Journal of Informatics and Web Engineering

Vol. 4 No. 3 (2025) eISSN: 2821-370X

The Role of Generative AI in e-Commerce Recommender Systems: Methods, Trends and Insights

Kai-Ze Liau[1*], Heru Agus Santoso[2]

[1]Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia

[2]Department of Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia.

corresponding author: (1211102809@student.mmu.edu.my; ORCiD: 0009-0005-5781-2047)

Abstract - Recommender systems have existed for decades, shaping how people consume digital content, receive information, and engage in day-to-day activities, among others. Undoubtably, recommender systems also play a crucial role in e-commerce applications as well, with industry players like Amazon, AliBaba, eBay using recommender systems within their ecosystems to give suitable and value-driven insights. However, recommender systems face some main concerns such as data sparsity, cold-start problems and so on. As a result, research is currently ongoing to solve these issues and provide high-quality recommendations to consumers. This review aims to identify prevailing gaps surrounding these issues by analysing existing research on generative Artificial Intelligence (AI) recommender systems within an e-commerce context. It explores the underlying framework of common generative AI techniques such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Transformers, diffusion models and so on. VAEs and Transformers hold great potential within e-commerce as noted by most researchers due to their ease of training and qualitative generations. This review intends to enhance recommender systems better to improve the quality of life of digital users, providing better recommendations in e-commerce as well as maximizing the value of stakeholders. It also includes potential future work for researchers to advance existing knowledge in this sector.

Keywords—Recommender Systems, Generative Artificial Intelligence, e-Commerce, Generative Adversarial Networks, Variational Autoencoders, Transformers, Diffusion

Received: 21 February 2025; Accepted: 25 May 2025; Published: 16 October 2025

1. INTRODUCTION

The advent of the Internet enables people to engage in commerce online, reducing the need for physical stores and can cut down on operating costs. Thus, e-commerce was born and has benefitted both customers and shop owners alike since then. However, one of the main drawbacks of e-commerce involves customers not being able to find what they like due to the vast arrays of products available throughout an e-commerce platform. Therefore, recommender systems are commonly employed to recommend items that customers are likely to buy. Recommender systems are

MULTIMEDIA UNIVERSITY

Journal of Informatics and Web Engineering

https://doi.org/10.33093/jiwe.2025.4.3.3

Published by MMU Press. URL: journals.mmupress.com/jiwe

tasked with recommending related items for a particular user based on previous interactions such as likes or dislikes, comments, browsing history and so on. This enables the shopping experience to be distinctly catered to a particular customer's interests, making the shopping experience more interesting and informative, not to mention being highly personalised.

Various recommender systems are typically employed within the e-commerce domain. Industry players such as Amazon and eBay use state-of-the-art recommender systems within their systems. Traditional recommender systems were often used as an essential tool for many decades, but with the advent of generative AI, it can provide individualised advice catered to a particular individual by generating human-like text and context, thus creating a more appealing, detailed and persuasive such that users can find what they exactly need. As a result, this review will be focused on modern generative AI applications within the recommender system domain.

Generative AI recently obtained mainstream attention when ChatGPT gained mainstream popularity. It is a generative AI model with a Large Language Model (LLM) based on a Transformer model that generates text and other outputs based on an input by a user. This framework is the basis of generative AI that can be used in a recommender system context. As such, this paper aims to utilise this to enhance recommender systems further to help with accuracy and efficiency in recommendations. When effectively implemented and used, it can have enormous potential benefits by improving revenue, time efficiency and market differentiation in an industrial standpoint. Thus, it is crucial to improve research in this domain, and this paper could be a stepping stone for further advancement by highlighting the potential, obstacles, and future research directions.

To aid in this endeavour, several research questions and objectives are formulated as a point of reference in this review, namely:

- What are the current state-of-the-art methods and techniques in generative AI recommender systems?
  Section 2 highlights these methods in brief in order to better understand the subject matter.
- What are the prevailing generative AI frameworks that is commonplace in e-commerce recommender applications?
  Section 3 meticulously detail current trends in e-commerce recommender systems by conducting a thorough literature review.
- What are the outstanding research gaps and trends that can be found?
  Section 4 provides a comprehensive analysis on current trends and gaps that are underlying within this domain.

This review underlines various generative AI applications within e-commerce, encompassing concepts such as GAN, VAE, Transformers and diffusion models. These methods will be discussed in brief and recent research using these frameworks are also discussed, dating from 2020 to 2025. A thorough examination on how the frameworks aid in recommendation tasks, such as recommendation fairness, product recommendation, cross-domain recommendations, multi-modal use cases and more. The evaluation metrics used to identify the effectiveness of the various proposed frameworks are also discussed and examined. This paper serves to contribute to the field of e-commerce recommender systems by giving a thorough review of state-of-the-art contributions by researchers, whereby it is then analysed by their advantages, disadvantages, and relevance in e-commerce recommenders. It also provides insight into crucial challenges encountered by generative AI recommender systems, and further outlines research directions to better advance research in this field.

The paper is arranged as follows: Section 2 illustrates the theoretical background of the core concepts surrounding generative AI recommender systems; Section 3 provides the literature review of current research, specifically focusing on five core frameworks: GANs, autoencoders, GAN-VAE hybrids, Transformers, and diffusion models; Section 4 provides the analysis obtained from the selected papers, including comparisons, identifying gaps, trends and so on; Section 5 summarises and concludes the review as a whole.

2. BACKGROUND

2.1 Overview of Recommender System

A recommender system is a system that provides suggestions or recommendations that are helpful to the user based on previous inputs and decisions made by the user [1], [2]. Recommender systems play a significant role in maximizing value in different sectors, and their importance is increasing year after year. In 2009, Netflix created the Netflix Prize of $1,000,000 to identify the best algorithm created by teams to predict user ratings for films based on some given information, thereafter highlighting the importance of recommender systems. A yearly conference by the Association for Computing Machinery (ACM) named the ACM Conference on Recommender Systems where it encourages academic discussion and advances in this sector [3]. This further highlights the imperativeness of recommender systems in technology.

Briefly, recommender systems use user feedback in the form of direct or indirect methods towards a particular product or service and tailor it uniquely for the user. Direct forms of feedback include ratings for products or movies and commenting on social media posts, while indirect forms of feedback involve the customer buying a product that is related or having similar characteristics to a previously purchased product in the e-commerce domain. On top of that, watch time for video streaming is also categorised as an indirect form of feedback [2]. Various relationships can be deduced from user feedback and product characteristics and can be categorised as follows. A user-product relationship is the most common, where a user prefers a product that they need. For example, if a user prefers action movies, the recommender system recommends more action movies. Next, a product-product relationship occurs when products have similar attributes and qualities, where the recommender system will group them, like books from the same author or music from the same artist. Finally, a user-user relationship materialises when users exhibit similar tendencies and aspects, such as mutual friends or the same demographic. In turn, the recommender system will tend to recommend relevant items to the users [4]

There are multiple ways of examining the performance of different recommender systems. The most imperative source would be the general accuracy of the model. In other words, is the recommender system capable of providing accurate results when given a certain set of inputs. Accuracy is usually calculated with various formulas and metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), while usage prediction is measured by precision and recall, Normalised Discounted Cumulative Gain (NDCG) and so on [5-8].

The main challenges surrounding recommender systems is a cold start. A cold start is when the system has little to no starting information regarding users or items and thus is unable to function properly. Three prominent cases of cold start usually happen when there are new communities, when the start of a recommender, community and users provide no information, making it hard for recommenders to perform reliably. New users can also hinder recommenders with zero interactions with the system thus having no datapoints for it to function. Newly introduced items also pose a threat as only product metadata is known, and interactions with users have not yet occurred. Collaborative filtering (CF) is more prone to novel items than Content-Based (CB) filtering due to its high reliance on interactions. To mitigate the impact of cold start, hybrid filtering is often used along with multiple strategies such as profile completion for users to aid with both CF and CB filtering. Generative AI methods significantly reduce the problems surrounding cold starts as the recommender can instant obtain instant feedback [2], [9].

Recommender systems exist everywhere in our online services, such as YouTube for videos and content recommendations, Spotify for music suggestions and Amazon for e-commerce applications. For this review, we will focus on recommender systems' e-commerce segment. E-commerce is electronically purchasing or selling products and services across the Internet. Normally, products such as online shopping and services such as music consumption are transacted across the Internet, bypassing physical commerce's time and space requirements, thus improving transaction efficiency [10]. Currently, the most popular e-commerce platforms are Amazon, with its online shopping services; eBay, with online auction capabilities; and Netflix, with media streaming. All the examples provided currently employ a recommender system to ensure maximum value is obtained for each transaction and customer.

2.2 Phases in RS

2.2.1 Traditional RS

First, CF was one of the crucial methods used before generative AI was developed. It can be further split into two main categories: user-based CF and item-based CF. Firstly, user-based CF at the most basic level, involves comparing users, such as searching for other users that give ratings close to the active user and then calculating and predicting the ratings of the active user. One specific application using user-based CF would be the Nearest Neighbour Algorithm. Next, item-based CF calculates the similarity between items by using people's ratings of those items. It was created and used by Amazon in 1998 and was published in a conference in 2001. It was agreed that item-based CF accomplished better than user-based CF. One class of CF used in the Netflix Prize would be matrix factorization that decompose the user-item interaction matrix into the product of two lower-dimensionality rectangular matrices [11]-[13].

Next, CB filtering is the second of two main categories of traditional recommender systems. CB filtering methods are based on a description of the item, such as name, location, description, and a profile of the user's preferences, such as background, ethnicity, location, and gender. Keywords are used to describe items and associate them with user profiles. This approach has foundations in information retrieval and information filtering research. A user profile is generated from a model of the user's preference and previous interactions with the recommender system. A weighted item vector will be computed and assigned to the items and specific users. Simple calculations will involve averages of the item vector, while complex calculations usually involve machine learning (ML) techniques such as Bayesian Classifiers, cluster analysis, decision trees, and artificial neural networks (ANN). A key problem of CB filtering is that attributes for one specific item might not be translated into similar attributes for another item. For example, music recommendations might be accurate, but when transitioning into e-commerce, preferences might not be accurately reflected, although the items from music and e-commerce share similar attributes. This issue could be rectified by using both CF and CB filtering, thus creating a hybrid filtering mechanism. CB filtering could include opinion-based filtering, such as users leaving feedback on items. This would improve the accuracy of metadata attributed to a specific item due to the wide range and accuracy of aspects and features described by users. Deep learning, sentiment analysis, information retrieval, and text mining would help in this endeavour [14], [15].

A combination of CF and CB filtering and other filtering methods is classified as a type of hybrid filtering. Some studies have suggested that hybrid filtering is superior to standalone CF and CB filtering in terms of accuracy and solving the cold start problem. Netflix employed hybrid filtering in their algorithms where they compare watching and searching statistics of users and compare them among similar users in addition to suggesting films that share characteristics with other films with favourable ratings, but deep learning methods, but newer deep learning approaches are currently employed [16]. Several hybridization methods exist, including weighted scores of different recommendation structures: Switching which is choosing between recommendation components and then applying the selected one; Mixed involves recommendations taken from other recommenders which are presented together to provide the recommendation results; Cascade, where recommenders are given priority, with the lower priority ones breaking ties in the recommendation scoring of the higher ones; Finally, meta-level recommendation technique is applied and produces a model, which is then the input used by the following techniques [17].

To end, Knowledge Graphs (KGs) have recently gained in popularity as a traditional recommender technique. KGs is classified into two main categories: embedding-based and path-based methods. Embedding-based uses pre-processed KGs using KG embedding algorithms, which turns information into a vector from which the system can learn [18], [19]. Path-based KGs use connection patterns within the graph to deduce relationships between attributes. Many methods are used to compute relationships, including matrix factorization, deep learning, and some model user-user, item-item, and user-item relationships [20], [21].

2.2.2 Generative AI based RS

In this review, the focus will be directed on generative AI based recommender systems. Traditional recommender systems typically obtain data points from a few certain sources and usually span a shallow comprehension of the customer base in general. With the advent of generative AI, it can now model and sample a wide variety of datasets, including user-item interactions, text, images, and videos, making novel recommendation tasks easier. Key advancements in generative AI within recommender systems include interaction-driven generative models the use of LLMs and textual data for natural language recommendation, and the integration of multi-modal models for generating and processing images/videos in recommender systems [22]. Not to mention the various challenges encountered by using traditional models such as data sparsity, cold start, and diversity to name a few, still existed.

GANs were introduced in 2014 by [23]. They pit two identical neural networks, a discriminator, and a generator, against each other in a zero-sum game where one neural network's gain is another one's loss. The generator is tasked to generate quality and accurate data with the provided random noise as input, which it will then convert into intended data such as images, videos, text, and so on. The generated data will be compared against the discriminator to determine whether it is real or generated data, where a failed comparison will mean the discriminator has been "fooled", which is the ultimate objective of the generator. For example, a GAN trained on photographs can generate brand new photographs that are somewhat authentic to human observers due to their realism [23]. Though originally suggested as a form of a generative model for use in unsupervised learning, GANs also proved beneficial for reinforcement learning [24], semi-supervised learning [25], and fully supervised learning [26]. The main benefit of a GAN is that it produces high-quality results given the proper training and criteria, especially in image synthesis, video synthesis, music synthesis, and other tasks. Apart from that, GANs are especially versatile and can handle a myriad of different challenges, including generative AI. However, the drawback of a GAN includes the computational cost used in training a GAN model, especially when introduced to a large and complex dataset. Unstable convergence, mode collapse, and the vanishing gradient problem are also core concerns surrounding GANs. Moreover, fairness and bias relating to the usage of GANs can be observed by many researchers [22], [27]. Figure 1(a) illustrates a simple GAN structure.

An autoencoder is an ANN used to learn efficient coding of unlabelled data. It learns an efficient coding for a set of data to generate lower dimensional representations used by other ML algorithms [28]. Many variants exist to make the generated representations useful, especially VAEs used for generative AI, first introduced by Kingma & Welling in 2014. In the heart of a VAE lies an encoder and decoder, where the encoder transforms raw input data and translates it into a probability distribution inside a latent space. The decoder network takes a sample point from the distribution and reassembles it into data space [29]. Training and calibration are required to ensure reconstruction loss is minimised. To optimise the VAE model, it requires a delicate balance of two critical components, which are reconstruction loss and Kullback-Leibler (KL) divergence [30]. Moreover, KL divergence uses Evidence Lower Bound (ELBO) to minimise KL divergence by maximizing the ELBO, which is part of its training process [31]. The main concern about VAEs is that data generated may not be as accurate, whereas its frequently compared counterpart, GANs, typically produce better quality data. On the contrary, VAEs are easier to train than GANs because they have more structured latent space [32]. Due to the various pros and cons presented by both VAEs and GANs, some researchers opt for a hybrid version of both methods and found that it has enormous potential and has produced promising results in anomaly detection and topological data generation, to quote some examples [33], [34]. Figure 1(b) illustrates a simple VAE model.

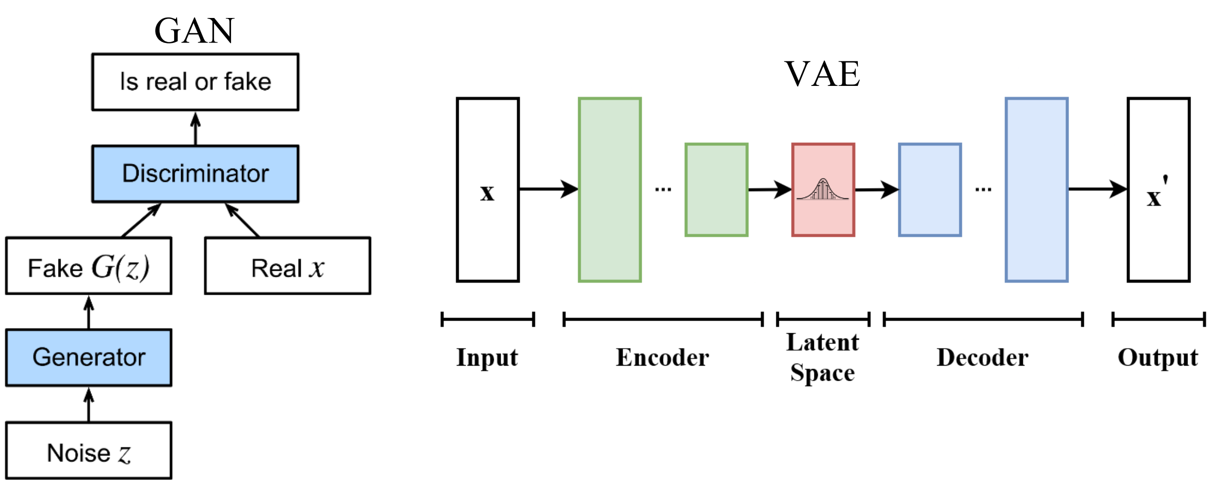


Figure 1. A General Illustration of a (a) GAN, (b) VAE.

Diffusion models also play a significant role in generative AI currently, especially in the generation of image generators like Stable Diffusion and DALL-E. Apart from computer vision, diffusion models also boast use cases in Natural Language Processing (NLP) [35] that include text generation [36], [37] and summarization [38], reinforcement learning [39], [40], and sound generation [41]. Its capability in NLP, especially in text generation, provides an alternative to generative AI in certain use cases. The main components of a diffusion model include a forward diffusion and reverse diffusion process or a diffusion and denoising process. Its main objective is to train the model to reverse the diffusion process and predict the noise added [42]. The main benefit of a diffusion model is that the data produced is often of high quality, sometimes surpassing GANs. Still, unlike GANs, training diffusion models are easier and more stable compared to GANs. Its main drawback, however, is the computational cost and complexity when using diffusion models, on top of its slow sampling speed. Despite the pros and cons, more generative AI projects use diffusion models, including Google and Meta, in their video generators [43], [44].

A landmark paper published by Vaswani et al. in 2017 proposed a new model of generative AI named Transformer, where text is converted into a numerical representation called tokens to be quickly processed. Transformers have the same dominant components: tokenisers that convert text and information into tokens and an embedding layer that converts tokens and token positions into vector representations. Transformer layers undergo repeater transformations of the vector representations, which extract linguistic information. Lastly, the un-embedding layer converts the final vector representation back into a probability distribution of the tokens [45]. This general architecture persists as the backbone of many mainstream generative AI applications. The famous Generative Pre-trained Transformer (GPT) generative AI lineup by OpenAI became the state-of-the-art generative AI usage, where the Transformer architecture was used. Since then, Transformers have been used in multiple sectors, such as vision Transformers [46], speech recognition [47], robotics [48], and multi-modal learning [49]. Image and video generators like DALL-E [50], Stable Diffusion 3 [51], and Sora [52], are based on the Transformer architecture [53].

3. LITERATURE REVIEW

3.1 GAN-Based

In their paper, Bock & Maewal [54] proposed a conditional, coupled GAN (RecommenderGAN) for recommender systems. The dataset used by Bock & Maewal is an e-commerce dataset which comprises of 2,756,101 behavioural events observed on 1,407,580 individual visitors, with 417,053 distinct items in 1669 product categories represented in the data. Product conversion rate and category similarity were used to evaluate the model, where the conversion rate is the number of items suggested and bought compared to the count of all items recommended, and category similarity is defined as the Jaccard index. The findings, however, are preliminary and may benefit customers and digital retailers. Computed conversion rate statistics range from 1.323% to 1.723%, which is significant compared to null hypothesis testing results and comparable to published conversion rates across industries and product types. Disadvantages were found, such as numerical efficiency, where the algorithm used involves high computational cost and large compute times. A more efficient programming language such as C++ might curb the issue. Moreover, there is no ranking of recommended results in the model used, as the GAN produces binary-valued information.

Zhou et al. [55] proposed a Positive-Unlabelled Recommendation (PURE), adopting work by Kiryo et al. [56]. It includes the positive-unlabelled risk minimiser to train an unbiased positive-unlabelled discriminator. Movielens100k, Movielens1M, and Yelp datasets were used in their experimentation. It was then compared to traditional CF, matrix factorization methods, along with modern neural CF, GAN recommender and PU-learning recommenders. It outperforms the closest baselines with a 1%-2% improvement on average, using metrics like precision, NDCG, Mean of Average Precision (MAP), and mean reciprocal rank (MRR). Running time is also desirable, following a log scale time complexity. It was observed that other GAN-based methods are susceptible to converging failure even with meticulous parameter tuning. This is due to perceiving the unobserved data as negative samples without the negative sampling procedure, ensuing in an unbalanced training data problem, notably in ultra-sparse data. Next, continuous space sampling was not done, and generating with discrete sampling will end up with bad model expressiveness, especially when handling with sparse large-scale datasets.

Li et al. [57] proposed a Multi-modal Adversarial Representation Network (MARN) for predicting Click-Through Rate (CTR), an essential metric for e-commerce implementations. It was experimented on datasets from Amazon and Taobao, along with comparisons with multiple baselines, encompassing various ML and deep learning frameworks such as Logistic Regression, Recurrent Neural Networks, Long Short-Term Memory (LSTM) networks and so on [58]. A receiver operating characteristic curve (ROC curve) was employed as a metric by measuring the area under the curve (AUC), along with online, real-world A/B testing. The improvement over the best baseline model is a 0.76%, 0.84% and 0.88% improvement for AUC scores. Online testing also shows improvements, with the model improving CTR by 5.23% and Gross Merchandise Volume (GMV) by 2.26%.

Yuan et al. [59] proposed a GAN recommender named Convolutional Generative CF (Conv-GCF). It includes an effective perturbation mechanism (adversarial noise layer) for convolutional neural networks (CNN). Four datasets were used in the experiments, namely: MovieLens-1M, Ciao, GoodBooks, and YahooMusic, where GoodBooks is an e-commerce platform. Two metrics were used to compare Conv-GCF against five selected baselines: Hit Ratio (HR) and NDCG. It outperforms all selected datasets, and the authors conclude by stating that building a new GAN-based model using pre-trained embeddings is more practical than adding noises and training using the original model in which the generator and discriminator are the same.

Li et al. [60] suggested a novel approach to fairness in recommendations, named FairGAN, which consists of two components. Firstly, a ranker models user preferences based on observed interactions. Next, a controller utilises the distribution of item's exposures based on the ranking generated by the ranker. The controller then dynamically generates fairness signals, enabling the ranker to allocate an item's exposure fairly. The proposed controller is to generate various fairness signals based on varying tasks [61], [62]. In summary, FairGAN aims to allocate exposure to items fairly and preserve users' utilities, adopting only positive feedback in implicit feedback and not negatively treating unobserved interactions. Experiments demonstrate that it outperforms selected methods of recommender systems as well. Four distinct datasets were used from Amazon [63], and were evaluated using precision, recall and NDCG, while fairness is assessed by Individual Exposure Disparity (IED). The lower the IED, the fairer the recommendations. The results proved that FairGAN exhibits an average improvement of 9.62%, 12.52%, 7.07% and 5.62% on recommendation quality and 36.15%, 24.02%, 17.90% and 14.82% on fairness on all four datasets, respectively. Future work by the authors will include investigating the issues of fairness across users and simultaneously exploring methods to improve items and user's fairness.

Shafqat & Byun [64] used a hybrid GAN approach in their research based on the architecture of conditional GAN [65], Wasserstein GAN with gradient penalty (WGAN-GP) [66], and PacGAN [67] called CWGAN-GP-PacGAN that is tasked to generate tabular data with categorical and numerical data. This method combines the auxiliary classifier (AC) loss and the Wasserstein loss with gradient penalty to tackle both categorical and numerical data where three architectures are combined for developing the generator, discriminator, and AC. The data used in the experiments is obtained from an online shopping platform in Jeju, South Korea called the eJeju Mall. The authors use metrics such as Fréchet Distance (FD), correlation coefficient (CORR), RMSE, mirror column association, MAE, and percent root mean square difference (PRD) for evaluating synthetic data, whereas recall, accuracy, MRR, and F-score were used for performance evaluation. To conclude, the novel architecture of the model enables the authors to condition the discriminator and the AC loss simultaneously. As such, the inclusion of AC further improves upon the performance and inhibits the error rate of the recommender system remarkably. It has been demonstrated that CWGAN-GP-PacGAN generates better synthetic data, and recommendations have been considerably enhanced. REN et al. [68] mentioned that high-dimensional data reduced the generative capacity of the model, causing a loss in quality and diversity. Moreover, it is susceptible to mode collapse, where the discriminator fails to cover all categories present in the data distribution. As such, the generated samples tend to be similar or lacking in quality.

Wei et al. [69] designed a Double GAN (Double-GAN) to solve issues of sparse consumer behavioural data. It consists of a double-layer iteration mechanism that iteratively compensates the original data of the e-commerce platform to mitigate data scarcity. A dataset of two bookstore platforms was used, dating from 2016 to 2019. This is then compared against four baseline models, along with two evaluation metrics, precision, and MAP. Consequently, it has alleviated the cold-start issue, along with helping platforms to achieve better user management, enhance user portraits, achieve user authentication, and administer network security. Future work includes fusing data from two domains to analyse multi-source recommendation scenarios.

3.2 Autoencoder-Based

Liu et al. [70] proposed a Deep Global and Local Generative Model (DGLGM) based on existing VAEs. Initially, a deep global generative recommendation model (DGGM) was proposed, and it was extended further by introducing DGLGM. DGLGM is based on Wasserstein autoencoder frameworks [71], [72], and it adopts a non-parametric Mixture Gaussian distribution with various components that capture the diverse users' preferences. Two main parts make up the proposed DGLGM model, namely a Beta-Bernoulli distribution as to model the implicit feedback of all users and secondly a Mixture Gaussian distribution which comprises of several Beta-Bernoulli structures to capture the diversity of user preferences sufficiently. Various datasets, such as MovieLens, Netflix, Yelp, and Epinions, were used. The users range from 49,290 to 1,182,626, while items range from 17,770 to 156,638 for all four datasets. MovieLens and Netflix have around 20 million and 100 million interactions, respectively, while Epinions and Yelp both have the highest percentage of near-cold-start users, with values above 69% for both. Various metrics, such as recall and NDCG, were used in the author's experiments. The results were compared against multiple baselines of distinct backgrounds. As a result, DGLGM was superior to all baselines and showed strength in highly sparse datasets such as Epinions and Yelp. However, in the author's future research [73] which was also based on the DGLGM model, inter-user preference similarity and intra-user preference diversity by investigating observed-level and latent-level disentanglement, it does not perform as well as expected. Furthermore, the generalization ability and convergence properties are further improved with the newer research.

Drif et al. [74] proposed a VAE-based recommender system named The Ensemble VAE framework for recommendations (EnsVAE). It aims to reduce the user-item interaction bias and improve the collaborative systems' effectiveness. It consists of recommenders (sub-recommenders), with their rating matrices values tweaked to output interest probabilities. The matrices are then merged by using a probabilistic aggregation function and given to the VAE to analyse possible user-item interaction patterns. Newly introduced users or resources do not affect the trained recommender system with few datapoints. Two datasets were used to evaluate the proposed model: MovieLens [75] and Amazon [63], which have a sparsity of 95.5% and 99%, respectively. Two metrics were used in the research: MAP and NDCG. As a result, the proposed model outperforms the baselines used despite a simple aggregation function that provides conclusive predictions without incurring major penalties to the overarching performance. Future work proposed by the authors includes exploration of better aggregation functions on the sub-recommenders, such as Bayesian approaches. According to the authors, context-aware recommender systems are also worth exploring as to give better recommendations specific to different contexts.

Shao et al. [76] introduced a fine-grained controllable generative model named *Apex*, that is tasked to generate product descriptions and item recommendations in Taobao. It employs a variant of the proportional, integral, and derivative (PID) controller to alter the diversity/accuracy trade-off in generated text which is then inserted into a conditional VAE (CVAE). Real-world e-commerce datasets were used, obtained from Taobao, then it is compared with baselines of various backgrounds such as GANs and Transformers. A/B testing showed a CTR improvement of 13.17% and an item recommendation improvement of 6.89% of CTR. It is able to achieve this due to an ideal manipulation of KL divergence, a crucial measure in VAE optimizations. As a result, it can achieve commendable accuracy and diversity, while avoiding the KL vanishing problem.

Truong et al. [77] proposed a Bilateral VAE (BiVAE), which treats users and items similarly, thus being “bilateral”. However, it may suffer from over-regularised latent space, which is posterior collapse [78], similar to the original VAE model proposed. To mitigate this issue, the authors introduce a constrained adaptive prior (CAP) for learning user and item-dependent prior distributions. Datasets from MovieLens, Amazon, and Epinions were used to evaluate performance. It is then compared with four baselines by using evaluation metrics such as NDCG and recall. Thus, it can be found that BiVAE outperforms the baselines, with CAP alleviating the posterior collapse issue. Future work includes identifying other methods of building informative priors, implementing BiVAE to other types of dyadic data like document word matrices, and alternative tasks such as co-clustering [79].

Hasumoto & Goto [80] used a VAE to extract latent features from purchase histories as explanatory variables. It is done in three stages. Firstly, data is prepared with Recency, Frequency, and Monetary values (RFM) measures along with VAE input matrices. Next, the inputs are then churned by the model, giving outputs that are then analysed further. The dataset was provided by a business platform with 60k customers in total. It is then evaluated with accuracy, precision, recall, and F-score. A 20% improvement over the baseline is recorded, thus demonstrating the effectiveness of the model. To conclude, adding latent variables as explanatory variables to a churn prediction model enhance the prediction performance. Some limitations are noted, where the model is planned to capture complex user behaviour on a platform business, where it will limit the effectiveness of the model if applied to a single business. Next, if users have a longer tenure, the behaviour of the customers before churn may differ and may not be efficiently captured by the model. Lastly, the model can be enhanced by using different models, like CNNs and LSTM networks, as well as network models that better fit the selected datasets.

Chen et al. [81] aimed to solve the common bottleneck of VAEs which is the softmax computations [82], where they decompose the softmax probability with the inverted multi-index. They also implement efficient sampling procedures for the approximate softmax distributions, where items can be sampled independently in sublinear or constant time. The designed model is then used in four real-world datasets, namely MovieLens10M, Gowalla, Netflix, and Amazon. It is then compared with other CF models as baselines, achieving 2.61% and 1.72% improvements in NDCG and recall metrics, respectively. Superior efficiency is also noted with computational efficiencies of sublinear or even constant time complexity.

Zhu & Chen [83] targeted the issues of sparsity and inefficiencies in user-oriented autoencoders (UAEs), where they suggest a mutually-regularised dual collaborative VAE (MD-CVAE) by replacing randomly initialised last layer weights of the vanilla UAE with stacked latent item embedding. Three datasets were used in comparisons, namely: citeulike-a [84], MovieLens [75], and Amazon [63]. Recall and NDCG are used as evaluation metrics. Various factorization-based and autoencoder-based baselines were used for comparisons. To conclude, typical issues encountered by UAEs, such as sparsity and cold-starts, are addressed by MD-CVAE. It also can be easily generalised to multi-class classification tasks where new classes constantly appear after model training and deployment.

Xia et al. [85] suggested using a Graph Neural Network (GNN) for CF use cases. However, current methods overly rely on manually generating effective contrastive views for heuristic-based data augmentation that does not generalise over varying datasets. Thus, Automated CF (AutoCF) is proposed with the addition of a masked graph autoencoder that captures the global collaborative relationships for reconstructing the masked user-item subgraph structures. Gowalla, Yelp and Amazon is used as datasets for this research, with comparisons from various baselines such as: Conventional CF Methods, Autoencoder-based, GNN-based CF, Disentangled Representation-enhanced GNN Model, and SOTA Self-Supervised Recommendation Methods. Recall and NDCG is chosen as metrics, with AutoCF improving over 14 chosen baselines. Computation efficiency is also superior to all baselines.

Yang et al. [86] proposed a framework based on existing VAEs called Memory Pool VAE (MPVAE) along with the usage of memory pools which is a new attention mechanism that simultaneously aggregates information and establishes similarities between each other, giving efficiency in computation. The authors use three real-world datasets from Amazon, are Movies, Music, and Books, and each user or item will have at least five ratings. MAE and RMSE were used in their evaluation, and it has been found that an average of a 13.31% and 8.55% overall improvement on all three tasks can be found. Despite the promising results, the model does not yet properly incorporate target information, which presents challenges when dealing with recommendations pertaining to non-sparse information instead of benefitting from sparse target data. On top of that, only single-target cross-domain recommendations (CDR) was achieved, and dual-target CDR is left for future work by the authors by merging the memory pool.

Gandhudi et al. [87] proposed an explainable causal VAE-based equivariant GNN, which combines causal modelling technique to identify important causal relationships, with VAEs to learn latent data representations, and GNNs to predict the e-commerce purchase behaviour of consumers. Datasets include marketing data, and sale analysis from Kaggle. As a result, it accurately recommends purchases, surpassing baselines with low mean squared error (MSE) of 4.49, MAE of 0.74, RMSE of 2.11, mean absolute percentage error (MAPE) of 4.75, and high R-Squared (R^2) of 97.17, an R^2 improvement of 10.72% over baselines.

Zheng et al. [88] proposed an LLM-based recommendation model called LC-Rec, along with a Residual-Quantised VAE (RQ-VAE) for generating item indices. It first uses LLMs, specifically LLaMA [89] in encoding the text information for an item, and use text embeddings as the initial item representation. The RQ-VAE is then trained based on information gathered prior. Amazon datasets [90] were chosen, along with nine other comparative baselines from various backgrounds. HR and NDCG is used as evaluation metrics, with LC-Rec having improvements of 7.39% to 68.62% over closest baselines. Future work includes exploring ways to extend the current approach in a multi-turn conversational setting, such that it supports more flexible user interaction.

3.3 Transformer-Based

Zhu et al. [91] used a multi-modal, transformer-based recommender named K3M in their proposal, consisting of three layers. First, the modal-encoding layer is tasked to separately encode individual information of each modality. Second, the modal-interaction layer aims to model the interaction between different modalities. Third, modal-task layer, and there are different pretraining tasks for varying modalities. Transformer-based image and text encoders were used in all layers involved. A dataset from Taobao is used in the experiments, and results are compared with image and text modality training baselines. Another stage of comparisons, which involves knowledge modality, is also carried out in addition to images and texts. As a result, advancements over chosen baselines can be seen using F-score as a metric. Future work from the authors includes applying K3M to more downstream usages and exploring its capability on more general datasets.

Dong et al. [92] applied a Transformer model towards five distinct modalities: table, text, image, video, and audio. The proposed Masked Region Prediction task (MRP) and the Masked Language Modelling task (MLM) within the overarching model is tasked to deal with image and text modalities. In contrast, Mask Entity Modelling task (MEM), Mask Frame Prediction task (MFP), and Mask Audio Modelling task (MAM) are tasked with table, video, and audio modalities. It is compared with similar multi-modal models by using a dataset crawled from an e-commerce company, and an improvement of 2% can be found using accuracy, MAP, and precision as metrics. However, generative capacity of modal representations is lacking, and image and caption generation may be potential avenues to explore.

Deng et al. [93] explored a novel Personalised Answer GEneration method (PAGE) to solve Product Question Answering (PQA) issues in e-commerce platforms. It combines a Transformer and the Bidirectional Attention Flow (BiDAF) [94] as the encoder-decoder architecture. Datasets from Amazon are used which includes a Question/Answer dataset paired with a product dataset [63]. It is then compared with other prevailing generative PQA models along with personalised text generation models. ROUGE F1 (R-1, R-2, R-L) and Embedding-based Similarity (ES) [95] is used as evaluation metrics, as well as Persona Coverage(C_{per}), Users-Distinct ($uDist$), and User-Language-Perplexity ($uPPL$) [96]. These metrics are commonly used to evaluate PQA and personalised text generation use cases. As a result, it significantly outperforms the baselines in answer generation. It also effectively generates highly diverse personalised answers of user-centric information as well as user-preferred language styles. For future work, the authors note that employing a multi-modal approach such as implementing user-item interaction modelling could be beneficial. Besides, PQA can be extended to other platforms like user-centric forums, such as Stack Overflow or Reddit.

Geng et al. [97] presented a "Pretrain, Personalised Prompt, and Predict Paradigm" (P5) for recommendations, utilizing five different task families: rating, sequential recommendation, explanation, review, and direct recommendation to discover information about users and items. Amazon and Yelp datasets were used in evaluation, where it is compared against various baselines of diverse backgrounds. RMSE and MAE is used to determine errors, while HR and NDCG is used for performance evaluation. As such, P5 achieved commendable results in all five task families. Future exploration may include expanding the model size of P5 and employing more superior models such as GPT-3, OPT, and BLOOM. Building on top of this paper, a future paper proposes Visual P5 (VIP5) to improve on P5 itself [98]. It provides multimodal personalised prompts to support the modalities. It gives the capability of parameter-efficient tuning instead of pre-training in existing recommendation foundation models such as P5, along with further improvements of performance with both less training time and less memory usage. Advancements over the prior P5 model can be seen, with improved efficiency as well. Future work includes further scaling up the backbone model, incorporating more modalities, and exploring better prompting strategies.

Rajput et al. [99] provided a new standard of generative retrieval models named as Transformer Index for GEnerative Recommenders (TIGER). The framework consists of two stages. The first is semantic ID generation using content features, which involves encoding item content features to embed vectors and converting them into a tuple of semantic codewords. The tuple is then referred to as the Semantic ID. A Transformer model is then trained using sequences of Semantic IDs. Amazon Product Reviews dataset [63], which contains metadata from May 1996 to July 2014, was used. Recall and NDCG were used as evaluation metrics. It is judged that TIGER outperforms the benchmarks provided with an improvement of 15% to 29%. A cold start simulation is also experimented on, where it outperforms the baseline, along with experiments on recommendation diversity, where it also excels. A main detriment of this model is that it is possible to predict invalid Semantic IDs where an ID does not map to anything in the provided dataset. Despite that, the model predicts valid IDs most of the time, where the percentage of invalid IDs varies from 0.1% to 1.6% as observed. Prefix matching of Semantic IDs as an extension could be considered as future research to fix this issue and could improve the recall and NDCG metrics further. Moreover, it is noted that the TIGER algorithm is more computationally expensive compared to conventional Approximate Nearest Neighbours-based models due to usage of beam search for autoregressive decoding. Memory cost, however, is better when compared to other traditional recommender systems since each item requires embedding for each item while TIGER only requires an embedding for each semantic codewords.

Chu et al. [100] used a Transformer in their proposal of RecSysLLM, a pre-trained model based on LLMs. It includes a novel mask mechanism, span order, and positional encoding to insert inter- and intra-entity knowledge into the LLM, which leverages prior work by Du et al. [101], ChatGLM. The dataset used is from Alipay, with comparisons against industry LLMs, ChatGPT and GPT-4, with HR and NDCG used as metrics. Surprisingly, the proposed model did not outperform baselines, a unique case study in this review. ChatGPT and GPT-4 outperforms RecSysLLM on almost all scenarios, with RecSysLLM outperforming the others in only one scenario. Despite that, the model overpowers ChatGLM in all scenarios, superseding prior research by Du et al. Further experimentation on a wider range of tasks might show the strengths and limitations of the proposed approach, particularly on a more diverse domain set could provide insight on how robust the learned representations are.

Li et al. [102] proposed a Graph Transformer (GFormer) that gives parameterised collaborative rationale discovery [103], [104] for selective augmentation while maintaining global-aware user-item interactions. Yelp, Ifashion, and LastFM were used as datasets for experimentation. Three main approaches of recommenders were chosen as comparative baselines, namely: Non-GNN CF Approaches, GNN-based Recommendation Methods without self-supervised learning (SSL), and SSL-enhanced Recommendation Models. Recall and NDCG is chosen as evaluation metrics, with GFormer regularly outperforming all given baselines, including strong SSL-enhanced methods. GFormer is also robust against artificial noise and sparse data conditions. The prevailing open question that surrounds

GFormer is adapting it in other recommendation scenarios, like social-aware recommendations and knowledge graph-enhanced recommenders.

Li et al. [105] proposed a prompt-based Recommendation Language Model (RLM) called Personalised Automatic Prompt for RECommendation (PAP-REC) that is based on work by Geng et al. [97]. The authors design surrogate metrics to tackle issues stemming from recommendation-specific metrics and develop the token update schedule to solve issues from inflating personalised tokens, as well as leveraging the gradient to generate effective automated prompts. Amazon was chosen the as dataset, and the experiments were evaluated using HR and NDCG. Superior performance is noted against chosen baselines in most cases. The authors note other methods for automated prompt generation, like reinforcement learning and LLMs. Potential research also includes searching the best prompts for pre-training or fine-tuning, which could bring new application scenarios for automated prompt generation that enhance the performance of LLMs.

Zhang et al. [106] proposed an LLM-based approach for recommender systems, called InstructRec. It is based on work by Chung et al. [107], which has been fine-tuned on the Text-to-Text Transfer Transformer (T5) architecture [108]. The authors tune the proposed model based on recommendation-oriented instruction data, instead of the original instruction data that is not catered to recommender systems. Datasets from Amazon is used for experimentation, and based on prior studies, they enhance the sequential data through data augmentation, extracting sub-sequences from the full user behavioural sequence to better capture dynamic changes in user preferences [90], [109], [110]. HR and NDCG is used as evaluation metrics, with improvements of 2.98% to 44.64% when compared to various baselines. A multi-turn interaction scenario is considered future work, where users can engage in interactions with the systems conversationally. It is also challenging to model excessively long user behavioural sequences directly, as the context length is severely hindered.

Ugurlu et al. [111] proposes Style4Rec, a recommender that uses style and shopping cart data to enhance existing Transformer-based sequential product recommenders due to existing models being unable to utilise product image information and shopping cart information effectively. It uses a neural style transfer algorithm [112] to obtain style information from item images, that are utilised as embeddings by the algorithm. The shopping cart information is then used in the training and validation stages, not in testing. This gives a better evaluation of real-world situations. An e-commerce dataset was used, with Style4Rec being compared with BERT4Rec and SASRec [113], [114], similar to many previous examples. Improvements range from 0.4% to 9.4%, with HR, NDCG, and MRR as evaluation metrics. Future work involves determining the performance of the model in scalable functions, as well as more complex recommendation scenarios.

3.4 Diffusion Models

Wang et al. [115] used the newer diffusion models and proposed a novel Diffusion Recommender Model (DiffRec) to learn the generative process in a denoising way. DiffRec reduces added noises and avoids corrupting user's interactions to retain personalised information. It also aims to tackle high computational costs and temporal shifts in traditional diffusion models [116]. As such, the authors propose two extensions of DiffRec, namely Latent DiffRec (L-DiffRec) where it clusters items for dimension compression and executes the diffusion process in latent space and Temporal DiffRec (T-DiffRec) where it reweights user interactions using timestamps to encode temporal information. Three real-world datasets were employed: Amazon Books, Yelp and MovieLens 1M. Two metrics were used in their evaluation: recall and NDCG. DiffRec was compared to various baselines and was found to outperform the baselines under clean training with an average of a 11.56%, 3.49% and 5.42% improvement over second-best baselines in Amazon, Yelp and MovieLens datasets, respectively. For noisy training, DiffRec also performs better against the generative model, even under large noise. The computational and memory cost is also greatly improved compared to other baseline recommenders. The authors recommend several research directions for future work, such as devising better model compression and encoding temporal information such as Transformers. Moreover, controllable, or conditional recommendations like guiding the prediction based on a pre-trained classifier is advisable. Finally, exploring different noise assumptions other than Gaussian distribution and diverse model structures is preferred.

Next up, Wu et al. [117] also employed a diffusion framework in their work called Diff4Rec. It consists of a diffusion model suited for modelling user-item interactions in a latent space and a curriculum scheduler that progressively adjusts user-item interactive sequences with generated data. The model is pre-trained on recommendation data through diffusing and denoising the interactions in latent space. At the same time, the curriculum scheduler is tasked to progressively provide the generated samples into the sequential recommenders within two levels, namely interaction augmentation and objective augmentation. Four datasets were used in the experiments across diverse backgrounds: MovieLens 1M, Amazon Beauty, Steam and Yelp with at least a 95% sparsity for all datasets. Evaluation metrics employed include HR, NDCG and MRR, where Diff4Rec improved over selected baselines of 10.14%, 18.57%, 11.46% and 6.20% for MovieLens, Amazon, Steam and Yelp datasets, respectively. The authors recognise the tendency for diffusion models to be often used for image generation, where it is difficult to integrate into product recommenders. Moreover, a diffusion model trained for user-item relations cannot guarantee that it can always bring benefits towards recommendation use cases.

Li et al. [118] attempted to adapt diffusion models in sequential recommendations by proposing DiffuRec, for item representation construction and uncertainty injection. It models item's representations as distributions, instead of fixed vectors, which mirror the user's multiple interests and the item's various aspects adaptively. It is then paired with another deep neural network (DNN) acting as a Approximator to reconstruct target item representation for training, for example, a Transformer. Amazon Beauty, Amazon Toys, Movielens-1M, and Steam is used as datasets in this experiment. Baselines are used to compare against DiffuRec, consisting of Conventional Sequential Neural Models, Multi-Interest Models, and VAE and Uncertainty Models. The model achieves up to 57.26% and 56.72%, improvements for HR and NDCG, respectively.

Zolghadr et al. [119] builds on similar work by Li et al. [118] by introducing several enhancements, specifically adding offset noise in the diffusion flow to enhance robustness and incorporating a cross-attention mechanism in the Transformer-based Approximator to capture relevant user-item interactions better. It not only learns temporal dependencies; it also learns more convoluted and complex relationships between past user-item interactions. Datasets from Amazon and MovieLens were used to evaluate the proposed improvements, with HR and NDCG acting as evaluation metrics. The subtle improvements aided in an average accuracy improvement of 1.5% compared to closest baselines, as well as better convergence training times, suggesting better efficiency gained. It is noted that shorter dataset sequences may lack sufficient information to accurately predict user preferences, while overly long sequences cause challenges for model performance. Averaging the predictions generated from random seeds could solve the issue by accounting for various aspects of behavioural uncertainty through the aggregation of diverse recommendation outcomes.

Jiang et al. [120] proposed a multi-modal graph diffusion model for recommendation named DiffMM to address prevailing issues such as self-supervised learning for recommenders that usually depend on simple random augmentation or intuitive cross-view information that introduces unimportant noise and fail to align the multi-modal context with user-item interaction modelling accurately. TikTok and Amazon datasets were used to compare performance with traditional CF methods, diffusion methods, and so on. Recall, precision, and NDCG serves as evaluation metrics, with improvements noted across all fields against the closest baseline, BM3 [121]. Integrating LLMs to guide the diffusion process with their powerful semantic understanding is a potential future work by the authors.

3.5 Hybrids

In their paper, Li et al. [122] introduced a deep sparse autoencoder predictor based on GAN learning for cross-domain recommendations (DSAP-AL) model. It is based on CDR systems that are created to deal with data sparsity problems. There are four major steps in their proposal, namely:

a. Joint matrix factorization, integrates and aligns the latent factor spaces through the GAN model (GAN training).
b. A deep sparse autoencoder model transfers the factor knowledge and weights using cross-domain representations.
c. Irregular noise factors are identified and removed from the factor spaces through the proposed algorithm, and solid user and item factors given as output.
d. Regularization constraints are added into the model for learning, and the final recommendation results are obtained.

Multiple datasets were used in their experiments, such as AmazonBooks, Epinions, and MovieLens. RMSE and MAE were used as evaluation metrics for DSAP-AL, where the lower the respective values, the smaller the prediction error. In the experiments, precision, recall and F-score are also used as metrics; the larger the value, the better the recommendations. The results show that the proposed model was better against the frameworks used and is more robust when compared to several types of sparse data. The time complexity of the model used is also better when comparing all the models used. On the contrary, static rating profiles are primarily used in their paper, and they cannot sustain dynamic requirements of most recommenders. Thus, the author's future research will be focused on the effect of temporal and contextual information on cross-domain recommender systems.

Xiao et al. [123] employed a GAN-VAE hybrid in their model, Generative Multi-modal Fusion Framework (GMMF), for predicting CTR, which includes a Difference-Set Network (DSN) paired with a Modal-Interest Network (MIN). DSN is tasked to eliminate the redundancy of multi-modal information, while MIN is to model the modality-specific user's preference for CTR prediction. Datasets from Amazon is used, and the performance is compared with several baselines. Offline and online testing were conducted, with AUC chosen as a metric. It achieved 0.53%, 0.87%, 1.51%, and 0.02% improvements in all four datasets over the best baseline. Online A/B testing is also done, with a noted CTR improvement of 6.64%.

Liu et al. [124] presented a hybrid of VAEs, Transformers, and attention mechanisms in their paper, called the VATA model. According to the authors, VAEs can obtain the implicit characteristics of customer behaviour, while Transformer models can more comprehensively obtain the dependencies between behaviours. The e-commerce dataset summarises the shopping history of users on e-commerce platforms, which is aggregated from multiple platforms [125]. Accuracy, Recall, F-Score, and AUC is used as evaluation metrics, with VATA outperforming selected baselines. In terms of efficiency, it has the least inference time and training time compared with others. Limitations include complex shopping scenarios, specifically in terms of performance fluctuations and model stability. As such, future work is to be focused on improving the model's robustness and generalization ability, and the interpretability of the model to better understand complex shopping decision-making processes.

4. ANALYSIS

4.1 General Overview

The papers are selected based on several criteria: the papers must be in the English language; dated from 2020 to 2025; pertinent to e-commerce, generative AI, and containing new and novel contributions. The papers were searched and screened through Google Scholar, containing keywords such as "VAE," "GAN," "generative AI," "Transformers," "diffusion models," and "e-commerce." Then, a preliminary overview on the abstract and keywords given to judge whether the paper is suitable to be utilised. The majority of papers were sourced from ACM linked journals, followed by arXiv, ScienceDirect journals, Institute of Electrical and Electronics Engineers (IEEE) journals, and other publishers like Springer, The Conference and Workshop on Neural Information Processing Systems (NeurIPS), and MDPI. A total of 36 papers were used in the literature review, with the distribution of seven GAN-based papers, 11 VAE-based papers, 10 Transformer-based papers, five diffusion models, and 3 hybrids of various generative models. From this sample size alone, VAEs are the preferred model for generative AI recommenders, with Transformers being closely followed. Figure 2(a) shows the various model distributions.

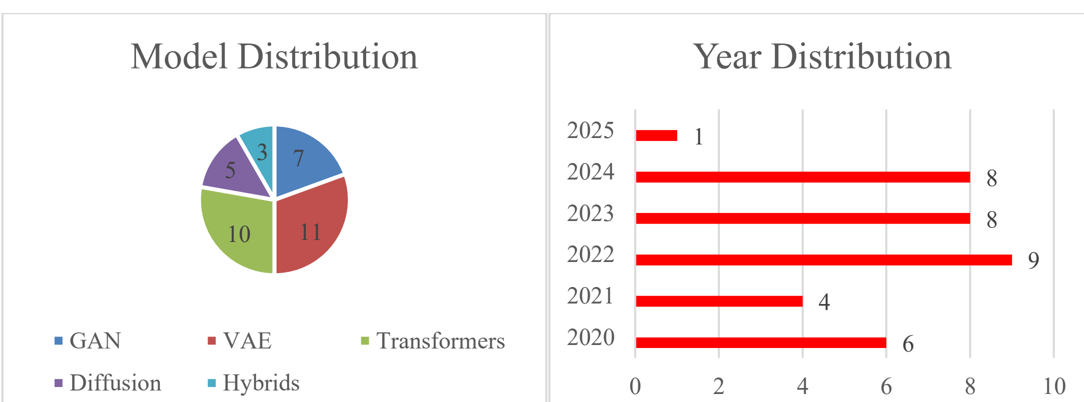


Figure 2. Distribution of the Selected Papers in Terms of (a) Model, and (b) Year

Papers from 2020 to 2025 were selected due to their recency, and it signifies the current trends within generative AI recommenders. The distribution of papers by year is detailed in Figure 2(b). Six papers dated to 2020, four in 2021, nine in 2022, eight in 2023, eight in 2024, and finally, only one as of March 2025, for a total of 36.

Various evaluation metrics were used in the experiments, NDCG, MAP, MAE, RMSE, precision, recall are among the common evaluation metrics used by all the authors. Figure 3 highlights the distribution of metrics used. NDCG is the most prevalent metric, with recall and HR being the second and third most common. It is worth noting that miscellaneous metrics are also prevalent, such as MAPE, accuracy, R^2, CORR, AUC, online A/B testing and so on, classified under "Misc.".

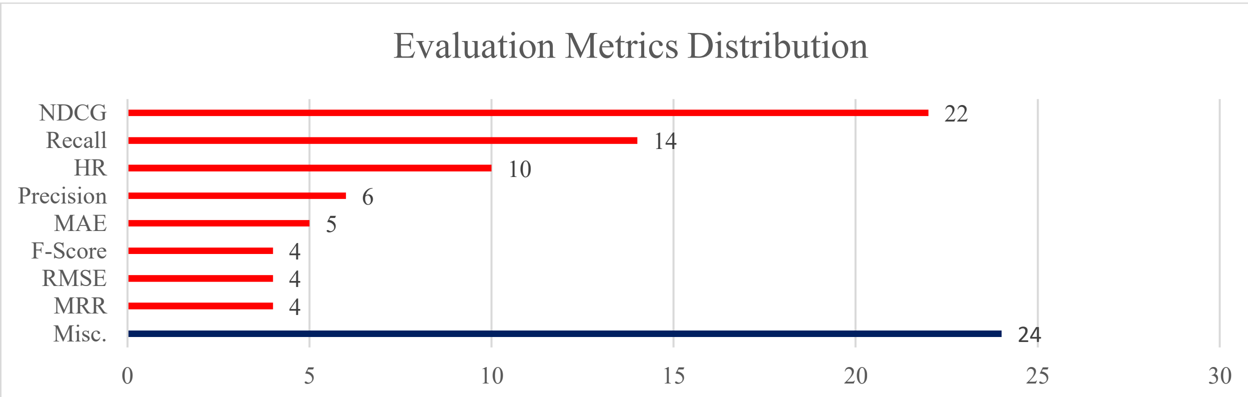
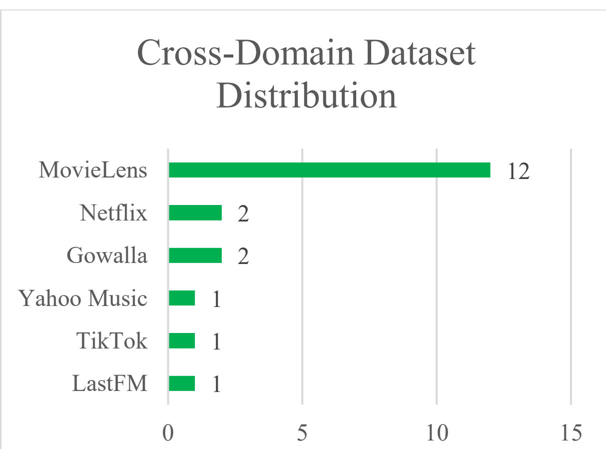


Figure 3. The Distribution of Evaluation Metrics Across 36 Selected Papers.

Datasets pertaining to e-commerce is the common feature that can be found in all papers, along with many other domains of data that the authors also choose to include, i.e. recommenders that does not focus solely on e-commerce. Common e-commerce platforms include Amazon, Yelp, and Taobao, each of them showcasing real-world scenarios, often in highly sparse situations.

Figure 4(a) lists the datasets used. Cross-domain recommender systems are also common in the reviewed papers, with the majority having an e-commerce recommender paired with a movie recommender as well. These cross-domain datasets are also included in Figure 4(b).



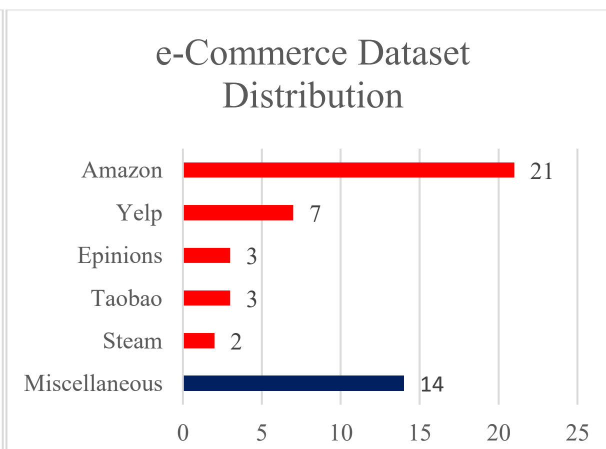


Figure 4. Distribution of the Dataset Used in Terms of (a) e-commerce, and (b) Miscellaneous.

A summary of the selected papers is detailed in Table 1, sorted by model type.

Table 1. A Summary of 36 Selected Papers

Core Framework	Paper	Year	Dataset	Metrics
GAN	[54]	2020	E-commerce platform	Jaccard Index
	[55]	2020	Yelp, MovieLens	Precision, NDCG, MAP, MRR
	[57]	2020	Amazon, Taobao	AUC, online A/B
	[59]	2020	GoodBooks, MovieLens, Ciao, YahooMusic	HR, NDCG
	[60]	2022	Amazon	Precision, recall, NDCG, IED
	[64]	2022	eJeju Mall	CORR, RMSE, FD, MAE, mirror column association, PRD, accuracy, recall, F-score, MRR
	[69]	2023	Bookstores	Precision, MAP
VAE	[70]	2020	Yelp, Epinions, MovieLens, Netflix	Recall, NDCG
	[74]	2020	Amazon, MovieLens	MAP, NDCG
	[76]	2021	Taobao	Online A/B
	[77]	2021	Amazon, Epinions, MovieLens	Recall, NDCG
	[80]	2022	Business Platform	Accuracy, precision, recall, F-score
	[81]	2022	Amazon, MovieLens, Gowalla, Netflix	Recall, NDCG
	[83]	2022	Amazon, MovieLens, citeulike-a	Recall, NDCG
	[85]	2023	Amazon, Gowalla, Yelp	Recall, NDCG
	[86]	2024	Amazon	MAE, RMSE
	[87]	2024	Kaggle e-commerce	MSE, MAE, RMSE, MAPE, R^2
	[88]	2024	Amazon	HR, NDCG
Transformers	[91]	2021	Taobao	F-score
	[92]	2022	E-commerce platform	Accuracy, MAP, precision
	[93]	2022	Amazon	ROUGE F1, ES, C_{per}, $uDist$, $uPPL$
	[97]	2022	Amazon, Yelp	RMSE, MAE, HR, NDCG
	[99]	2023	Amazon	Recall, NDCG
	[100]	2023	Alipay	HR, NDCG
	[102]	2023	Yelp, Ifashion, LastFM	Recall, NDCG
	[105]	2024	Amazon	HR, NDCG
	[106]	2024	Amazon	HR, NDCG
	[111]	2025	E-commerce platform	HR, NDCG, MRR
	[115]	2023	Amazon, Yelp, MovieLens	Recall, NDCG

Diffusion	[117]	2023	Amazon, Steam, Yelp, MovieLens	HR, NDCG, MRR
	[118]	2023	Amazon, Steam, MovieLens	HR, NDCG
	[119]	2024	Amazon, MovieLens	HR, NDCG
	[120]	2024	Amazon, TikTok	Recall, precision, NDCG
Hybrids	[122]	2021	Amazon, Epinions, MovieLens	RMSE, MAE, precision, recall, F-score
	[123]	2022	Amazon	AUC, Online A/B
	[124]	2024	E-commerce platform	Accuracy, recall, F-score, AUC

4.2 Papers on GAN-based

The GAN-based papers used in this review are more varied and unique when compared to other core frameworks. It can be seen in the datasets used, metrics employed, advantages and disadvantages and so on, there is no specific pattern that can be picked up. Contrary to the overall trend, only a few authors chose to use Amazon as the dataset of choice, preferring for more obscure or proprietary datasets. As is the case for evaluation metrics used, precision and NDCG stand as the prevalent metrics, veering from the general trend.

In terms of benefits provided, each of the models proposed surpassed previous iterations and versions of recommenders in terms of performance, along with other added features added such as fairness determination, acceptance of categorical and numerical data and so on. On the contrary, similar drawbacks such as computational efficiency, mode collapse, and quality of data generated can be found in most papers. Table 2 provides a comprehensive list of pros and cons from the papers.

Table 2. Detailed Overview of GAN-based Papers

Paper	Model	Pros	Cons/Research Gaps
[54]	Conditional, coupled GAN	Superior to baselines.	Large time and computational cost, no ranking of generated results.
[55]	Positive-unlabelled risk minimiser	Superior to baselines.	Susceptible to converging failure, generating with discrete sampling with sparse data will cause poor model expressiveness.
[57]	Multi-modal Adversarial Representation Network	Superior to baselines.	*Not available.*
[59]	Effective perturbation mechanism (adversarial noise layer) for CNNs	Superior to baselines.	*Not available.*
[60]	Fairness generation	Superior to baselines, fairness towards items ensures equal recommendation.	Requires fairness distribution to all users, lacking simultaneous fairness rating in both item and users.
[64]	Conditional GAN, Wasserstein GAN with gradient penalty, PacGAN hybrid.	Superior to baselines, includes auxiliary classifier, which is great in generating synthetic data, excels in imbalanced data.	High-dimensional data and mode collapse diminishes output quality and diversity [68].
[69]	Double-layer iteration mechanism	Superior to baselines, alleviated the cold-start issue.	Requires fusing data from two domains to analyse multi-source recommendation scenarios.

4.3 Papers on VAE-Based

The VAE-based papers consist of the majority of the 36 papers chosen in this review, with 11 papers making up this category, or over 30% of the overall composition, suggesting the viability of VAEs in e-commerce recommenders. The majority of papers feature Amazon in their experimentations, with seven papers in total. Even though it is outside the subject topic, MovieLens is also prevalent in this collection, with four papers using it as datasets. The recall and NDCG pair of evaluation metrics are also featured many times in this collection, with five papers using this pairing exclusively. NDCG by itself is also used in most cases, with seven appearances. The benefits and drawbacks of the 11 papers are noted by all authors, with the main benefits being superior over selected baselines, better computational and time efficiency, and excelling in sparse data, among others. Some authors also combine alternate features within their proposals, such as adding GNNs in tandem with VAEs, LLMs to aid in encoding data, Beta-Bernoulli components to capture preference diversity and so on. Some gaps can also be identified, such as the KL vanishing problem, posterior collapse and so on. Table 3 outlines the underlying methods, advantages, and disadvantages in detail.

Table 3. Detailed Overview of VAE-based Papers

Paper	Model	Pros	Cons/Research Gaps
[70]	Wasserstein autoencoder with non-parametric Mixture Gaussian distribution.	Improvement over baselines especially in highly sparse data (69.6% and 78.4% near-cold-start users in two datasets).	Lacking inter-user and intra-user preference diversity; generalization ability and convergence property were insufficient [73].
[74]	Sub-recommenders' rating matrices values were adjusted, then merged using a probabilistic aggregation function and given to the VAE.	Performs well over baselines despite a simple aggregation function.	It requires better aggregation functions such as Bayesian approaches; context-aware recommender systems can be implemented.
[76]	A variant of the PID controller that alters the diversity/accuracy trade-off, then inserted into CVAE.	Improvement over real-world baseline due to ideal KL divergence.	*Not available.*
[77]	Bilaterally treating users and items similarly, with a CAP.	Superior to baselines, with CAP alleviating posterior collapse.	Applying to dyadic data, and co-clustering.
[80]	Adding latent VAE variables to churn prediction model.	Superior to baselines, designed to identify complex customer behaviour on a platform.	Viability limited if applied to single business model; Customers with longer tenure may not be captured.
[81]	Efficient sampling procedures for the approximate softmax distributions.	Superior to baselines, excellent time complexity and efficiency.	*Not available.*
[83]	Replacing randomly initialised last layer weights of the vanilla UAE with stacked latent item embedding.	Superior to baselines, typical issues encountered by UAEs such as sparsity and cold-starts are addressed.	*Not available.*
[85]	Masked graph autoencoder for reconstructing masked user-item subgraph structures.	Superior to baselines, excellent computation efficiency.	*Not available.*
[86]	Memory pools that simultaneously aggregates information and establishes similarities.	Superior to baselines, memory pool usage encourages simultaneous aggregation of information and establishment of similarity, efficient computations.	Poor in non-sparse data due to inadequately incorporating target information; only single-target CDR achieved instead of dual-target.
[87]	VAE-based equivariant GNN.	Superior to baselines.	*Not available.*
[88]	LLMs to encode information, RQ-VAE to train based on prior info.	Superior to baselines.	Exploring ways to further the current method in a multi-turn conversational setting.

4.4 Papers on Transformer-Based

The recent popularity of Transformer research surged when ChatGPT gained mainstream attention with the launch in 2022, with nine out of 10 papers published in 2022 and beyond. This also suggests Transformers could be the new norm in generative AI based applications, especially in e-commerce recommendations. Many of the selected papers feature multi-modal recommendations as well, a stark difference from other core frameworks, suggesting better flexibility in Transformer-based recommenders.

Similar phenomena to VAE-based papers can be seen, with Amazon datasets being the most used as well, with half of the papers using it in experiments. Similar evaluation metric usage can also be seen, with 70% of papers using NDCG, whereas half is using HR as metrics. However, Deng et al. [93] uses different metrics in their research, with metrics like C_{per}, $uDist$, $uPPL$ that are commonly used in PQA domains. Common benefits can be deduced from the research, including recommendation efficiency, cross-domain and multi-modal capabilities, extra degree of personalization and so on. Table 4 shows the detailed analysis of all 10 Transformer-based papers.

Table 4. Detailed Overview of Transformer-based Papers.

Paper	Model	Pros	Cons/Research Gaps
[91]	Three layers: modal-encoding, modal-interaction, modal-task multi-modal Transformer	Superior to baselines.	Further applying K3M in downstream tasks; Exploring performance on more general datasets.
[92]	Applying Transformer towards five modalities, using MRP, MLM, MEM, MFP, MAM	Superior to baselines.	Lacking capabilities of the modal representations; image and caption generation can be explored.
[93]	Combines a Transformer and the BiDAF [94] as the encoder-decoder architecture.	Superior to baselines with high diversity of user-centric information and user-preferred language styles.	Employing a multi-modal approach like implementing user-item interaction modelling could be beneficial.
[97]	Five task families: rating, sequential recommendation, explanation, review, and direct recommendation.	Superior to baselines with the ability of parameter-efficient tuning, with less training time and memory usage.	Further scaling up of the backbone model; incorporating more modalities, exploring better prompting strategies.
[99]	Semantic ID generation using content features, with Transformer model trained on Semantic ID.	Superior to baselines, efficiency in terms of memory cost compared to traditional recommender.	Possibility of invalid Semantic IDs due to clashing; high computation cost compared to Approximate Nearest Neighbours.
[100]	Pre-trained model based on LLM, includes a mask mechanism, span order, and positional encoding.	Superior to baselines, only losing out to ChatGPT and GPT-4.	Wider range of tasks might reveal the pros and cons, particularly on a more diverse set of domains.
[102]	Graph Transformer.	Superior to baselines, robust against artificial noise and data sparse condition	Adapting to different recommendation scenarios, like social-aware recommendations as well as knowledge graph-enhanced recommenders.
[105]	Prompt-based RLM.	Superior to baselines.	Reinforcement learning and LLM for automated prompt generation.
[106]	Tuned T5 with recommendation-oriented instruction data.	Superior to baselines.	Multi-turn interaction scenario, i.e. conversational; Challenging to directly model long user behavioural sequences due to limited context length.
[111]	Neural style transfer algorithm that is utilised as embeddings.	Superior to baselines.	To assess the performance of the model for scalable tasks, as well as more complex recommendation scenarios.

4.5 Diffusion-Based Papers

Only five diffusion-based papers were aggregated in this review. This is due to diffusion models being less suitable in recommender applications. Rather, they excel in image and video generation, as some researchers have noted [117]. All papers were published in 2023 and beyond, showing the relevant field has yet to mature.

All five diffusion-based papers show great consistency, in terms of datasets used and metrics employed. All papers used datasets from Amazon in their experimentation, with MovieLens as a supplementary dataset in three papers, similar to previous examples. All papers feature NDCG as a core metric, with HR in three papers as well, like the overall trend of this review. Similar benefits can be found throughout the research, with most of them noting the model's performance in noisy, sparse, and convoluted conditions.

Some researchers paired the diffusion model with other generative models such as Transformers and DNNs, suggesting diffusion as a secondary tool in a larger recommender system. Table 5 discusses the papers in detail.

Table 5. Detailed Overview of Diffusion-based Papers

Paper	Model	Pros	Cons/Research Gaps
[115]	L-DiffRec combined with T-DiffRec.	Great in clean and noisy datasets, better computation, and memory cost.	Needs better model compression and temporal information encoding, needs controllable and conditional recommendations.
[117]	Diffusion model combined with curriculum scheduler.	Curriculum scheduling effectively augment sequential behaviours, model can be combined with other base frameworks.	Diffusion mostly used in image generation that cannot easily be applied in recommendations.
[118]	Diffusion model paired with DNN as a Approximator	Superior against baselines.	*Not available.*
[119]	Adding offset noise in the diffusion flow and incorporating a cross-attention mechanism in the Transformer-based Approximator.	Superior against baselines, better convergence training times.	Shorter dataset sequences may lack sufficient information predict user preferences, while long sequences cause challenges for model performance.
[120]	Multi-modal graph diffusion model	Superior against baselines.	Integrating LLMs to guide the diffusion process is worth exploring.

4.6 Papers on Hybrid-Based

Lastly, hybrid recommenders have the least occurrence in this review, with only three papers in total, which may suggest difficulty and complexity in implementing multiple core frameworks to coordinate due to varying backgrounds. Two GAN-VAE hybrids and one VAE-Transformer hybrid is included. Each paper is published in different years, suggesting a sparse interest in hybrid recommender research.

Amazon is used as datasets in both GAN-VAE hybrids, with the latter having a general e-commerce dataset. Only three occurrences of AUC appear in this review, and two of them is used in hybrid-based recommenders. The other prevalent metrics that is used is recall and F-score. Table 6 shows a detailed analysis of the selected papers.

4.7 Summary

In summary, each core framework gives separate pros and cons, and no single framework provides an all-in-one solution to every requirement and use case. VAEs is the most prevalent model used in this review, with 11 papers in total. NDCG is the most common evaluation metric used, with 22 total appearances. Common datasets include Amazon in the e-commerce domain and MovieLens for cross-domain recommendation, with 21 and 12 instances used.

Table 6. Detailed Overview of Hybrid-based Papers

Paper	Model	Pros	Cons/Research Gaps
[122]	GAN-VAE hybrid	Performs well in highly sparse data with great knowledge transfer efficiency with superior prediction results.	Cannot handle dynamic demands due to static rating profiles, requires temporal and contextual information.
[123]	GAN-VAE hybrid, includes DSN paired with MIN	Superior to baselines.	*Not available.*
[124]	Hybrid of VAEs, Transformers, and attention mechanism.	Superior to baselines, with the least inference time and training time.	Intricate shopping conditions; In terms of performance variations and overall model stability.

Common characteristics pertaining to GAN models include alleviating cold-start issues, data sparsity, and superior recommendation quality. However, mode collapse, converging failure, and high computation cost remain core issues. VAEs boast simpler computations while having exceptional performance against sparse and cold-start datasets. Yet, KL vanishing issues and posterior collapse pose as core concerns for VAE models.

Transformers benefit cross-domain and multi-modal recommendations, while boasting great computational efficiency, but the fresh nature of Transformer research shows that it has yet to mature, with issues like prompting, adapting to other recommenders and so on. Diffusion models are deemed unsuitable in recommender applications due to its high computation cost and better performance in image and video generation instead of recommender systems. However, due to its generation quality, it performs ideally in sparse, noisy, and complicated conditions. Pairing diffusion models with other models could prove beneficial. Lastly, hybrid applications are less researched on, as models with diverse backgrounds might not scale well, impacting stability as a whole. Despite that, the reviewed papers note the performance achieved, especially in sparse data, doing so with commendable efficiency.

5. CONCLUSION

To conclude, this review provided deeper insight into the current state of recommender systems research. Generative AI emerges as a new avenue which provides a user-centric and personalised recommendations compared to traditional recommenders like CF and CB filtering. VAEs has presented itself as the more predominant framework in recommenders, with Transformers being a close second. Findings show that GANs is also a viable framework in recommenders, especially pertaining to high quality generations, but is typically hindered by high computation costs and overall model stability. VAEs provide better stability and efficiency than GANs, but the lower quality of generations might prove costly to some business applications. Transformers stood out by taking the best qualities of GANs and VAEs, boasting superior quality generations with minimal cost and robust model performance, not to mention better flexibility in inputs such as cross-domain and multi-modal recommendations. Despite that, more research needs to be done in order to fully expand the potential of Transformers, particularly in prompt engineering and contextual applications. Diffusion models can be alternatives in generative AI recommenders, especially in image and video applications, rather than product recommendations on e-commerce platforms.

Future research is required to apply diffusion models in recommender applications and ways to reduce costs in training. Hybrid models show immense potential in future research, combining benefits from all frameworks. However, the complex implementation due to varying framework backgrounds might deter future researchers, but prevailing research shows the underlying potential of hybrid models, suggesting a bright future in this research domain. In summary, this domain of recommender research has yet to reach its fullest potential, especially with research on novel and new domains like Transformers and diffusion models. Older discoveries such as GANs and VAEs still have space to mature, with promising research avenues on hybrid implementations as well. Continuing research and innovation in this field is critical to expand the potentials of these frameworks in user interaction and personalization, as well as efficient qualitative recommendations which aims to better societal and commercial interests.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for the suggestions to improve the paper.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Kai-Ze Liau: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation; Heru Agus Santoso: Project Administration, Writing – Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/.

REFERENCES

[1] F. Ricci, L. Rokach, and B. Shapira, "Recommender systems: Techniques, applications, and challenges," in *Recommender Systems Handbook*, New York, NY: Springer US, pp. 1–35, 2022, doi: 10.1007/978-1-0716-2197-4_1.

[2] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," *Journal of Big Data*, vol. 9, no. 1, pp. 59, Dec. 2022, doi: 10.1186/s40537-022-00592-5.

[3] T. Di Noia *et al.*, Eds., "RecSys '24: Proceedings of the 18th ACM Conference on Recommender Systems," in *18th ACM Conference on Recommender Systems*, New York, NY, USA: ACM, Oct. 2024.

[4] F. O. Momoh, S. Rakshit, and N. R. Vajjhala, "Exploratory study of machine learning algorithms in recommender systems," in *Advances in Intelligent Systems and Computing*, vol. 1406. Singapore: Springer, pp. 639–650, 2022, doi: 10.1007/978-981-16-5207-3_48.

[5] H. Abdollahpouri *et al.*, "Multistakeholder recommendation: Survey and research directions," *User Model User-adapt Interact*, vol. 30, no. 1, 2020, doi: 10.1007/s11257-019-09256-1.

[6] C. Gao, W. Lei, X. He, M. de Rijke, and T. S. Chua, "Advances and challenges in conversational recommender systems: A survey," *Artificial Intelligence Open*, vol. 2, pp. 100–126, 2021, doi: 10.1016/j.aiopen.2021.06.002.

[7] A. Gunawardana, G. Shani, and S. Yogev, "Evaluating recommender systems," in *Recommender Systems Handbook*, New York, NY: Springer US, pp. 547–601, 2022, doi: 10.1007/978-1-0716-2197-4_15.

[8] E. Zangerle and C. Bauer, "Evaluating recommender systems: Survey and framework," *ACM Computing Surveys*, vol. 55, no. 8, 2022, doi: 10.1145/3556536.

[9] S. Natarajan, S. Vairavasundaram, S. Natarajan, and A. H. Gandomi, "Resolving data sparsity and cold start problem in collaborative filtering recommender system using Linked Open Data," *Expert Systems with Applications*, vol. 149, 2020, doi: 10.1016/j.eswa.2020.113248.

[10] V. Jain, B. Malviya, and S. Arya, “An overview of electronic commerce (e-Commerce),” *Journal of Contemporary Issues in Business and Government*, vol. 27, no. 3, Apr. 2021, doi: 10.47750/cibg.2021.27.03.090.

[11] W. Song and C. Wang, “Hybrid recommendation based on matrix factorization and deep learning,” in *Proceedings of the 2022 ACM International Conference on Intelligent Computing and Its Applications*, 2022, doi: 10.1145/3538950.3538961.

[12] B. T. Imani and E. B. Setiawan, “Recommender system based on matrix factorization on Twitter using random forest (Case study: Movies on Netflix),” *International Journal on Information and Communication Technology (IJoICT)*, vol. 8, no. 2, pp. 11–21, Dec. 2022, doi: 10.21108/ijoict.v8i2.655.

[13] W.-E. Kong, T.-E. Tai, P. Naveen, and H. A. Santoso, “Performance evaluation on E-commerce recommender system based on KNN, SVD, CoClustering and ensemble approaches,” *Journal of Informatics and Web Engineering*, vol. 3, no. 3, pp. 63–76, Oct. 2024, doi: 10.33093/jiwe.2024.3.3.4.

[14] I. Karabila, N. Darraz, A. El-Ansari, N. Alami, and M. El Mallahi, “Enhancing collaborative filtering-based recommender system using sentiment analysis,” *Future Internet*, vol. 15, no. 7, 2023, doi: 10.3390/fi15070235.

[15] K. Deepa, D. Balamurugan, M. Bharath, S. Dineshkumar, and C. Vignesh, “Recommendation system using deep learning with text mining algorithm in e-commerce framework,” in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 141, 2023, doi: 10.1007/978-981-19-3035-5_27.

[16] H. Steck, L. Baltrunas, E. Elahi, D. Liang, Y. Raimond, and J. Basilico, “Deep learning for recommender systems: A Netflix case study,” *AI Magazine*, vol. 42, no. 3, 2021, doi: 10.1609/aimag.v42i3.18140.

[17] M. Sharaf, E. E. D. Hemdan, A. El-Sayed, and N. A. El-Bahnasawy, “A survey on recommendation systems for financial services,” *Multimedia Tools and Applications*, vol. 81, no. 12, 2022, doi: 10.1007/s11042-022-12564-1.

[18] Z. Shokrzadeh, M. R. Feizi-Derakhshi, M. A. Balafar, and J. Bagherzadeh Mohasefi, “Knowledge graph-based recommendation system enhanced by neural collaborative filtering and knowledge graph embedding,” *Ain Shams Engineering Journal*, vol. 15, no. 1, 2024, doi: 10.1016/j.asej.2023.102263.

[19] J.-C. Zhang, A. M. Zain, K.-Q. Zhou, X. Chen, and R.-M. Zhang, “A review of recommender systems based on knowledge graph embedding,” *Expert Systems with Applications*, vol. 250, pp. 123876, Sep. 2024, doi: 10.1016/j.eswa.2024.123876.

[20] Q. Guo *et al.*, “A survey on knowledge graph-based recommender systems,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 8, pp. 3549–3568, 2022, doi: 10.1109/TKDE.2020.3028705.

[21] J. Kwon, S. Ahn, and Y.-D. Seo, “RecKG: Knowledge graph for recommender systems,” in *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, New York, NY, USA: ACM, pp. 600–607, Apr. 2024, doi: 10.1145/3605098.3636009.

[22] Y. Deldjoo, T. Di Noia, and F. A. Merra, “A survey on adversarial recommender systems: From Attack/Defense Strategies to Generative Adversarial Networks,” *ACM Computing Surveys*, vol. 54, no. 2, pp. 1–38, Mar. 2021, doi: 10.1145/3439729.

[23] I. J. Goodfellow *et al.*, “Generative adversarial nets,” in *Advances in Neural Information Processing Systems*, 2014, doi: 10.1007/978-3-658-40442-0_9.

[24] R. Bhattacharyya *et al.*, “Modeling human driving behavior through generative adversarial imitation learning,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 3, 2023, doi: 10.1109/TITS.2022.3227738.

[25] X. Yang, Z. Song, I. King, and Z. Xu, “A survey on deep semi-supervised learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 9, 2023, doi: 10.1109/TKDE.2022.3220219.

[26] A. Haque, “EC-GAN: Low-sample classification using semi-supervised algorithms and GANs (Student abstract),” in *35th AAAI Conference on Artificial Intelligence, AAAI 2021*, 2021, doi: 10.1609/aaai.v35i18.17895.

[27] A. Borji, “Pros and cons of GAN evaluation measures: New developments,” *Computer Vision and Image Understanding*, vol. 215, 2022, doi: 10.1016/j.cviu.2021.103329.

[28] D. Bank, N. Koenigstein, and R. Giryes, “Autoencoders,” in *Machine Learning for Data Science Handbook*, Cham: Springer International Publishing, pp. 353–374, 2023, doi: 10.1007/978-3-031-24628-9_16.

[29] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” in *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, 2014, doi: 10.61603/ceas.v2i1.33.

[30] A. Asperti and M. Trentin, “Balancing reconstruction error and kullback-leibler divergence in variational autoencoders,” *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3034828.

[31] M. Rivera, “How to train your VAE,” in *2024 IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 3882–3888, Oct. 2024, doi: 10.1109/ICIP51287.2024.10647557.

[32] R. Wei, C. Garcia, A. El-Sayed, V. Peterson, and A. Mahmood, “Variations in variational autoencoders - A comparative evaluation,” *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3018151.

[33] S. M. Park, H. G. Yoon, D. B. Lee, J. W. Choi, H. Y. Kwon, and C. Won, “Topological magnetic structure generation using VAE-GAN hybrid model and discriminator-driven latent sampling,” *Scientific Reports*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-47866-3.

[34] Z. Niu, K. Yu, and X. Wu, “LSTM-based vae-gan for time-series anomaly detection,” *Sensors (Switzerland)*, vol. 20, no. 13, 2020, doi: 10.3390/s20133738.

[35] Y. Li, K. Zhou, W. X. Zhao, and J. R. Wen, “Diffusion models for non-autoregressive text generation: A survey,” in *IJCAI International Joint Conference on Artificial Intelligence*, 2023, doi: 10.24963/ijcai.2023/750.

[36] X. Han, S. Kumar, and Y. Tsvetkov, “SSD-LM: Semi-autoregressive simplex-based diffusion language model for text generation and modular control,” in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2023, doi: 10.18653/v1/2023.acl-long.647.

[37] W. Xu, W. Hu, F. Wu, and S. H. Sengamedu, “DeTiME: Diffusion-enhanced topic modeling using encoder-decoder based LLM,” in *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2023, doi: 10.18653/v1/2023.findings-emnlp.606.

[38] H. Zhang, X. Liu, and J. Zhang, “DiffuSum: Generation enhanced extractive summarization with diffusion,” in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2023, doi: 10.18653/v1/2023.findings-acl.828.

[39] C. Chi et al., “Diffusion policy: Visuomotor policy learning via action diffusion,” in *Robotics: Science and Systems (RSS 2023)*. Daegu, South Korea: RSS, 2023, doi: 10.15607/rss.2023.xix.026.

[40] M. Janner, Y. Du, J. B. Tenenbaum, and S. Levine, “Planning with diffusion for flexible behavior synthesis,” in *Proceedings of Machine Learning Research*, 2022.

[41] D. Yang *et al.*, “Diffsound: Discrete diffusion model for text-to-sound generation,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, 2023, doi: 10.1109/TASLP.2023.3268730.

[42] Z. Chang, G. A. Koulieris, and H. P. H. Shum, “On the design fundamentals of diffusion models: A survey,” Jun. 2023.

[43] C. Saharia *et al.*, “Photorealistic text-to-image diffusion models with deep language understanding,” in *Advances in Neural Information Processing Systems*, 2022.

[44] C. Zhou *et al.*, “Transfusion: Predict the next token and diffuse images with one multi-modal model,” Aug. 2024.

[45] A. Vaswani *et al.*, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, 2017.

[46] A. Dosovitskiy *et al.*, “An image is worth 16x16 words: transformers for image recognition at scale,” in *ICLR 2021 - 9th International Conference on Learning Representations*, 2021.

[47] A. Gulati *et al.*, “Conformer: Convolution-augmented transformer for speech recognition,” in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2020, doi: 10.21437/Interspeech.2020-3015.

[48] L. Chen *et al.*, “Decision transformer: Reinforcement learning via sequence modeling,” in *Advances in Neural Information Processing Systems*, 2021.

[49] K. Choromanski *et al.*, “Rethinking attention with performers,” in *ICLR 2021 - 9th International Conference on Learning Representations*, 2021.

[50] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, “Hierarchical text-conditional image generation with CLIP latents,” Apr. 2022.

[51] P. Esser *et al.*, “Scaling rectified flow transformers for high-resolution image synthesis,” Mar. 2024.

[52] T. Brooks *et al.*, “Video generation models as world simulators,” 2024, [Online]. Available: https://openai.com/research/video-generation-models-as-world-simulators

[53] P. Esser *et al.*, “Scaling rectified flow transformers for high-resolution image synthesis,” Mar. 2024.

[54] J. R. Bock and A. Maewal, “Adversarial learning for product recommendation,” *AI (Switzerland)*, vol. 1, no. 3, Sep. 2020, doi: 10.3390/ai1030025.

[55] Y. Zhou *et al.*, “GAN-based recommendation with positive-unlabeled sampling,” Dec. 2020, *arXiv preprint* arXiv:2012.06901, Dec. 2020, doi: 10.48550/arXiv.2012.06901.

[56] R. Kiryo, G. Niu, M. C. Du Plessis, and M. Sugiyama, “Positive-unlabeled learning with non-negative risk estimator,” in *Advances in Neural Information Processing Systems*, 2017.

[57] X. Li, C. Wang, J. Tan, X. Zeng, D. Ou, and B. Zheng, “Adversarial multimodal representation learning for click-through rate prediction,” in *The Web Conference 2020 - Proceedings of the World Wide Web Conference, WWW 2020*, 2020, doi: 10.1145/3366423.3380163.

[58] M. R. Abdurrahman, H. Al-Aziz, F. A. Zayn, M. A. Purnomo, and H. A. Santoso, “Development of robot feature for stunting analysis using Long-Short Term Memory (LSTM) algorithm,” *Journal of Informatics and Web Engineering*, vol. 3, no. 3, pp. 164–175, Oct. 2024, doi: 10.33093/jiwe.2024.3.3.10.

[59] F. Yuan, L. Yao, and B. Benatallah, “Exploring missing interactions: A convolutional generative adversarial network for collaborative filtering,” in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, New York, NY, USA: ACM, pp. 1773–1782, Oct. 2020, doi: 10.1145/3340531.3411917.

[60] J. Li, Y. Ren, and K. Deng, “FairGAN: GANs-based fairness-aware learning for recommendations with implicit feedback,” in *WWW 2022 - Proceedings of the ACM Web Conference 2022*, Association for Computing Machinery, Inc, pp. 297–307, Apr. 2022, doi: 10.1145/3485447.3511958.

[61] Y. Ge *et al.*, “Towards long-term fairness in recommendation,” in *WSDM 2021 - Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 2021, doi: 10.1145/3437963.3441824.

[62] T. Yang and Q. Ai, “Maximizing marginal fairness for dynamic learning to rank,” in *The Web Conference 2021 - Proceedings of the World Wide Web Conference, WWW 2021*, 2021, doi: 10.1145/3442381.3449901.

[63] Y. Hou, J. Li, Z. He, A. Yan, X. Chen, and J. McAuley, “Bridging language and items for retrieval and recommendation,” Mar. 2024.

[64] W. Shafqat and Y. C. Byun, "A hybrid GAN-based approach to solve imbalanced data problem in recommendation systems," *IEEE Access*, vol. 10, pp. 11036–11047, 2022, doi: 10.1109/ACCESS.2022.3141776.

[65] K. E. Smith and A. O. Smith, "Conditional GAN for timeseries generation," arXiv preprint arXiv:2006.16477, Jun. 2020, doi: 10.48550/arXiv.2006.16477.

[66] A. Koivu, M. Sairanen, A. Airola, and T. Pahikkala, "Synthetic minority oversampling of vital statistics data with generative adversarial networks," *Journal of the American Medical Informatics Association*, vol. 27, no. 11, 2020, doi: 10.1093/jamia/ocaa127.

[67] G. Douzas and F. Bacao, "Effective data generation for imbalanced learning using conditional generative adversarial networks," *Expert Systems with Applications*, vol. 91, 2018, doi: 10.1016/j.eswa.2017.09.030.

[68] J. REN *et al.*, "Multidimensional tensor-aware GAN based pseudo measurement data deduction in IoT-empowered distribution station," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, pp. 2024EAP1062, 2024, doi: 10.1587/transfun.2024EAP1062.

[69] S. Wei, X. Zhou, X. An, X. Yang, and Y. Xiao, "A heterogeneous e-commerce user alignment model based on data enhancement and data representation," *Expert Systems with Applications*, vol. 228, 2023, doi: 10.1016/j.eswa.2023.120258.

[70] H. Liu *et al.*, "Deep global and local generative model for recommendation," in *The Web Conference 2020 - Proceedings of the World Wide Web Conference, WWW 2020*, Association for Computing Machinery, Inc, pp. 551–561, Apr. 2020, doi: 10.1145/3366423.3380138.

[71] X. Zhang, J. Zhong, and K. Liu, "Wasserstein autoencoders for collaborative filtering," *Neural Computing and Applications*, vol. 33, no. 7, 2021, doi: 10.1007/s00521-020-05117-w.

[72] J. Omana, P. N. Jeipratha, K. Devi, S. Benila, and K. Revathi, "Personalized drug recommendation system using Wasserstein auto-encoders and adverse drug reaction detection with Weighted Feed Forward Neural Network (WAES-ADR) in Healthcare," *Journal of Informatics and Web Engineering*, vol. 4, no. 1, pp. 332–347, Feb. 2025, doi: 10.33093/jiwe.2025.4.1.24.

[73] H. Liu *et al.*, "Interpretable deep generative recommendation models," *Journal of Machine Learning Research*, vol. 22, pp. 1–54, 2021.

[74] A. Drif, H. E. Zerrad, and H. Cherifi, "Ensvae: Ensemble variational autoencoders for recommendations," *IEEE Access*, vol. 8, pp. 188335–188351, 2020, doi: 10.1109/ACCESS.2020.3030693.

[75] F. M. Harper and J. A. Konstan, "The MovieLens datasets," *ACM Transactions on Interactive Intelligent Systems*, vol. 5, no. 4, pp. 1–19, Jan. 2016, doi: 10.1145/2827872.

[76] H. Shao *et al.*, "Controllable and diverse text generation in e-commerce," in *The Web Conference 2021 - Proceedings of the World Wide Web Conference, WWW 2021*, 2021, doi: 10.1145/3442381.3449838.

[77] Q. T. Truong, A. Salah, and H. W. Lauw, "Bilateral variational autoencoder for collaborative filtering," in *WSDM 2021 - Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 2021, doi: 10.1145/3437963.3441759.

[78] Y. Wang, D. M. Blei, and J. P. Cunningham, "Posterior collapse and latent variable non-identifiability," in *Advances in Neural Information Processing Systems*, 2021.

[79] E. Battaglia, F. Peiretti, and R. G. Pensa, "Co-clustering: A survey of the main methods, recent trends, and open problems," *ACM Computing Surveys*, vol. 57, no. 2, pp. 1–33, Feb. 2025, doi: 10.1145/3698875.

[80] K. Hasumoto and M. Goto, "Predicting customer churn for platform businesses: using latent variables of variational autoencoder as consumers' purchasing behavior," *Neural Computing and Applications*, vol. 34, no. 21, 2022, doi: 10.1007/s00521-022-07418-8.

[81] J. Chen, D. Lian, B. Jin, X. Huang, K. Zheng, and E. Chen, "Fast variational autoencoder with inverted multi-index for collaborative filtering," in *WWW 2022 - Proceedings of the ACM Web Conference 2022*, 2022, doi: 10.1145/3485447.3512068.

[82] I. Goodfellow, Y. Bengio, and A. Courville, "6.2.2.3 Softmax units for multinoulli output distributions," in *Deep Learning*, MIT Press, ch. 6.2.2.3, pp. 180–184, 2016.

[83] Y. Zhu and Z. Chen, "Mutually-regularized dual collaborative variational auto-encoder for recommendation systems," in *WWW 2022 - Proceedings of the ACM Web Conference 2022*, 2022, doi: 10.1145/3485447.3512110.

[84] V. Stergiopoulos, M. Vassilakopoulos, E. Tousidou, and A. Corral, "Hyper-parameters tuning of artificial neural networks: An application in the field of recommender systems," in *Communications in Computer and Information Science*, 2022, doi: 10.1007/978-3-031-15743-1_25.

[85] L. Xia, C. Huang, C. Huang, K. Lin, T. Yu, and B. Kao, "Automated self-supervised learning for recommendation," in *ACM Web Conference 2023 - Proceedings of the World Wide Web Conference, WWW 2023*, 2023, doi: 10.1145/3543507.3583336.

[86] J. Yang, J. Zhu, X. Ding, Y. Peng, and Y. Zhang, "A memory pool variational autoencoder framework for cross-domain recommendation," *Expert Systems with Applications*, vol. 241, May 2024, doi: 10.1016/j.eswa.2023.122771.

[87] M. Gandhudi, A. P.J.A., V. Velayudham, L. Nagineni, and G. G.R., "Explainable causal variational autoencoders based equivariant graph neural networks for analyzing the consumer purchase behavior in E-commerce," *Engineering Applications of Artificial Intelligence*, vol. 136, pp. 108988, Oct. 2024, doi: 10.1016/j.engappai.2024.108988.

[88] B. Zheng *et al.*, "Adapting large language models by integrating collaborative semantics for recommendation," in *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, IEEE, pp. 1435–1448, May 2024, doi: 10.1109/ICDE60146.2024.00118.

[89] H. Touvron *et al.*, "LLaMA: Open and efficient foundation language models," Feb. 2023.

[90] Y. Hou, S. Mu, W. X. Zhao, Y. Li, B. Ding, and J. R. Wen, "Towards universal sequence representation learning for recommender systems," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2022, doi: 10.1145/3534678.3539381.

[91] Y. Zhu *et al.*, "Knowledge perceived multi-modal pretraining in e-commerce," in *MM 2021 - Proceedings of the 29th ACM International Conference on Multimedia*, 2021, doi: 10.1145/3474085.3475648.

[92] X. Dong *et al.*, "M5Product: Self-harmonized contrastive learning for e-commercial multi-modal pretraining," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2022, doi: 10.1109/CVPR52688.2022.02057.

[93] Y. Deng, Y. Li, W. Zhang, B. Ding, and W. Lam, "Toward personalized answer generation in e-commerce via multi-perspective preference modeling," *ACM Transactions on Information Systems*, vol. 40, no. 4, pp. 1–28, Oct. 2022, doi: 10.1145/3507782.

[94] Y. Yang, "BiEAF: An bidirectional enhanced attention flow model for question answering task," in *2021 2nd International Conference on Information Science and Education (ICISE-IE)*, IEEE, pp. 344–348, Nov. 2021, doi: 10.1109/ICISE-IE53922.2021.00086.

[95] T. C. Luo and J. T. Chien, "Variational dialogue generation with normalizing flows," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2021, doi: 10.1109/ICASSP39728.2021.9414586.

[96] B. Wu *et al.*, "Guiding variational response generator to exploit persona," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2020. doi: 10.18653/v1/2020.acl-main.7.

[97] S. Geng, S. Liu, Z. Fu, Y. Ge, and Y. Zhang, "Recommendation as Language Processing (RLP): A unified pretrain, personalized prompt & predict paradigm (P5)," in *RecSys 2022 - Proceedings of the 16th ACM Conference on Recommender Systems*, 2022, doi: 10.1145/3523227.3546767.

[98] S. Geng, J. Tan, S. Liu, Z. Fu, and Y. Zhang, "VIP5: Towards multimodal foundation models for recommendation," in *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2023, doi: 10.18653/v1/2023.findings-emnlp.644.

[99] S. Rajput *et al.*, "Recommender systems with generative retrieval," *Adv Neural Inf Process Syst*, vol. 36, pp. 10299–10315, 2023.

[100] Z. Chu et al., "Leveraging large language models for pre-trained recommender systems," *arXiv preprint* arXiv:2308.10837, Aug. 2023, doi: 10.48550/arXiv.2308.10837.

[101] Z. Du *et al.*, "GLM: General language model pretraining with autoregressive blank infilling," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2022, doi: 10.18653/v1/2022.acl-long.26.

[102] C. Li, L. Xia, X. Ren, Y. Ye, Y. Xu, and C. Huang, "Graph transformer for recommendation," in *SIGIR 2023 - Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, doi: 10.1145/3539618.3591723.

[103] Z. Zhang, B. Yu, X. Shu, M. Xue, T. Liu, and L. Guo, "From what to why: Improving relation extraction with rationale graph," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2021, doi: 10.18653/v1/2021.findings-acl.8.

[104] Y. X. Wu, X. Wang, A. Zhang, X. He, and T. S. Chua, "Discovering invariant rationales for graph neural networks," in *ICLR 2022 - 10th International Conference on Learning Representations*, 2022.

[105] Z. Li et al., "PAP-REC: Personalized automatic prompt for recommendation language model," *arXiv preprint* arXiv:2401.08735, Jan. 2024, doi: 10.48550/arXiv.2402.00284.

[106] J. Zhang, R. Xie, Y. Hou, X. Zhao, L. Lin, and J.-R. Wen, "Recommendation as Instruction Following: A large language model empowered recommendation approach," *ACM Transactions on Information Systems*, Dec. 2024, doi: 10.1145/3708882.

[107] H. W. Chung *et al.*, "Scaling instruction-finetuned language models," *Journal of Machine Learning Research*, vol. 25, no. 70, pp. 1–53, 2024, [Online]. Available: http://jmlr.org/papers/v25/23-0870.html

[108] C. Raffel *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, 2020.

[109] K. Zhou *et al.*, "S3-Rec: Self-supervised learning for sequential recommendation with mutual information maximization," in *International Conference on Information and Knowledge Management, Proceedings*, 2020, doi: 10.1145/3340531.3411954.

[110] W. X. Zhao *et al.*, "RecBole: Towards a unified, comprehensive and efficient framework for recommendation algorithms," in *International Conference on Information and Knowledge Management, Proceedings*, 2021. doi: 10.1145/3459637.3482016.zs

[111] B. Ugurlu, M.-Y. Hong, and C. Lin, "Style4Rec: Enhancing transformer-based e-commerce recommendation systems with style and shopping cart information," *arXiv preprint* arXiv:2501.01234, Jan. 2025.

[112] Q. Cai, M. Ma, C. Wang, and H. Li, "Image neural style transfer: A review," *Computers and Electrical Engineering*, vol. 108, 2023, doi: 10.1016/j.compeleceng.2023.108723.

[113] W. C. Kang and J. McAuley, "Self-attentive sequential recommendation," in *Proceedings - IEEE International Conference on Data Mining, ICDM*, 2018, doi: 10.1109/ICDM.2018.00035.

[114] F. Sun *et al.*, "Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer," in *International Conference on Information and Knowledge Management, Proceedings*, 2019, doi: 10.1145/3357384.3357895.

[115] W. Wang, Y. Xu, F. Feng, X. Lin, X. He, and T. S. Chua, "Diffusion recommender model," in *SIGIR 2023 - Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Association for Computing Machinery, Inc, pp. 832–841, Jul. 2023, doi: 10.1145/3539618.3591663.

[116] L. Yang *et al.*, "Diffusion models: A comprehensive survey of methods and applications," *ACM Computing Surveys*, vol. 56, no. 4, 2023, doi: 10.1145/3626235.

[117] Z. Wu *et al.*, "Diff4Rec: Sequential recommendation with curriculum-scheduled diffusion augmentation," in *MM 2023 - Proceedings of the 31st ACM International Conference on Multimedia*, Association for Computing Machinery, Inc, pp. 9329–9335, Oct. 2023, doi: 10.1145/3581783.3612709.

[118] Z. Li, A. Sun, and C. Li, "DiffuRec: A diffusion model for sequential recommendation," *ACM Transactions on Information Systems*, vol. 42, no. 3, 2023, doi: 10.1145/3631116.

[119] S. Zolghadr, O. Winther, and P. Jeha, "Generative diffusion models for sequential recommendations," *arXiv preprint* arXiv:2410.02584, Oct. 2024, doi: 10.48550/arXiv.2410.19429.

[120] Y. Jiang, L. Xia, W. Wei, D. Luo, K. Lin, and C. Huang, "DiffMM: Multi-modal diffusion model for recommendation," in *Proceedings of the 32nd ACM International Conference on Multimedia*, New York, NY, USA: ACM, pp. 7591–7599, Oct. 2024, doi: 10.1145/3664647.3681498.

[121] X. Zhou *et al.*, "Bootstrap latent representations for multi-modal recommendation," in *ACM Web Conference 2023 - Proceedings of the World Wide Web Conference, WWW 2023*, 2023, doi: 10.1145/3543507.3583251.

[122] Y. Li, J. Ren, J. Liu, and Y. Chang, "Deep sparse autoencoder prediction model based on adversarial learning for cross-domain recommendations," *Knowledge-Based Systems*, vol. 220, May 2021, doi: 10.1016/j.knosys.2021.106948.

[123] F. Xiao *et al.*, "From abstract to details: A generative multimodal fusion framework for recommendation," in *MM 2022 - Proceedings of the 30th ACM International Conference on Multimedia*, 2022, doi: 10.1145/3503161.3548366.

[124] Y. Liu, J. Hou, and W. Zhao, "Deep learning and user consumption trends classification and analysis based on shopping behavior," *Journal of Organizational and End User Computing*, vol. 36, no. 1, pp. 1–23, Mar. 2024, doi: 10.4018/JOEUC.340038.

[125] M. Aydogan and V. Kocaman, "TRSAv1: A new benchmark dataset for classifying user reviews on Turkish e-commerce websites," *Journal of Information Science*, vol. 49, no. 6, 2023, doi: 10.1177/01655515221074328.

BIOGRAPHIES OF AUTHORS

Kai-Ze Liau is a final year student taking a Bachelor of Computer Science, majoring in Data Science in Multimedia University, Malaysia. His projects involve applying data science concepts in business related domains, specifically in e-commerce applications. His research interests surround generative AI applications, machine learning, recommender systems and data science. He can be contacted at email: 1211102809@student.mmu.edu.my.

	**Heru Agus Santoso** is a faculty member at Dian Nuswantoro University in Semarang, Indonesia. His research interests primarily include ontology, information retrieval, text mining and deep learning. He has published several publications in these areas, showcasing contributions in data and knowledge systems. He can be contacted at email: heru.agus.santoso@dsn.dinus.ac.id.