-preserving methods like federated learning and differential privacy can be used to increase data security. Improvement of model robustness to limit bias and adversarial attacks is also a key to trustworthy AI systems. Additionally, the introduction of human-in-the-loop methods and tasks/specific fine-tuning in specific domains can make a personalization and usability boost. Advancements in hardware acceleration and optimization techniques will only aid in the scalable and realtime deployment to use generative AI in a dynamic and complex financial environment for widespread adoption.

REFERENCES

[1] Eckerli, F., & Osterrieder, J. (2021). Generative adversarial networks in finance: an overview. *arXiv preprint arXiv:2106.06364*.

[2] Selvaraj, A., Selvaraj, A., & Venkatachalam, D. (2022). Generative Adversarial Networks (GANs) for Synthetic Financial Data Generation: Enhancing Risk Modeling and Fraud Detection in Banking and Insurance. *Journal of Artificial Intelligence Research*, *2*(1), 230-269.

[3] Bai, X., Zhuang, S., Xie, H., & Guo, L. (2024). Leveraging generative artificial intelligence for financial market trading data management and prediction.

[4] Qian, K., Fan, C., Li, Z., Zhou, H., & Ding, W. (2024). Implementation of Artificial Intelligence in Investment Decision-making in the Chinese A-share Market. Journal of Economic Theory and Business Management, 1(2), 36-42.

[5] Zacharias, J. (2024). Public and Expert Insights into Generative AI: The potential for the Financial Industry. In *INFORMATIK 2024* (pp. 1491-1500). Gesellschaft für Informatik eV.

[6] Julia Matuszewska, Jerzy Biernacki(2024), Generative AI in Financial Services: Use Cases, https://www.miquido.com/blog/generative-ai-in-financial-services/

[7] Patil, D., Rane, N. L., & Rane, J. (2024). Applications of ChatGPT and generative artificial intelligence in transforming the future of various business sectors. *The Future Impact of ChatGPT on Several Business Sectors*, 1-47.

[8] Ooi, K. B., Tan, G. W. H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., ... & Wong, L. W. (2023). The potential of generative artificial intelligence across disciplines: Perspectives and future directions. *Journal of Computer Information Systems*, 1-32.

[9] Liu, X. Y., Wang, G., Yang, H., & Zha, D. (2023). Fingpt: Democratizing internet-scale data for financial large language models. *arXiv preprint arXiv:2307.10485*.

[10] Alwahedi, F., Aldhaheri, A., Ferrag, M. A., Battah, A., & Tihanyi, N. (2024). Machine learning techniques for IoT security: Current research and future vision with generative AI and large language models. *Internet of Things and Cyber-Physical Systems*.

[11] Linkon, A. A., Shaima, M., Sarker, M. S. U., Nabi, N., Rana, M. N. U., Ghosh, S. K., ... & Chowdhury, F. R. (2024). Advancements and applications of generative artificial intelligence and large language models on business management: A comprehensive review. *Journal of Computer Science and Technology Studies*, *6*(1), 225-232.

[12] Lee, J., Stevens, N., Han, S. C., & Song, M. (2024). A survey of large language models in finance (finllms). *arXiv preprint arXiv:2402.02315*.

[13] Rane, N., Choudhary, S., & Rane, J. (2024). Gemini or ChatGPT? Efficiency, Performance, and Adaptability of Cutting-Edge Generative Artificial Intelligence (AI) in Finance and Accounting. *Efficiency, Performance, and Adaptability of Cutting-Edge Generative Artificial Intelligence (AI) in Finance and Accounting (February 19, 2024)*.

[14] Naik, D., Naik, I., & Naik, N. (2024, November). Applications of AI Chatbots Based on Generative AI, Large Language Models and Large Multimodal Models. In *The International Conference on Computing, Communication, Cybersecurity & AI (The C3AI 2024). Springer.*

[15] Lo, A. W., & Ross, J. (2024). Generative AI from Theory to Practice: A Case Study of Financial Advice.

[16] Krause, D. (2023). Large language models and generative AI in finance: an analysis of ChatGPT, Bard, and Bing AI. *Bard, and Bing AI (July 15, 2023)*.

[17] Chen, B., Wu, Z., & Zhao, R. (2023). From fiction to fact: the growing role of generative AI in business and finance. *Journal of Chinese Economic and Business Studies*, *21*(4), 471-496.

[18] Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., ... & Mann, G. (2023). Bloomberggpt: A large language model for finance. *arXiv preprint arXiv:2303.17564*.

[19] Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, *40*(2), 806-841.

[20] S. Bengesi, H. El-Sayed, M. K. Sarker, Y. Houkpati, J. Irungu and T. Oladunni, "Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers," in *IEEE Access*, vol. 12, pp. 69812-69837, 2024, doi: 10.1109/ACCESS.2024.3397775

[21] S. Bengesi, H. El-Sayed, M. K. Sarker, Y. Houkpati, J. Irungu and T. Oladunni, "Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers," in *IEEE Access*, vol. 12, pp. 69812-69837, 2024, doi: 10.1109/ACCESS.2024.3397775.

[22] D. P. Kingma, S. Mohamed, D. J. Rezende and M. Welling, "Semi-supervised learning with deep generative models", *Proc. Adv. Neural Inf. Process. Syst.*, 2014, Aug. 7, 2023

[23] . M. Zong and B. Krishnamachari, "A survey on GPT-3", *arXiv:2212.00857*, 2022.

[24] C. Zhang, C. Zhang, S. Zheng, Y. Qiao, C. Li, M. Zhang, et al., "A complete survey on generative AI (AIGC): Is ChatGPT from GPT-4 to GPT-5 all you need?", *arXiv:2303.11717*, 2023.

[25] A. Aggarwal, M. Mittal and G. Battineni, "Generative adversarial network: An overview of theory and applications", *Int. J. Inf. Manage. Data Insights*, vol. 1, no. 1, Apr. 2021.

[26] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, et al., "Diffusion models: A comprehensive survey of methods and applications", *ACM Comput. Surv.*, vol. 56, no. 4, pp. 1-39, Nov. 2023.

[27] H. Cao, C. Tan, Z. Gao, Y. Xu, G. Chen, P.-A. Heng, et al., "A survey on generative diffusion models", *IEEE Trans. Knowl. Data Eng.*, 2024.

[28] Arshed, Muhammad Asad, Ştefan Cristian Gherghina, Dur-E-Zahra, and Mahnoor Manzoor. 2024. "Prediction of Machine-Generated Financial Tweets Using Advanced Bidirectional Encoder Representations from Transformers" *Electronics* 13, no. 11: 2222. https://doi.org/10.3390/electronics13112222

[29] Marchena Sekli, G. (2024). The research landscape on generative artificial intelligence: a bibliometric analysis of transformer-based models. *Kybernetes*.

[30] Dahal, S. B. (2023). Utilizing Generative AI for Real-Time financial market analysis opportunities and challenges. *Advances in Intelligent Information Systems*, *8*(4), 1-11.

[31] Franceschelli, G., & Musolesi, M. (2024). Reinforcement learning for generative ai: State of the art, opportunities and open research challenges. *Journal of Artificial Intelligence Research*, *79*, 417-446.

[32] J. A. O'Reilly and F. Asadi, "Pre-trained vs. Random Weights for Calculating Fréchet Inception Distance in Medical Imaging," *2021 13th Biomedical Engineering International Conference (BMEiCON)*, Ayutthaya, Thailand, 2021, pp. 1-4, doi: 10.1109/BMEiCON53485.2021.9745214.

[33] Arslan, B., Lehman, B., Tenison, C., Sparks, J. R., López, A. A., Gu, L., & Zapata-Rivera, D. (2024). Opportunities and challenges of using generative AI to personalize educational assessment. *Frontiers in Artificial Intelligence*, *7*, 1460651.

[34] Desai, A. P., Mallya, G. S., Luqman, M., Ravi, T., Kota, N., & Yadav, P. (2024). Opportunities and Challenges of Generative-AI in Finance. *arXiv preprint arXiv:2410.15653*.

[35] Soni, A., Arora, R., Kumar, A., & Panwar, D. (2024, August). Evaluating Domain Coverage in Low-Resource Generative Chatbots: A Comparative Study of Open-Domain and Closed-Domain Approaches Using BLEU Scores. In *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)* (Vol. 1, pp. 1-6). IEEE

[36] Nguyen, Q. H. (2024). *Automatic evaluation of companies' alignment with EU Taxonomy using Large Language Models* (Master's thesis, University of Twente).

[37] Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The power of generative ai: A review of requirements, models, input–output formats, evaluation metrics, and challenges. *Future Internet*, *15*(8), 260.

[38] Doshi, A. R., Bell, J. J., Mirzayev, E., & Vanneste, B. S. (2024). Generative artificial intelligence and evaluating strategic decisions. *Strategic Management Journal*.