

Survey on Serendipity in Recommender Systems

You

August 3, 2025

1 Introduction

Recommender Systems (RS) are specialized software tools that assist users in navigating vast amounts of information by suggesting items likely to match their interests [1]. Traditional RS prioritize accuracy, recommending items that closely match users’ historical profiles with which users are familiar. However, this approach may result in user boredom and dissatisfaction, often leading to ”filter bubble” or over-specialization problems [2]. Furthermore, RS often ignore unpopular or ”long-tail” items to improve accuracy [3]. Although accuracy is improved, the recommendations become limited to a narrow scope of items, which may negatively impact user satisfaction and long-term system performance [1], [2].

Therefore, many researchers have proposed the concept of serendipity to address the ”filter bubble” and the under-representation of long-tail items. Although there is currently no consensus on the definition of serendipity in RS, most researchers explain serendipity through three components: unexpectedness, novelty, and relevance, while some researchers consider diversity as an additional component of serendipity [1], [3]. Numerous studies have explored integrating these components into RS, including traditional algorithms such as content-based and collaborative filtering. In recent years, deep learning models have been introduced into RS due to their capacity and advantages in capturing deep relationships between users and items [3]. Architectures such as Multi-Layer Perceptrons, deep reinforcement learning, and Large Language Models are employed to learn users’ latent serendipity preferences.

Although traditional and deep learning methods excel at accuracy-based tasks [4], [5], several challenges remain: lack of consensus on serendipity definition, scarcity of available ground truth data, difficulty in representing serendipity in RS, and absence of well-established evaluation methods [3]. In the following sections, we discuss classic challenges of RS and outline the conventional RS pipeline.

2 Various Definitions of Serendipity in RS

Serendipity is generally defined as ”luck that takes the form of finding valuable or pleasant things that are not looked for” (Britannica Dictionary). In the field of RS, serendipity has emerged as a vital objective. It aims to enhance user satisfaction by providing recommendations that are not only accurate but also surprising and valuable [5], [6]. Traditional RS mainly prioritized precision and relevance, but researchers later recognized that fostering serendipity could help break ”filter bubbles” or over-specialization problems [5], [7], encourage user exploration, and increase user engagement. The earliest structured definition of serendipity in RS can be traced back to Herlocker *et al.* (2004), who characterized it as the interaction of surprise and relevance [8]. Over time, the conceptualization and modeling of serendipity in RS have become increasingly sophisticated, with researchers proposing diverse interpretations and no universally accepted definition. This section systematically reviews how serendipity has been defined in the RS literature, following a chronological trajectory.

2.1 Early Definition of Serendipity (2008-2015)

In the early stage of serendipity research in RSs, researchers started to notice that serendipity is more than just novelty. Most literature mentions that, by definition, serendipity is novel. During this period, there was generally no clear formulaic definition of what constitutes serendipity.

Iaquinta *et al.* (2008) defined serendipity as recommending items that had not been previously searched

by the user but were still useful [9]. This definition emphasizes novelty in the sense of unsearched content, along with usefulness, but does not yet account for surprise or unexpectedness. We can interpret their definition of serendipity essentially as unexpectedness (i.e., a superset of novelty) + usefulness (i.e., relevance).

Another definition was proposed by Ge *et al.* (2010): serendipity arises when an RS suggests items that are both unexpected and useful [10]. In their view, unexpectedness is defined as a surprising item that is not yet discovered and not expected by the user. Usefulness is concretely described as the item being interesting, relevant, and valuable to the user. Thus, we can formalize their definition of serendipity as unexpectedness + usefulness (i.e., relevance).

By 2012, this two-component formulation was subsequently adopted and applied by Lu *et al.* [11], who followed the definitions proposed by Iaquinta *et al.* and Ge *et al.*, treating them as equivalent [9]–[11]. In their paper, Lu *et al.* explicitly proposed evaluating serendipity through the dimensions of unexpectedness and usefulness [11].

A few years later, Manca *et al.* (2015) proposed that serendipitous recommendations should be both successful (i.e., of interest to the user) and unexpected (i.e., different from the user’s historical interests) [12]. Essentially, the notion of ”successful” in this context aligns with usefulness. As such, their definition of serendipity can also be abstracted as: unexpectedness + usefulness (i.e., interestingness) [12].

Across these early studies, two main definitional elements began to take shape:

- **Novelty/(Emerging) Unexpectedness:** The item has not been discovered or not expected by the user.
- **Usefulness:** The item must be of interest, relevant or valuable to the user.

However, these components were not yet formally integrated or clearly differentiated from one another.

2.2 Divergence and Enrichment of the Definition (2016-2020)

Between 2016 and 2020, the conceptual understanding of serendipity in RSs evolved significantly. Researchers increasingly began to differentiate serendipity from novelty, and definitions shifted from two-component models (unexpectedness + usefulness) toward three-component or even multidimensional frameworks.

A key milestone came from a survey by Kaminskas *et al.* (2017), who formalized serendipity as the intersection of surprise and relevance, synthesizing earlier perspectives [13]. At same time, a comprehensive survey by Kotkov *et al.* systematically examined serendipity-related beyond-accuracy objectives [14]. They proposed a tripartite model for serendipity, formulated as $serendipity = Relevance + Novelty + Unexpectedness$ emphasizing recommendations that are useful, previously unseen, and surprising to users. Following this, many studies began to build on this three-part definition, by adding, removing, or replacing components. However, notable exceptions existed. Paiva *et al.* (2017) defined serendipity as chance and sagacity. Afridi (2018) focused on surprising recommendations not discovered by users themselves, implicitly relating to novelty and usefulness. Yang *et al.* (2018) and Xu *et al.* (2020) defined it via low interest and high satisfaction. Nishioka *et al.* (2019) combined novelty and positive surprise, while Cerri *et al.* (2020) defined it as unintentional discovery with positive value.

2.3 Definition Maturation and (Lack of) Stabilization (2021-2025)

During this period, the definition of serendipity in RS research underwent maturation but still lack of stabilization. Recent studies have largely converged on a three-component model involving relevance (usefulness), novelty, and unexpectedness (surprise). Some studies, however, reflect earlier two-component definitions or introduce nuanced variations.

- **Relevance:** Most researchers agree that relevance is a crucial aspect; serendipitous recommendations are not random and should align with user’s profile and interests. Items should remain pertinent to the user’s interests [5].
- **Novelty:** Serendipitous recommendations should often be new to users. Novelty usually describes two aspects: new items in the RS, and items the user has not encountered before. How-

ever, novelty alone doesn’t guarantee serendipity if the new item is very similar to the user’s existing interests. It often refers to new or unknown items [5], [14].

- **Unexpectedness (Surprise):** Many researchers consider surprise a key aspect, where the item violates users’ expectations in a positive way [21]–[23]. Unexpectedness is often treated as low similarity to a user’s profile or popular items [24], or items that cannot be generated by a normal RS [24]. It can involve items that meet a user’s short-term demand [25], deviate from typical interests [25], or are low-interest items that turn out to be satisfying [26].
- **Diversity:** Some researchers also consider diversity a component of serendipity, as providing different and diverse kinds of items can enhance serendipity by offering a broader range of options [22], [27], [28]. However, the relationship is complex; Wang *et al.* found diversity correlated with serendipity in movies [24] but anti-correlated in e-commerce [30], suggesting further research is needed. A common assumption is that user historical data represents all interests, but users might have unobserved interests. Diversity, by introducing more kinds of items, can help uncover these.
- **User Curiosity:** Xu *et al.* tied user curiosity to serendipity, suggesting high-curiosity users might seek more serendipitous experiences, while low-curiosity users prefer familiar recommendations. This allows for adjusting the degree of serendipity for different users [31], [32].

Recently, Kotkov *et al.* argued that the original meaning of serendipity is broader, referring to accidental discoveries that are valuable, even if not novel or unexpected, suggesting serendipity can manifest in multiple ways beyond the common ”relevance + unexpectedness + novelty” definition Kotkov *et al.*, Kotkov *et al.* While serendipity lacks a single consensus definition, it generally implies that a recommendation delights the user with the unexpected by introducing something new yet relevant. In practice, researchers often consider an item serendipitous if it is relevant to the user and provides a pleasant surprise through novelty or unexpectedness.

Component	Definition of Serendipity	Study
Unexpectedness	Items meet the user’s short-term demand	[3]
	Items deviate from the user’s typical interests	[4]
	Low-interest items	[25]
	Surprise and accident	[34]–[36]
	items the user thinks they would not found without RS	[6], [37]
Novelty	New or unknown item	[6], [34], [36], [37]
	Novel and unrepeated recommendations	[4]
Relevance	Items remain pertinent to the user’s interests	[3], [4], [34], [36]
Diversity	variation among the recommended items	[23], [38]

Table 1: Various Components Definitions of Serendipity

Paper	Definition of Serendipity
Kotkov <i>et al.</i> (2024) [6]	relevance + novelty + unexpectedness
Fu <i>et al.</i> (2025) [34]	relevance + unexpectedness
Binst (2024) [36]	novelty + surprise + valuable + taste broadening
Xu <i>et al.</i> (2024) [25]	low interest and high satisfaction
Xi <i>et al.</i> (2025) [39]	novelty + unexpectedness + relevance
Wang <i>et al.</i> (2024) [40]	relevance + unexpectedness
Tokutake <i>et al.</i> (2024) [41]	relevance + unexpectedness + novelty + diversity
Ping <i>et al.</i> (2024) [23]	novelty + unpopularity + unexpectedness
Wang <i>et al.</i> (2023) [42]	novelty + timing
Gyewon Jeon <i>et al.</i> (2024) [37]	novelty + unexpectedness + relevance
Hassan <i>et al.</i> (2024) [43]	unexpectedness + relevance + diversity
Hasan <i>et al.</i> (2023) [44]	item surprise + user similarity

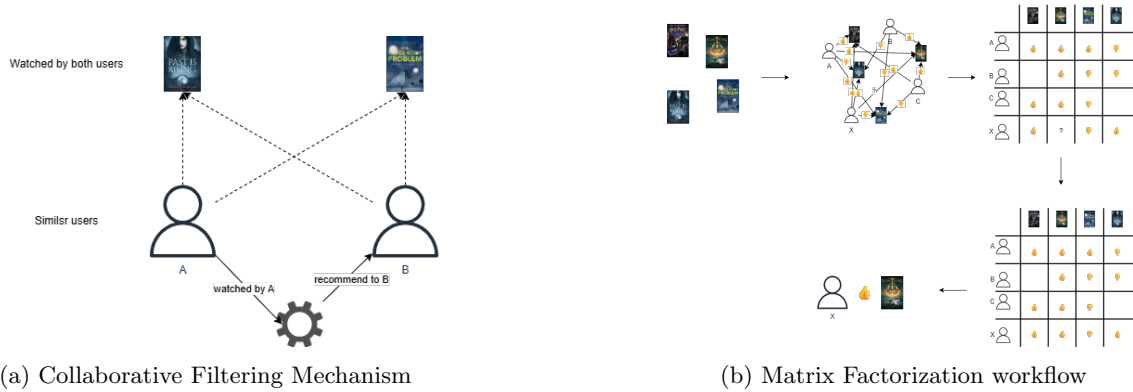
Table 2: Various Definitions of Serendipity

In summary, although these components may have overlapping definitions and computations, they are widely believed to contribute to serendipitous recommendations. The definition of serendipity in RS remains ambiguous, and this lack of consensus makes current research difficult to compare and generalize [5].

the user-based CF is hypothesized to indirectly facilitate serendipity by suggesting PAs liked by similar users, which can inherently introduce novel and unexpected items beyond the user’s direct interaction history [45]. Results showed that although CF produced objectively more diverse recommendations, users in the CBF group perceived their chosen PAs as increasingly more serendipitous over time [47].

3.1.1.2 Matrix Factorization (MF) based CF

MF decompose user-item interaction matrix into two latent factor matrices, enabling accurate prediction of missing ratings and addressing data sparsity by capturing user preferences and item attributes [45]. Kotkov *et al.* proposed Serendipity-Oriented Greedy (SOG) algorithm to introduce serendipity at re-rank stage. First they build a SVD accuracy-oriented RS, generate the initial recommendation list, then they adopt SOG to inject the serendipity recommendations, rebuild the list. SOG selects items based on a weighted score that combines predicted relevance from initial RS, diversity dissimilarity to items already selected for the new list, profile dissimilarity to promote unexpectedness, and item unpopularity to promote novelty. Experimentally, this SOG applied on SVD outperformed baseline algorithms in terms of serendipity and diversity, and also surpassed other serendipity-oriented algorithms in accuracy, although it was less accurate than purely accuracy-focused methods like SVD alone. A key finding was that increasing diversity can improve serendipity up to a point, but often at the cost of accuracy, and excessive diversity can negatively impact serendipity [48]. Fu *et al.* proposed SerenCDR-MF for cross-domain serendipity recommendations, utilizing MF to generate initial pre-calculated user and item embeddings from rating matrices. In the training stage it introduce serendipity by dual-loss functions. Bayesian Personalized Ranking (BPR) to learn novelty and auxiliary loss function to learn unexpectedness items, tackles the inherent data sparsity of serendipitous events through cross-domain knowledge transfer by decomposing embeddings into domain-shareable and domain-specific parts with learnable alignment matrices, even without overlapping users/items between domains [34].



3.1.2 Content-based Filtering (CBF)

The CBF mechanism involves two key components: item profiles and user profiles. Item profiles are constructed from the features or attributes of the items, such as movie genres, actors, textual descriptions, or product specifications. A user profile is then created by aggregating the features of items the user has liked, effectively building a model of their specific interests. Recommendations are generated by matching the attributes of unrated items against the user’s profile and suggesting the items with the highest similarity score. The primary strength of CBF lies in its user independence; it does not require data from other users, thus avoiding the “new user” cold-start problem common in collaborative methods [45], [49]. However, its core mechanism leads to a significant and widely recognized limitation: over-specialization. By design, CBF excels at finding more of the same, reinforcing a user’s existing preferences but rarely expanding them. This creates a “filter bubble” where users are shown a narrow, predictable range of content, making it fundamentally difficult for the system to generate novel, diverse, or serendipitous recommendations [45].

Researchers have explored methods to augment or redesign the CBF process to deliberately introduce novelty and surprise. These efforts represent a progression from adding simple heuristics to employing more complex generative and evolutionary techniques. Iaquina *et al.* (2008), was to introduce serendipity by adding a heuristic-based module to a CBF system. Their method did not alter the

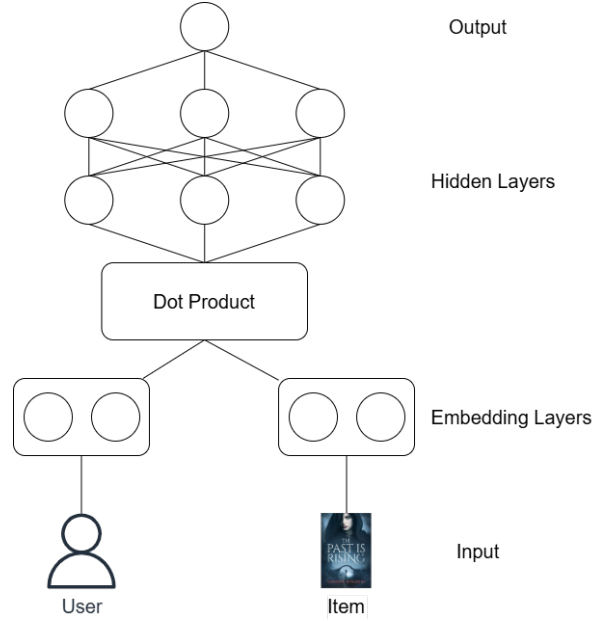


Figure 3: General Deep Learning Based RS

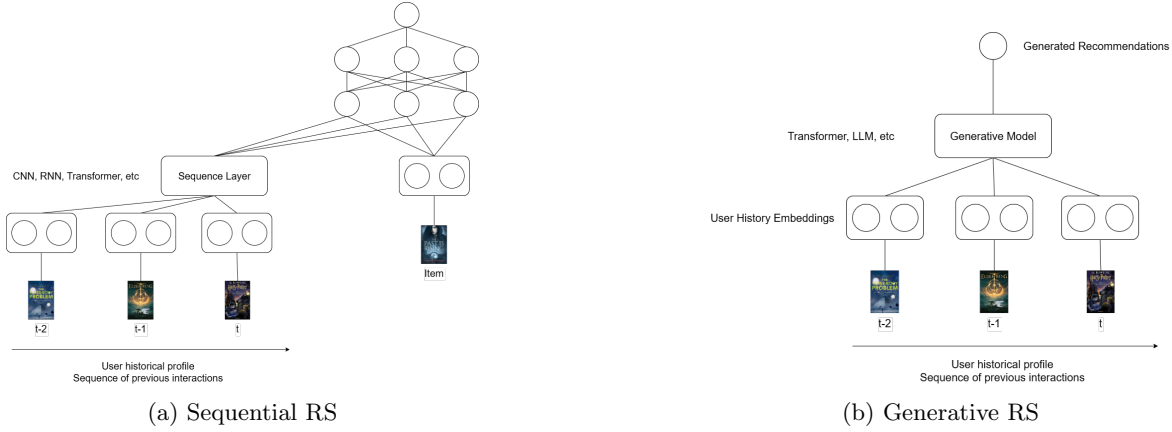
core CBF model but instead identified potentially serendipitous items by searching for those where the system was most uncertain in its classification. The underlying assumption is that an item the system struggles to categorize as either "like" or "dislike" for a user is likely unfamiliar to them and thus a candidate for a surprising discovery. This use of classification uncertainty serves as an elegant computational proxy for the subjective human experience of surprise [9].

A more advanced and integrated approach is demonstrated by [50] (2022) with their RRSQA model, which employs a Genetic Algorithm (GA) to fundamentally reshape the recommendation generation process. Instead of scoring items individually, RRSQA treats entire recommendation lists as "chromosomes" within a population. The system then introduces serendipity through two key evolutionary operators: Fitness Function: Each list is evaluated based on a fitness function that rewards not only relevance to the user's profile but also diversity and novelty. Crossover and Mutation: The GA combines strong "parent" lists and, crucially, introduces random changes to items within a list. This mutation operator serves as a direct and explicit mechanism for injecting unexpectedness, breaking the deterministic cycle of similarity-matching and forcing the system to explore items far outside the user's immediate profile. By evolving a population of candidate lists, RRSQA moves beyond simple ranking to search for an optimal list that holistically balances relevance and serendipity. This represents a shift from using a simple heuristic proxy to employing a formal optimization technique aimed at generating serendipitous outcomes [50].

3.2 Deep Learning Approaches

3.2.1 Multi-Layer Perceptron (MLP)/ DNN-based Approaches

Deep Neural Networks (DNNs) based on the Multi-Layer Perceptron (MLP) have become a cornerstone of modern RS. Their primary advantage over traditional models like MF is the ability to model non-linear and complex user-item interactions. While MF models are inherently linear, MLPs use multiple layers of neurons with non-linear activation functions to learn hierarchical and intricate feature representations from scratch. In a typical MLP-based RS, embeddings of users and items, often concatenated with other features, are fed through a series of hidden layers, with the output layer predicting a final score, such as a rating or click probability [5], [51]. This capacity to capture complex patterns makes DNNs a powerful tool for moving beyond simple relevance prediction and toward more nuanced recommendation objectives like serendipity. Recent research has leveraged the non-linear modeling capabilities of MLP/DNNs to explicitly generate serendipitous recommendations. These approaches use deep learning not as a monolithic solution but as a targeted tool to model specific facets of serendipity,



such as novelty, unexpectedness, or user curiosity.

The NSR model by Xu *et al.* (2020) exemplifies this by combining a linear MF component to ensure accuracy with an MLP to capture novelty. In this framework, the MLP’s task is to predict satisfaction from the complex, non-linear signals in user-item interactions, while MF predicts interest based on more linear patterns. Serendipity is then defined as items with high predicted satisfaction but low predicted interest. The MLP is crucial here because only a non-linear model can effectively learn the subtle patterns that indicate a user might be satisfied by an item that deviates from their established interest profile [18]. More advanced applications use DNNs to model a user’s dynamic state or intent, which then guides the recommendation process. The JD-Seren proposed by Wang *et al.* (2023) uses a DNN in its first stage as a category novelty scorer. This model analyzes a user’s behavior history to output a score quantifying how novel a product category is to that user. This personalized novelty score is then fed into a downstream uplift model that determines if and how much novelty to inject into the recommendations at that moment [42]. Similarly, Xu *et al.* (2025) use an MLP to predict the usefulness (relevance) of an item, but the final serendipity score is a dynamically weighted combination of this usefulness and a separate unexpectedness score. The weight is determined by a computational model of the user’s ”diversive curiosity,” which is estimated from their long- and short-term preferences. In these systems, the DNN provides the personalized, data-driven signal of what constitutes ”unexpected” or how much ”relevance” is required for a specific user, enabling a more tailored approach to serendipity. Another emerging direction is to use DNNs as the core predictive engine within a user-in-the-loop system [31]. Gyewon Jeon *et al.* (2024) propose an interactive feedback loop where users can provide counterfactual feedback. In their system, a powerful DNN acts as the click-through rate predictor. Serendipity is not generated by the DNN directly but is elicited through user interaction that modifies the input features fed to the model. The DNN’s role is to process these counterfactually modified user profiles and generate new, updated recommendations. The MLP component of DeepFM is essential for capturing the high-order feature interactions that emerge from these user-driven ”what-if” scenarios, thereby producing results that are both relevant and surprisingly different from the original recommendations [37].

3.2.2 RNN-based Approaches

Recurrent Neural Networks (RNNs), including their advanced variants like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), are uniquely suited to model the temporal dynamics of user behavior. Unlike feed-forward networks, RNNs maintain an internal hidden state that acts as a memory, allowing them to process sequences of user-item interactions chronologically. By learning the transitional patterns between items, RNN can effectively predict a user’s next likely interaction, making it powerful for tasks like session-based and next-item recommendation where sequential context is paramount. The inherent nature of RNNs to ”predict a perpetuation of past observed behavior” creates a fundamental tension with the goal of serendipity. Because standard RNNs are optimized to predict the next most likely item, they excel at capturing normative patterns but, when used naively, can inadvertently place the user in a ”filter bubble.” [5], [45], [51] Research in this area has therefore evolved from identifying this limitation to using RNNs as sophisticated components within more complex serendipity-aware architectures.

The challenge was clearly articulated by Pardos *et al.* (2020) in their study of a university course recommender. They found that their production RNN-based system, while accurate at predicting enrollment, exhibited a "dramatic lack of novelty." The RNN recommended predictable courses that fulfilled major requirements, which students already knew about. This study underscored that an RNN's optimization objective for sequential accuracy is often misaligned with the goal of fostering surprising discoveries [35]. However, rather than abandoning RNNs, subsequent research has demonstrated their value as powerful tools for modeling the dynamic user state, which is a crucial input for personalizing serendipity. The PURS model by Li *et al.* (2020) exemplifies this approach. It uses a bidirectional GRU with a self-attention mechanism to process a user's behavior sequence. The output of this GRU is not the final recommendation but a user interest embedding that captures the relevance component of a hybrid utility function. This relevance signal is then combined with a separately calculated, personalized "unexpectedness" score to generate the final serendipitous recommendation. Here, the RNN provides the necessary, context-aware foundation upon which unexpectedness can be intelligently layered [52]. Similarly, Xu *et al.* use an LSTM model to specifically capture a user's short-term preferences from their recent interaction history. This dynamic, short-term profile is then contrasted with a stable, long-term profile using SVD to computationally derive a "diversive curiosity" score for each user. This curiosity score then acts as a personalized weight, determining the final balance between usefulness (relevance) and unexpectedness in the re-ranking process. In both cases, the RNN is not tasked with generating the surprise itself but with modeling the user's evolving context with high fidelity [31]. This allows a separate, dedicated mechanism to effectively introduce unexpectedness that is still personally relevant.

3.2.3 CNN-based Approaches

Ziarani *et al.* (2021) [26] introduce SDNet, a serendipity oriented recommendation framework that augments traditional accuracy driven models through integration of a one dimensional convolutional neural network together with Particle Swarm Optimization and Serendipitous Personalized Ranking. SDNet begins by extracting six user level features from the Serendipity Dataset 2018—including the serendipity label, average rating, mean popularity, weighted deviation mean average, average negative feedback and openness—to form a six-by-one input vector . A convolutional neural network composed of three filters of kernel size three, unit stride and Rectified Linear Unit activation processes this input to generate joint predictions of unexpectedness and relevance per user . In the candidate generation phase, Particle Swarm Optimization explores the neighborhood defined by predicted unexpectedness within a tolerance threshold to identify unrated items whose relevance deviation falls below a predefined parameter . Finally, Serendipitous Personalized Ranking employs a Serendipity AUC objective that balances accuracy and item popularity via a tunable parameter to re-rank PSO candidates . Empirical evaluation on serendipity rate, hit ratio and Normalized Discounted Cumulative Gain demonstrates that SDNet delivers relative improvements of at least 38 percent and 31 percent over leading baseline methods respectively, achieving an NDCG of approximately 0.88 while maintaining a mean absolute error below 0.07 across varying dataset sizes . Despite its strong performance, the model relies on historical ratings, which limits its effectiveness for new users and new items, and it does not incorporate temporal weighting of interactions, potentially reducing responsiveness to evolving user preferences. Chen *et al.* (2024) [53] proposed CKLF, in this framework CNN plays a critical and specific role at the output stage of the SAN module, where it functions as a sophisticated feature fusion and abstraction layer to create a final, serendipity-aware item representation. The input is a multi-channel matrix formed by stacking two distinct representations for each item, the initial, dense word embeddings that capture linguistic meaning, and the new, KG-enhanced entity embeddings generated by the spreading activation process to synthesize the diverse and unexpected signals from SAN into a single, holistic item embedding. Through its feature extraction and max-pooling layers, it distills the most potent signals—both from the item's core content and its newly discovered relational context—into a final, fixed-size vector. At final, serendipity-aware embedding is then used for the downstream matching task against the user's profile [53].

3.2.4 GAN-based Approaches

Generative Adversarial Networks (GANs) framework operates through a minimax game between two neural networks: a Generator (G), which learns to create synthetic data that mimics a target distribution, and a Discriminator (D), which learns to distinguish the generator's synthetic data from

real data. This adversarial training process forces the generator to produce increasingly realistic and high-quality data. In RS, GANs are typically employed for generative data augmentation, where the generator creates synthetic but plausible user-item interaction data to enrich sparse training sets, and generating hard negative samples, which can improve the robustness of discriminative ranking models. This approach fundamentally treats the GAN as a sophisticated module for enhancing the data or training process of a downstream recommendation model [51]. The inherent ability of GANs to generate novel data distributions makes them a compelling tool for fostering serendipity. Instead of simply replicating existing interaction patterns, a well-designed GAN can learn the underlying collaborative structures and generate interactions that represent novel yet plausible user interests, directly addressing the core of a serendipitous recommendation.

Xu *et al.* (2024) propose GS-RS, a framework that uses a GAN as a generative pre-processing module to explicitly model serendipity. The core of their method is a Twin Conditional GAN that learns to generate two distinct fine-grained user preference signals: interest and satisfaction. Serendipity is computationally defined as items with a combination of low predicted interest and high predicted satisfaction. The GAN’s role is to generate these preference scores for unrated items, creating ”virtual but convincing” user preference profiles. This generated data is then used to create an enhanced, denser user-item matrix which serves as a superior input for a traditional downstream recommender system. The GAN does not produce the final recommendation but enriches the input data with explicit, computationally derived serendipity signals, enabling the downstream model to better balance accuracy and novelty [25]. Hassan *et al.* take a different approach with GCZRec, a framework that uses GANs to directly synthesize the user-item interaction matrix, achieving serendipity implicitly. Their model employs a dual-generator architecture based on conditional Wasserstein GANs (cWGANs). One generator learns to produce interaction scores for all users given a news item, while the other generates scores for all news items given a user. Serendipity emerges from the model’s ability to learn and generate based on latent collaborative signals. By training on historical interactions, the GAN captures the complex, non-obvious relationships between users and items, allowing it to generate diverse recommendations that expand a user’s interests. For cold-start scenarios, a separate zero-shot classifier first predicts a user or item’s ”class”, which is then used as the condition for the cWGAN to generate plausible interactions, making this approach particularly effective at introducing serendipity for new users and items [43]. The application of GANs to serendipity, while powerful, is still in its early stages and presents several key challenges and future research directions. The current approaches demonstrate two different philosophies: GS2-RS models serendipity explicitly through a defined objective, while GCZRec achieves it implicitly as an emergent property of collaborative data generation. A significant future direction is to bridge these approaches by designing GANs with serendipity-aware loss functions. This would involve incorporating a regularizer into the GAN’s objective that explicitly rewards the generator for producing outputs that are not only realistic but also diverse, novel, or unexpected according to a defined metric.

3.2.5 Knowledge Graph-based Approaches

Knowledge-Based RS are a powerful class of systems that leverage rich, structured domain knowledge, often in the form of ontologies or Knowledge Graphs (KGs), to generate recommendations. KB can reason over complex relationships between entities. In a typical KG-based RS, entities are represented as nodes and their relationships as edges. Recommendation is often framed as a link prediction task, where the model learns to infer plausible but missing connections between a user and items. This approach is particularly effective at addressing data sparsity and the cold-start problem, and its inherent structure provides a high degree of explainability for its recommendations. The structured, relational knowledge in a KG makes it an exceptionally well-suited foundation for generating serendipitous recommendations [51]. By traversing multi-hop paths and exploring non-obvious connections, a KG-based system can discover items that are relevant but unexpected, directly burst the ”filter bubble.”

Inoue *et al.* (2024) pioneered a novel approach by combining a KG with an Interactive Genetic Algorithm (IGA) to induce ”serendipitous learning” at school. Their system recommends sequences of scientific discoveries which are represented as traversals through a KG. Serendipity is introduced through the IGA, an evolutionary process where the user’s feedback serves as the fitness function. Crossover and mutation operators create new, potentially novel learning paths by combining and altering existing high-rated paths. Crucially, if a generated path does not exist in the KG, the sys-

tem uses Dynamic Time Warping (DTW) to find the most semantically similar existing path. This constrains the exploration, ensuring that even novel recommendations remain coherent and relevant. This interactive, user-in-the-loop approach empowers user agency and uses the KG as a structured playground for serendipitous discovery [54]. Building on the idea of KG traversal, Chen *et al.* (2024) developed the Cognitive-based Knowledge Learning Framework (CKLF), which integrates principles from cognitive psychology to more explicitly model and generate serendipity. The framework’s core component for serendipity is the Spreading Activation Network (SAN). This module simulates the associative nature of human memory on a KG. Instead of just considering immediate neighbors, the SAN propagates activation signals along diverse semantic paths to discover more distant, high-order entities. This ”spreading activation” is a direct mechanism for unearthing unexpected connections. For example, it might connect a user interested in a specific political figure (e.g., Trump) to a related but non-obvious entity (e.g., Giuliani) through a multi-hop path, which is then used to recommend a relevant but surprising news article. By systematically exploring these high-order relationships, the SAN enriches item embeddings with diverse semantic information, which is then synthesized by a CNN into a final, serendipity-aware representation [53].

3.2.6 Transformer-based Approaches

Transformer-based RS have marked a significant advancement in sequential recommendation by processing sequences of user-item interactions non-sequentially, primarily through self-attention mechanisms [51], [55]. These mechanisms assign adaptive weights to all items within a user’s historical sequence, allowing the model to discern the varying importance of past interactions and capture complex, long-range dependencies. This approach overcomes the limitations of traditional sequential models like RNNs and facilitates efficient parallel processing of extensive interaction data, leading to enhanced recommendation quality by more accurately modeling dynamic user preferences [55]. Seminal models like BERT4Rec exemplify this paradigm, treating recommendation as a sequence modeling task akin to masked language modeling in NLP [56].

While the base Transformer architecture excels at predicting a user’s next likely action, this inherent focus on perpetuating past behavior can lead to over-specialization. Consequently, recent research has focused on adapting the powerful sequence modeling capabilities of Transformers to explicitly foster serendipity, which requires a deliberate and intelligent deviation from the most predictable path [5]. An initial approach is to frame serendipity as a multi-task learning problem directly within a Transformer architecture. The SNPR model by Zhang *et al.* (2021) for Point-of-Interest (POI) recommendation is a prime example. Zhang *et al.* first established a quantitative definition of serendipity as a trade-off between relevance and unexpectedness. Then designed a dedicated neural network with a Transformer encoder to capture the complex, non-sequential interdependencies between POIs in a user’s check-in history. The model is trained with a multi-task objective to simultaneously optimize for both relevance and unexpectedness, demonstrating that the Transformer’s ability to model the entire sequence context can significantly improve relevance without sacrificing the unexpectedness crucial for serendipity. Building on this, Fu *et al.* developed a series of works that use Transformers as a base architecture and augment them with specialized modules. First, Fu *et al.* (2023), proposed a self-enhanced module SerenEnhance that can be coupled with any existing RS. This module injects serendipity during training by learning the fine-grained facets of unexpectedness and relevance through two auxiliary loss functions. Crucially, it mitigates the sparsity of serendipity data by using self-generated labels for these auxiliary tasks, demonstrating superior performance in predicting serendipity with a relatively low sacrifice to relevance. Later, Fu *et al.* (2025) further advanced this work to the more challenging context of cross-domain serendipity. This model utilizes pre-trained BERT embeddings and tackles extreme data sparsity by transferring knowledge between a source and target domain, even without user or item overlap. Methodologically, it injects serendipity during training by decomposing embeddings into domain-shareable and domain-specific parts and employing a similar dual-loss mechanism to capture unexpectedness and relevance in a cross-domain context, achieving superior serendipity prediction over single-domain baselines. The most recent evolution in this area is the use of Decision Transformers to frame multi-objective recommendation as a sequence modeling problem solvable with reinforcement learning (RL) principles. The Hierarchical Decision Transformer (HDT) framework by Wang *et al.* (2024) models user’s long-term and short-term preferences. It introduces hierarchical unexpected returns as a reward signal, allowing the model to dynamically balance accuracy with diversity, novelty, and serendipity. This approach leverages the Transformer’s ability to model long sequences

of states, actions, and rewards, enabling it to learn a policy that can generate recommendations to satisfy complex, multi-faceted objectives like serendipity over the long term.

3.3 Generative Recommendation

Recent advancements in generative models, particularly Large Language Models (LLMs), are catalyzing a paradigm shift in RS, moving from traditional discriminative models that rank a pre-filtered set of items to generative frameworks that directly produce recommendations or related content. This is achieved by representing users and items as sequences of tokens—either as natural language (e.g., titles, metadata) or as learned, tokenized "semantic IDs"—and reformulating recommendation as a sequence-to-sequence task [58], [59].

Two primary approaches have emerged: 1) LLM-based Recommendation, which adapts large, pre-trained LLMs to RS tasks through techniques like zero-shot prompting or instruction tuning to align them with user behaviors and collaborative signals, and 2) Semantic ID-based Recommendation, which trains specialized generative models to directly output tokenized item IDs, offering scalability and efficiency by eliminating massive embedding tables. In both cases, the core working principle involves framing recommendation as a generative process: user data, item context, and task-specific instructions are combined into a prompt, which is then processed by the LLM to auto-regressively generate the desired output [58], [59]. This generative approach is particularly promising for serendipity, which aims to provide users with unexpected yet relevant discoveries that break the "filter bubble." By moving beyond simple ranking, generative models can facilitate richer, more human-like interactions and generate novel explanations that guide users toward these serendipitous encounters.

Tokutake *et al.* (2024) directly investigated "Can Large Language Models Assess Serendipity in Recommender Systems?", to research whether the LLM's understanding of a concept as subjective as serendipity aligns with human emotional responses. They use LLMs like GPT-3.5 and Llama2 as zero-shot binary classifiers to predict human-annotated serendipity labels from the Serendipity-2018 dataset. Their findings reveal that while the models' judgments do not perfectly align with human assessments, they can achieve performance comparable to or better than traditional serendipity-oriented baseline methods. This demonstrates the potential of LLMs as evaluators or proxies for human judgment, though it also underscores the need for further alignment to capture the nuanced and personal nature of a serendipitous experience [41]. Given that LLMs are sensitive to how they are prompted, another critical research area is understanding how to "ask" for serendipity effectively. The work by Fu *et al.* (2024) on SerenPrompt dive deep into this, exploring various prompt engineering strategies. They compared discrete (natural language), continuous (trainable tokens), and hybrid prompts. More importantly, they contrasted a direct style (e.g., "Is this item serendipitous?") with an indirect style that breaks the question down into its core components ("Is this item unexpected? And is it relevant?"). Their results show that the indirect style, which explicitly guides the LLM's reasoning process, performs significantly better. This suggests that decomposing the abstract concept of serendipity into its more tangible facets is a more effective way to elicit high-quality serendipitous recommendations from LLMs [1]. Bridging the gap between theoretical models and practical application is a major hurdle, especially given the high latency of LLMs, research group from Alibaba Xi *et al.* (2025) proposed SERAL, an industrial framework deployed on Taobao. Their approach is a pragmatic hybrid system that leverages an LLM SerenGPT in a nearline capacity. First, user behavior is compressed into "Cognition Profiles" to manage long histories. Then, SerenGPT is aligned with human serendipity judgments using preference optimization and generates a cache of high-quality serendipitous candidates offline. These candidates are then injected into the traditional online ranking pipeline. This methodology successfully breaks the filter bubble and enhances user engagement in a live environment, demonstrating a viable path for integrating powerful but slow generative models into latency-sensitive industrial recommender systems [39].

The intersection of generative models and serendipity is ripe with opportunities. Future work should aim to move beyond the current hybrid and evaluative systems toward fully leveraging the unique capabilities of generative models. Current approaches like SERAL still rely on a traditional downstream ranker. A key future direction is to develop true end-to-end generative models that can directly produce a finalized, high-quality list of serendipitous recommendations from a prompt, completely bypassing the conventional multi-stage pipeline. This would require innovations in constrained decoding and training objectives that jointly optimize for relevance, unexpectedness, and list-level coherence. A major risk,

especially when generating novel items or explanations, is hallucination. A generated recommendation must be grounded in reality (e.g., a recommended restaurant must exist) or, if novel, must be plausible and useful. Future research needs to focus on developing robust methods for constrained generation and retrieval-augmentation, ensuring that generative outputs are factually accurate, trustworthy, and do not lead to user frustration.

3.4 Transfer Learning

Transfer learning (TL) has emerged as a powerful technique in RS, primarily aimed at improving the performance of a target model by transferring knowledge from a different but related source domain. This approach is particularly effective at mitigating data sparsity problems, such as the cold-start issue for new users or items. This paradigm is exceptionally well-suited for the challenge of serendipity recommendations. Serendipity—providing unexpected yet relevant items—is notoriously difficult to model due to the extreme scarcity of ground-truth data. While large datasets with relevance signals are abundant, datasets with explicit user feedback on serendipity are rare. This creates a natural transfer learning scenario: the abundant relevance data can serve as a source domain for a proxy task, and the sparse serendipity data can serve as the target domain [60].

By carrying information across domains or tasks, TL expands the candidate pool and relationships considered during recommendation generation, helping to break out of the “filter bubble” of a user’s history. The application of transfer learning to serendipity recommendations is a nascent but promising field, with two pioneering papers demonstrating its viability through distinct approaches.

The first major work by Pandey *et al.* (2018) proposed SerRec, which introduced a classic pre-training and fine-tuning transfer learning methodology. They first trained a deep neural network on a large-scale general recommendation dataset MovieLens, to learn general user-item interaction patterns. Subsequently, they froze the initial layers of the network and fine-tuned the final layers using a much smaller, serendipity-oriented dataset Serendipity-2018. The pre-training on relevance scores acted as a crucial proxy task, allowing the model to develop a robust understanding of user preferences before specializing in the more complex task of identifying serendipity. This two-stage approach was shown to significantly improve the system’s ability to produce serendipitous item suggestions [24].

Building on such ideas, Fu *et al.* (2025) proposed a more sophisticated deep transfer learning model, SerenCDR, which frames the problem as cross-domain recommendation between different item categories from books to movies without requiring overlapping users or items. Methodologically, SerenCDR achieves this by decomposing user and item embeddings into domain-shareable and domain-specific parts, allowing it to learn a transferable “essence” of serendipity-seeking behavior. It employs a dual-loss mechanism where a main loss learns from direct serendipity labels and an auxiliary loss computationally models the core facets of serendipity unexpectedness and relevance. This approach yielded positive results in enabling “unexpected but valuable” discoveries for users by alleviating the inherent sparsity of serendipity data. In sum, transfer learning provides the bridging knowledge that helps RS suggest items outside a user’s typical realm while still aligning with their latent interests [34].

By borrowing information from large auxiliary datasets or domains, TL allows serendipity-oriented models to learn patterns that would be impossible to infer from the scant serendipity-labeled data alone, leading to more robust models that can recognize what might pleasantly surprise users. Knowledge transfer can connect disparate domains of user behavior, increasing the diversity of items considered. This enables recommendations that are both relevant and novel, improving user satisfaction and engagement. Experiments have reported that models using TL produce more “pleasant surprises,” giving users increased chances of encountering valuable unexpected items [24], [34]. Despite its promise, applying TL for serendipity also presents challenges. As previously discussed, serendipity is a complex construct, and explicit feedback data is extremely scarce. While TL can leverage proxy signals like relevance ratings, obtaining or synthesizing high-quality serendipity labels for fine-tuning remains an open challenge.

3.5 Comparative Analysis of Methods

The preceding review of SOTA approaches reveals a clear and rapid methodological evolution in the pursuit of serendipitous recommendations. While early research laid the conceptual groundwork, the field has transitioned from augmenting traditional models with clever heuristics to developing

sophisticated, end-to-end deep learning frameworks. This section provides a comparative analysis of these methods, synthesizing their underlying principles to identify overarching trends, highlight the current situation, and outline key challenges and future directions.

3.5.1 Methodological Trajectory: From Heuristics to Generative

The history of serendipity-oriented RS can be understood as a three-stage trajectory, reflecting broader trends in machine learning. Stage 1: Heuristic Augmentation and Formal Optimization. Initial efforts focused on modifying traditional RS, like CBF to counteract their tendency for over-specialization. This was achieved either through heuristic proxies, such as formal optimization techniques like the Genetic Algorithm in RRSQA [50], which evolved entire recommendation lists. These methods treated serendipity as an external objective to be optimized, often in a post-processing or re-ranking stage, rather than an intrinsic property learned from the data itself.

Stage 2: The Deep Learning Shift and In-Processing Integration. The advent of deep learning marked a significant paradigm shift. Models based on MLP, RNN, and CNN architectures enabled the learning of complex, non-linear user and item representations, allowing for a more nuanced modeling of serendipity’s core facets. This moved serendipity from a post-processing step to an in-processing objective. Hybrid models like NSR [18] exemplify this by using an MLP as a dedicated “novelty engine” alongside an MF component for accuracy. Sequential models like PURS [52] and knowledge-based models like CKLF [53] further refined this by integrating serendipity directly into their network architectures and training objectives.

Stage 3: The Rise of End-to-End Generative Recommendation. The most recent transformation is the move from discriminative deep learning models to end-to-end generative frameworks. Powered by architectures like GANs, Transformers, and LLMs [58], this new paradigm redefines the problem itself. Instead of scoring a pre-filtered list of candidates, these models aim to generate a recommendation directly from a vast, often unstructured, input space. This approach holds immense promise for serendipity, as generation is inherently more aligned with discovery than discrimination is with ranking.

3.5.2 Current Situation

The current SOTA is characterized by deep learning models that decompose serendipity into its constituent parts—typically balancing relevance with a form of unexpectedness, novelty, or diversity. We can comparatively analyze these dominant approaches along two key axes:

1. Architectural Specialization for Serendipity Facets: There is a clear trend of specializing different deep learning architectures for the specific facets of serendipity they are best suited to model.
 - Relevance Modeling: RNNs and Transformers are the architectures of choice for ensuring relevance, especially in sequential contexts. By modeling the temporal dynamics of user behavior, models like PURS (GRU) and SNPR (Transformer) can create a highly accurate, context-aware user profile. This profile serves as the necessary anchor of relevance, against which surprising suggestions can be safely made.
 - Unexpectedness and Novelty Modeling: To generate the “surprise” element, researchers have employed a wider variety of architectures. MLPs are used in models like NSR to capture the complex, non-linear relationships that might indicate a novel yet satisfying item. Knowledge Graph traversal mechanisms, such as the Spreading Activation Network in CKLF, explore multi-hop, non-obvious semantic paths to discover unexpected connections. Generative Adversarial Networks (GANs), as seen in GS2-RS and GCZRec, are used to generate synthetic interactions that represent novel but plausible user interests.
2. Explicit vs. Implicit Serendipity Injection: The methods also differ in how they operationalize serendipity.
 - Explicit Modeling: Many models, such as NSR and GS2-RS, explicitly define serendipity through a computational objective (e.g., low predicted interest + high predicted satisfaction). They then train a dedicated component of the model to optimize for this pre-defined metric. This approach offers interpretability and control but relies heavily on the correctness of the initial definition.

- **Implicit Modeling:** Other models, particularly generative ones like GCZRec, achieve serendipity implicitly as an emergent property. The GAN is trained to replicate the distribution of collaborative signals, and in doing so, it learns the complex latent relationships that naturally lead to diverse and unexpected suggestions. This approach requires less feature engineering but offers less direct control.
- **Human-in-the-Loop Modeling:** The most recent LLM-based approaches (e.g., SERAL, SerenPrompt) represent a new hybrid. Here, the abstract concept of serendipity is operationalized through a combination of sophisticated prompt engineering and preference alignment with human judgments, effectively teaching the model a human-centric understanding of a "good surprise."

3.5.3 Future Directions and Open Challenges

Like previously discussed, serendipity research becomes increasingly intertwined with generative AI, user interaction, and reinforcement learning, though this path is fraught with significant open challenges. The most significant future direction involves the transition from the current hybrid systems, where generative components often act as pre-processing modules, to true end-to-end generative models. The ultimate goal is to develop frameworks that can directly produce a finalized, high-quality list of serendipitous items from a prompt, completely collapsing the traditional multi-stage pipeline. This will require innovations in areas like constrained decoding and sequence-level training objectives that can jointly optimize for relevance, unexpectedness, and list-level coherence.

However, this ambition immediately surfaces the most critical research question in the field: the formalization of the serendipity objective function. Whether defined through a reinforcement learning reward signal, a preference score from a supervising LLM, or a computational metric of user curiosity, the challenge of teaching a model what constitutes a "good surprise" that aligns with long-term user satisfaction remains a fundamental and formidable obstacle.

Beyond this central conceptual problem, the practical implementation of powerful generative models introduces a triad of engineering and ethical challenges. First, scalability remains a major hurdle, as these computationally intensive models must operate at the speed and latency required by production systems. Second, the risk of hallucination is particularly acute; a generated recommendation must be grounded in reality to prevent user frustration and maintain trust, necessitating robust methods for constrained generation and retrieval-augmentation. Finally, the "black box" nature of these models requires a renewed focus on explainability, not just for debugging, but for building user trust by making the logic behind a surprising suggestion transparent.

Ultimately, the most sophisticated solutions may not be fully autonomous but will instead embrace the user as a collaborative partner. The future of serendipity recommendation likely lies in creating interactive systems that blend the automated exploration of knowledge graphs, the user-driven feedback of counterfactuals, and the preference-guided optimization of evolutionary algorithms. This vision shifts the goal from simply building a better algorithm to designing a true human-AI partnership for discovery, where the system and the user work together to navigate the vast space of information in a meaningful and delightful way.

Model (Paper)	Method	Advantages	Limitations	Suitability for Serendipity
TBSSRS [44]	Context-based + CF	Online learning handles non-stationary user interests	1) Manual labelling of serendipitous items without user engagement; 2) Small annotation sets limit generalizability	Aligns surprise and value by multiplying Bayesian surprise with positive user feedback; outperforms diversity-based baselines.
SerenCDR-MF [34]	MF	Fast pre-computation without text processing	Reliant on ratings only; struggles with cold-start items and users	Suited for datasets lacking textual data; outperforms baselines in rating-only recommendations.

continued on next page

Table 3 – *continued from previous page*

Model (Paper)	Method	Advantages	Limitations	Suitability for Serendipity
NSR [18]	MLP + MF + Filtering	1) Defines serendipity as high satisfaction but low interest; 2) Mitigates data sparsity in serendipity	Complex implementation with many hyperparameters; increases computational cost	Predicts interest via MLP and satisfaction via MF, then filters low-interest/high-satisfaction items to balance accuracy and serendipity.
SUCRS [35]	RNN + CF	Includes offline & online user studies; uses catalog descriptions & interaction history	Success depends heavily on how recommendations are presented	Directly measures user surprise and novelty through empirical studies.
DESR [3]	GMM + Capsule Network	Back-routing explanations build user trust; captures short-term interests	Explanations limited to historical items only	Combines long- and short-term behaviors into a serendipity vector.
PURS [52]	MLP + GRU	Linear scalability; validated via online A/B testing	Deep architecture increases complexity and cost	Measures divergence from user interests to inject serendipity.
SOG [48]	MF + SOG	Improves serendipity and diversity via re-ranking; easily applied to existing RS	Worst-case $O(n^3)$ complexity; some trade-off in accuracy	Re-ranks candidates by unpopularity, dissimilarity, and diversity.
SRS [27]	CBF	E-QSEM metric with user feedback balances relevance and novelty	Limited to a MovieLens subset; requires careful calibration	Validated qualitatively; injects serendipity without sacrificing satisfaction.
Coppens <i>et al.</i> 's method [47]	User-based CF/CBF	Measures both objective and subjective serendipity; tracks user learning over time	Cold-start and sparsity challenges; diversity-accuracy trade-offs	Evaluated in a physical activity RS; subjective ratings align with metrics.

Table 3: Traditional approaches for enhancing serendipity in RS

Model (Paper)	Method	Advantages	Limitations	Suitability for Serendipity
SerenCDR - BERT (Fu <i>et al.</i> (2025) [34])	BERT based	1) Powerful language understanding to capture deep patterns; 2) Handles item cold-start; 3) Dual loss functions address serendipity sparsity	1) Requires sizable, well-written reviews; 2) High computation and latency due to BERT encoding and dual-loss training	Suited for rich textual review datasets; outperforms baselines in cross-domain serendipity recommendations.
SRSCUC (Xu <i>et al.</i> (2025) [31])	SVD + MLP + LSTM + Clustering	Curiosity-weighted personalization adapts to behavior changes	Potential trade-off in accuracy; high computational cost	Dynamically balances unexpectedness based on user curiosity.
ISER (Gyewon Jeon <i>et al.</i> (2024) [37])	Content-based + MF + MLP	1) Near-zero latency; 2) Engages users with transparent interactions	1) Binary interactions cannot represent intensity; 2) Validated only on MovieLens	Maintains novelty and unexpectedness while ensuring relevance through proxy serendipity metrics.
GS-RS (Xu <i>et al.</i> (2024) [25])	Conditional GAN	1) Addresses cold-start and filter-bubble simultaneously; 2) Integrable with existing RS; 3) Negative-matrix injection mitigates sparsity	1) Complex computation from twin-CGAN training; 2) Relies solely on ratings, not handling grey-sheep users	Generates recommendations outside user interest bubbles with high satisfaction potential.

continued on next page

Table 4 – *continued from previous page*

Model (Paper)	Method	Advantages	Limitations	Suitability for Serendipity
GCZRec (Hassan <i>et al.</i> (2024) [43])	Conditional GAN	1) Jointly tackles cold-start and sparsity via matrix injection; 2) Easily integrable into existing RS pipelines	1) Requires manual threshold tuning; 2) Rating-only reliance; 3) Additional offline training cost	Explicitly designed for serendipity, showing superior performance to baseline models.
JD-Seren (Wang <i>et al.</i> (2023) [42])	MLP	1) Offline and online serendipity evaluation; 2) Investigates timing effects; 3) Lightweight re-ranking for low latency	1) Human-defined thresholds needed; 2) Dependent on rich interaction history	Balances novelty and relevance, exploring contextual timing for serendipity.
CKLF (Chen <i>et al.</i> (2024) [53])	CNN + KG + attention	1) KG side information alleviates sparsity and cold-start; 2) Optimizes CTR while boosting diversity and serendipity	1) User cold-start remains; 2) High latency from KG embedding computation	Defines serendipity via relevance, novelty, and unexpectedness using rich side data.
SerenEnhance (Fu <i>et al.</i> (2023) [57])	Transformer based	1) Easy RS integration; 2) Mitigates data sparsity; 3) Addresses item cold-start partially	1) User cold-start persists; 2) Higher computational requirements; 3) Serendipity enhancement may reduce accuracy	Uses loss balancing of unexpectedness and relevance, with SerenEnhance outperforming baselines.
SerGNN (Boo <i>et al.</i> (2023) [61])	GNN	1) Plug-and-play RS component; 2) Tunable parameter for trade-off balancing	1) Requires rich history; 2) Struggles with sparse data; 3) Short-term accuracy drop	Integrates unexpectedness and relevance, balancing serendipity and accuracy.
SPARKS (Wang <i>et al.</i> (2024) [40])	Transformer based	1) Simple integration into existing RS	1) Dual-transformer training overhead; 2) Needs extensive history	Defines serendipity as "relevant and beyond expectation", showing strong real-world performance.
SDNet (Ziarani <i>et al.</i> (2021) [4])	CNN + PSO + SPR	Balances serendipity and accuracy; fine-tunable via PSO; rich feature engineering	High complexity; sensitive hyperparameters; limited adaptability	Detects expectation-breaking yet relevant items.
SNPR (Zhang <i>et al.</i> (2021) [21])	Transformer based	λ hyperparameter balances relevance and unexpectedness	Complex multi-task learning; careful λ tuning required	Injects serendipity to combat over-specialization, outperforming baselines.
IEC (Inoue <i>et al.</i> (2024) [54])	Knowledge-based + IGA	Interactive feedback generates novel learning paths via KG & genetic algorithms	Computationally intensive; quality of KG and user effort are constraints	Produces unexpected yet relevant educational recommendations iteratively.

Table 4: Deep learning approaches for enhancing serendipity in RS

Model (Paper)	Method	Advantages	Limitations	Suitability for Serendipity
SerenPrompt (Fu <i>et al.</i> (2024) [1])	LLM based	1) No heavy fine-tuning required; only prompt engineering; 2) Leverages world knowledge to assess serendipity without user surveys; 3) Mitigates item cold-start and data sparsity	1) Performance sensitive to prompt design and hyperparameters; 2) Hybrid methods add pre-training complexity	Experiments demonstrate LLM prompting can outperform state-of-the-art serendipity models.

continued on next page

Table 5 – *continued from previous page*

Model (Paper)	Method	Advantages	Limitations	Suitability for Serendipity
LLM4Seren (Tokutake <i>et al.</i> (2024) [41])	LLM based	1) Utilizes LLM knowledge for serendipity assessment without surveys; 2) Partially solves cold-start and sparsity issues	1) Moderate alignment with human judgment; 2) Output sensitivity to prompts; 3) Limited output interpretability	Shows promise as a serendipity assessment tool, outperforming basic baselines.
SERAL (Xi <i>et al.</i> (2025) [39])	LLM based	1) Combines prompt engineering and fine-tuning with domain knowledge; 2) Evaluated offline and online; 3) Handles data sparsity; 4) Near-line adaptation for real-time requirements	1) High GPU and inference costs; 2) Limited scalability for tail items	Captures relevance and unexpectedness in e-commerce, boosting revenue.

Table 5: Generative approaches for enhancing serendipity in RS

4 Classic RS Challenges

4.1 Cold Start

A long-standing challenge is the cold start problem, which occurs when new users or new items enter the system with little or no historical interaction data. In this scenario, the RS cannot accurately predict user preferences or item relevance due to lack of information, often resulting in poor or trivial recommendations [46].

4.2 Sparsity

In online platforms, user-item interaction matrices are typically sparse, as most users have rated or interacted with only a tiny fraction of the available catalog. This data sparsity creates difficulty in finding overlapping user preferences or item similarities, leading to less precise recommendations [46].

4.3 Latency

The latency problem refers to delays in the recommendation cycle, especially in systems that update models in batch or require significant re-computation when new data arrives. For example, collaborative filtering may not recommend new items until enough users have rated them, causing a lag [46].

4.4 Synonym Problem

The synonymy issue arises when the same or very similar item is described in different ways (e.g. “NYC” vs. “New York City”). Traditional algorithms, especially collaborative filtering, may treat these entries as distinct and fail to aggregate their feedback. This leads to diluted item popularity and missed connections between user groups [46].

4.5 Grey Sheep Problem

Grey sheep users are those with idiosyncratic preferences that do not align well with any user community or segment. A grey-sheep user receives poor recommendations because the RS cannot find a reliable “neighbor” group with similar tastes [46].

4.6 Scalability

Modern RS must serve massive catalogs and user bases, introducing scalability concerns. Algorithms that work well on small datasets may become impractically slow or resource-intensive at scale. Platforms like Amazon handle tens of millions of items and customers [46].

5 Evaluation

5.1 Offline Evaluation

5.1.1 Datasets

Proper datasets are crucial for studying serendipity, in practice, there are only three publicly available serendipity-oriented dataset in small scales, so many researchers use proxy metrics to simulate serendipity, train and evaluate on proxy datasets.

- **Serendipity-2018**¹: this dataset was collected by user survey based on **MovieLens**, contains user answers to our questions and additional information, which have 10 million relevance ratings, and 2,150 serendipity labels on movies [62].
- **Taobao Serendipity dataset**²: this dataset was collected on Taobao, China largest e-commerce platforms, contains 11,383 users’ feedbacks through the user survey [30], [63].
- **SerenLens**³: this dataset is based on **Amazon Reviews Data**, Fu *et al.* utilize crowd-sourcing annotate the serendipity labels, contains 265,037 serendipity labels on books and 74,967 labels on movies [57], in 2025, Fu *et al.* extend it to **SerenCDRLens** for cross-domain serendipity-oriented recommendation tasks [34].

Dataset	Models	Metrics	Improvements
Serendipity-2018	LLM4Seren [41]	Serendipity@n	Serendipity@3: SOG(0.305), SPR(0.291), SVD(0.241), TD(0.227)
	SerRec [24]	$NDCG_{seren}@k$	$NDCG_{seren}@10$: POP(0.4078), Random(0.5674), SVD(0.6281), UAUM(0.6749), UNPOP(0.7040), SPR(0.7369), SerRec(0.7614)
	SDNet [4]	Serendipity factor (SRDP@k)	SRDP@5: SDNet(0.678), POP(0.051), Random(0.140), SVD(0.213), UAUM(0.204), SPR(0.268), SerRec(0.226), SOG(0.230)
	SC-CF [64]	Unexpectedness, Diversity	NA
Taobao Serendipity	TUR [65]	Unexpectedness (MAE), (RMSE)	MAE: TUR(0.3267), PURS(0.5260), SOG(0.3831) RMSE: TUR(0.4083), PURS(0.6373), SOG(0.4716)
	Ser-CF [63]	Serendipity	Serendipity: HOT(1062.11), Rel-CF(1142.97), Nov-CF(1182.73), Ser-CF(1306.05)
SerenLens	SerenEnhance [57]	$HR_{seren}@k$, $NDCG_{seren}@k$	$NDCG_{seren}@10$: SerenEnhance: 0.364, PURS: 0.170, DESR: 0.178, SNPR: 0.192; $HR_{seren}@10$: SerenEnhance: 45.63%, PURS: 38.81%, DESR: 36.67%, SNPR: 32.20%
	SerenCDR-MF [34]	$HR_{seren}@k$, $NDCG_{seren}@k$	SerenCDR-MF $HR_{seren}@10$: 31.71% $NDCG_{seren}@10$: 0.162
	SerenCDR-BERT [34]	$HR_{seren}@k$, $NDCG_{seren}@k$	SerenCDR-BERT $HR_{seren}@10$: 35.89% $NDCG@10$: 0.182
	SerenPrompt [1]	$HR_{seren}@k$, $NDCG_{seren}@k$	$NDCG_{seren}@10$: SerenPrompt(0.398), SerenEnhance(0.364), PURS(0.170), DESR(0.178), SNPR(0.192), RAND(0.043); $HR_{seren}@10$: SerenPrompt(46.15%), SerenEnhance(45.63%), PURS(38.81%), DESR(36.67%), SNPR(32.20%), RAND(9.16%)

Table 6: Serendipity-Oriented Dataset and Evaluation Metrics for serendipity research

Besides these three serendipity-oriented datasets, several smaller closed datasets from user studies exist, where researchers label which recommendations were serendipitous. For example, Hasan *et al.* extracted data from **Goodreads** and manually annotated 449 book-reading events for 4 users [44], illustrating the difficulty of obtaining serendipity labels at scale. Consequently, some researchers collaborate with industrial companies to train and evaluate on proprietary datasets, though their methods and ideas are still considered valuable [6], [39], [42].

¹<https://grouplens.org/datasets/serendipity-2018/>

²<https://github.com/greenblue96/Taobao-Serendipity-Dataset>

³<https://github.com/zhefu2/SerenLens>

Though serendipity-oriented datasets are few and limited in scope, they are useful for validation but insufficient for training complex models. Hence, researchers often rely on general RS datasets combined with proxy methods for serendipity evaluation. This approach allows for large-scale experimentation and benchmarking, although it requires careful and explicit definitions of what constitutes a "serendipitous" outcome.

Among these, the **MovieLens**⁴ datasets are arguably the most widely used in RS research. These datasets contain millions of movie ratings from thousands of users, providing a rich, stable benchmark for evaluating recommendation algorithms. While MovieLens lacks explicit serendipity labels, its rich user-item interaction matrix allows researchers to simulate serendipity through various proxy metrics. This typically involves defining serendipity as a combination of relevance and a computationally derived measure of unexpectedness or novelty [4], [31]. In SDNet, Ziarani *et al.* defining a Serendipity Factor (SRDP) where unexpectedness is calculated as an inverse function of item popularity. This approach treats less popular items as inherently more surprising, a common proxy for novelty [4]. Peng *et al.* take a different approach with CHESTNUT, evaluating the unexpectedness of their recommendations by comparing their output against mainstream memory-based collaborative filtering techniques [66]. Li *et al.* in their DESR framework operationalizes unexpectedness as a measurable deviation from a user's established taste profile by calculating the Euclidean distance between recommended items and the user's historical profile in an embedding space [3].

Other prominent datasets used for proxy-based evaluation include the **Amazon Reviews Data**⁵, valued for its scale and rich textual content, and the **MIND (Microsoft News dataset)**⁶, which provides a massive corpus for studying serendipity in the dynamic context of news recommendation.

5.1.2 Metrics

Evaluating serendipity in RS is inherently complex, as it involves both objective and subjective dimensions. Objective metrics often focus on computational approximations of serendipity, while subjective metrics rely on user feedback to capture the true essence of a "happy surprise."

Objective Metrics Objective metrics attempt to computationally quantify serendipity by decomposing it into measurable components. The most widely adopted framework models serendipity as a function of two core dimensions: *relevance* (how well the item matches user preferences) and *unexpectedness* (how surprising the item is given the user's history). This decomposition enables researchers to leverage existing accuracy metrics while introducing novel measures for surprise.

The most common approach defines serendipity as:

$$Serendipity(u, i) = Relevance(u, i) \times Unexpectedness(u, i) \quad (1)$$

where relevance is typically measured through traditional metrics like rating prediction accuracy or ranking quality, while unexpectedness is operationalized in various ways. Popular unexpectedness measures include: 1) *Popularity-based*: treating less popular items as more unexpected [4], computed as $Unexpectedness(i) = 1 - \frac{popularity(i)}{max_popularity}$; 2) *Distance-based*: measuring the semantic or embedding distance between recommended items and the user's historical profile [3]; and 3) *Prediction-based*: using the inverse of a baseline model's confidence as a proxy for surprise [66].

Building on this foundation, researchers have developed various serendipity-specific metrics. The *Serendipity Factor* (SRDP@k) combines relevance with inverse popularity weighting [4], while $NDCG_{seren}@k$ adapts the normalized discounted cumulative gain metric to use serendipity labels instead of relevance ratings [24]. The $Hit Rate_{seren}@k$ measures the proportion of recommended items that users explicitly rated as serendipitous [57].

More sophisticated approaches recognize serendipity as a multi-faceted construct. Chen *et al.* define serendipity through three components: relevance, novelty, and unexpectedness, requiring items to score highly on all dimensions [53]. Xu *et al.* operationalize serendipity as items with low predicted interest but high predicted satisfaction, explicitly modeling the "pleasant surprise" aspect [25].

⁴<https://grouplens.org/datasets/movielens/>

⁵<https://amazon-reviews-2023.github.io/>

⁶<http://msnews.github.io/>

Subjective Metrics Despite the convenience of objective metrics, the subjective nature of serendipity necessitates direct user evaluation. Subjective metrics capture the phenomenological experience of serendipity through explicit user feedback, typically collected via surveys, interviews, or interactive studies.

The most direct approach involves post-consumption surveys where users rate recommendations on serendipity-related dimensions. Kotkov *et al.* pioneered this methodology by asking MovieLens users binary questions about whether specific movies were "serendipitous" [62]. More nuanced approaches decompose the experience: Kim *et al.* use multi-item Likert scales to separately measure perceived novelty, unexpectedness, and satisfaction [67].

Recognizing the fragmentation in measurement approaches, Binst recently proposed the "Serendipionaire"—a standardized instrument for measuring perceived serendipity [36]. This tool addresses the lack of consistency across studies and provides validated questions that capture the multidimensional nature of serendipitous experiences.

Interactive evaluation methods offer real-time feedback during the recommendation process. Gyewon Jeon *et al.* developed ISER, which allows users to provide immediate binary feedback on recommendation components, enabling dynamic adjustment of serendipity-relevance trade-offs [37]. Similarly, Inoue *et al.* employ interactive genetic algorithms where user feedback serves as a fitness function, iteratively evolving toward more serendipitous recommendations [54].

5.2 Online Evaluation

Given the inherent limitations of offline proxy metrics, the research community increasingly emphasizes that the ultimate arbiter of serendipity is the user’s own subjective experience [6], [36]. As recent work has underscored, serendipity is an emergent property that resides in the user’s mind, making online, user-centric evaluation a crucial and indispensable methodology [36]. This approach moves beyond computational approximation to directly measure key indicators of a "happy surprise," such as user satisfaction, engagement, and perceived novelty. Online evaluations are primarily conducted through two main avenues: large-scale A/B testing, which is the standard in industry, and controlled user studies, which are more common in academic research.

5.2.1 A/B Testing

In production environments, A/B testing provides business-critical evidence of a model’s real-world impact. This method involves deploying a serendipity-enhancing algorithm to a treatment group of live users and comparing their behavioral metrics against a control group. Rather than relying on explicit user feedback, the success of a serendipity-enhancing model is measured through its influence on implicit behavioral signals and key business metrics. This approach is invaluable as it demonstrates not only that an algorithm can generate serendipitous items but also how users actually respond to them in a natural setting. Two recent studies from large-scale e-commerce platforms, JD.com and Taobao, provide powerful examples of this methodology in practice.

Wang *et al.* (2023) deployed a personalized serendipity framework on JD.com and observed a +0.8% increase in aggregate user clicks and a +3.23% increase in interactions with novel items [42]. While these figures may appear modest in isolation, such lifts are considered substantial in a large-scale recommender system, representing significant gains in user engagement.

Even more striking results were reported by Xi *et al.* (2025) from the deployment of SERAL, an LLM-based framework on Taobao. Their online experiments showed that the framework improved the click-through rate on serendipitous items by 29.56% and increased transactions on such items by 27.6%. Critically, this boost in exploratory behavior translated directly to overall business value, increasing the total transaction volume by 0.98% and the aggregate CTR by 0.05pt [39]. These industrial-scale online experiments demonstrate conclusively that systematically engineering for serendipity is not merely an academic exercise; it can positively and measurably impact user engagement, platform revenue, and long-term user satisfaction.

5.2.2 User Study

In academic research, where large-scale, continuously running systems for A/B testing are typically unavailable, controlled user studies remain the primary method for user-centric evaluation. This approach is essential for gathering authentic, explicit feedback by presenting recommendations to real users and directly asking them about their serendipitous experiences. These studies vary widely in their design and objectives, from collecting ground-truth data to testing complex theoretical models of user psychology.

A foundational goal of user studies is to collect ground-truth data for what users perceive as serendipitous. The work of Kotkov *et al.* (2018) is a prime example, where they systematically surveyed MovieLens users about their reactions to specific movies, breaking down the concept into components like novelty and unexpectedness. This led to the creation of the invaluable Serendipity-2018 dataset, which provides a public benchmark for validating offline metrics [62]. Similarly, the large-scale survey by Chen *et al.* (2019) on the Taobao platform provided crucial, in-the-moment user feedback, linking perceptions of serendipity to factors like timeliness and user curiosity [63].

Beyond dataset creation, user studies are critical for testing hypotheses about the impact of serendipity on the overall user experience. For example, Zhao *et al.* (2022) conducted a survey with 401 TikTok users to investigate how serendipity contributes to the platform’s ability to create a “flow experience”—a state of deep user immersion and engagement. Using a multi-item scale to measure “perceived recommendation serendipity,” they found it to be the strongest predictor of flow. Their study also revealed a crucial moderator effect: the positive influence of serendipity was significantly stronger for high-experience users, whereas recommendation accuracy was more important for new users [2]. This demonstrates how user studies can uncover nuanced, dynamic relationships that offline analysis would miss. Furthermore, user studies are used to build and validate causal models of user behavior. Kwon *et al.* (2020) in a survey of 212 Netflix subscribers, employed structural equation modeling to test a model of user retention. They operationalized “perceived serendipity” as a key variable and demonstrated that it significantly increases “decision satisfaction,” which in turn is a strong predictor of users’ “intention to continue a subscription.” This work provides strong academic evidence linking the subjective experience of serendipity directly to a critical business outcome [28]. User studies in controlled lab settings can be used to dissect the psychological mechanisms that underpin serendipity. Kim *et al.* (2021) conducted a series of experiments that systematically manipulated the context of recommendations. Their findings reveal the core ingredients of a serendipitous encounter: it must be perceived as positive (a negative surprise is not serendipitous), unexpected, and, crucially, involving a degree of chance or randomness [67]. When an encounter was framed as being deterministically chosen by a marketer, the feeling of serendipity was diminished.

While these studies provide rich and varied insights, they also collectively highlight a central challenge: the fragmentation of measurement. As Binst (2024) notes, each study tends to use its own ad-hoc questions or scales, making direct comparison of results difficult and reinforcing the need for standardized instruments like the proposed “Serendipionnaire” [36]. Moreover, even with direct user feedback, the concept remains elusive. Kotkov *et al.* (2024) in a recent observational study, found that users’ own opinions of what was serendipitous did not align strongly with any single computational definition, exposing a “dark matter” of serendipity that current evaluation methods still fail to capture [6], [33]. Thus, while user studies are indispensable, they also confirm the complexity of the phenomenon and have their own limitations, including small sample sizes, potential experimental biases, and the inherent subjectivity of user opinions.

5.3 Future Directions and Open Challenges

The evaluation of serendipity in RS remains an active area of research characterized by significant methodological challenges and opportunities. While substantial progress has been achieved in developing beyond-accuracy metrics, several fundamental issues persist that impede the field’s maturation. This section identifies the most critical open challenges and proposes promising research directions that warrant investigation to advance both theoretical understanding and practical implementation of serendipity evaluation.

5.3.1 Standardization of Evaluation Methodologies

One of the most pressing challenges in serendipity research is the lack of methodological consensus across studies. The fragmentation of evaluation approaches has created substantial barriers to reproducibility and cross-study comparison, ultimately hindering scientific progress in the field.

The establishment of standardized evaluation protocols requires coordinated efforts across multiple dimensions. First, the development of validated psychometric instruments for measuring perceived serendipity represents a fundamental necessity. The "Serendipionnaire" proposed by Binst (2024) exemplifies this direction, offering a standardized questionnaire that could serve as the equivalent of established metrics like NDCG for ranking accuracy [36]. Such instruments would enable more rigorous user studies and facilitate meta-analyses across different research contexts.

Second, the creation of comprehensive benchmark datasets with explicit serendipity annotations across diverse domains is essential for advancing the field. While existing datasets such as Serendipity-2018 provide valuable foundations, their limited scale and domain coverage constrain their utility for training sophisticated models and conducting comprehensive evaluations. Future efforts should prioritize the development of large-scale, multi-domain datasets that capture the nuanced nature of serendipitous experiences across different application contexts including music, e-commerce, news, and educational content.

5.3.2 Reconciling Computational Proxies with Human Perception

A fundamental tension exists between the computational tractability of offline evaluation metrics and their alignment with authentic human experiences of serendipity. This gap represents one of the most significant methodological challenges in the field, as highlighted by recent research revealing substantial discrepancies between algorithmic predictions and user judgments.

Research efforts must systematically address the validation and calibration of computational proxies against empirical user data. This involves establishing correlational relationships between various operationalizations of unexpectedness—including popularity-based, distance-based, and prediction-based measures—and ground-truth user reports of surprise [62]. Such validation studies would enable the identification of more psychologically valid offline metrics and improve the reliability of large-scale evaluations.

The "dark matter" phenomenon identified by Kotkov *et al.* (2024), wherein user perceptions of serendipity diverge from computational definitions, suggests that current proxy measures capture only a limited subset of the serendipitous experience [6]. This observation indicates the need for more sophisticated modeling approaches that can account for the multifaceted and contextual nature of serendipity perception.

The emergence of Large Language Models as evaluation tools presents a promising avenue for bridging this gap. Recent investigations by Tokutake *et al.* (2024) demonstrate that LLMs can approximate human judgment with performance comparable to traditional metrics, offering a scalable approach to human-aligned evaluation [41]. Future research should explore the optimization of LLM-based evaluation through improved prompt engineering, fine-tuning strategies, and integration with domain-specific knowledge.

5.3.3 Towards Personalized and Context-Aware Evaluation

Contemporary evaluation methodologies typically report aggregate performance metrics that obscure important individual and contextual variations in serendipity perception. This approach fails to capture the inherently personal and situational nature of serendipitous experiences, limiting both theoretical understanding and practical applicability.

Research evidence increasingly demonstrates that serendipity preferences vary significantly across user segments. The findings of Zhao *et al.* (2022) regarding differential serendipity effects for experienced versus novice users exemplify this heterogeneity [2]. Future evaluation frameworks should incorporate user segmentation based on relevant characteristics such as domain expertise, curiosity levels, and platform experience to provide more nuanced performance assessments.

Similarly, contextual factors significantly influence serendipity perception, as demonstrated by Kim *et*

al. (2021), who showed that the perceived randomness and valence of encounters critically determine serendipitous experiences [67]. Advanced evaluation methodologies should account for these contextual dimensions, incorporating factors such as temporal context, user mood, task urgency, and social setting into their assessment frameworks.

5.3.4 Longitudinal and Multi-Objective Evaluation Paradigms

Current evaluation practices predominantly focus on immediate, isolated interactions, failing to capture the long-term effects and multi-objective nature of serendipitous recommendations. This limitation represents a significant gap between evaluation methodologies and the ultimate goals of RS.

The development of longitudinal evaluation frameworks is essential for understanding how serendipitous encounters influence user behavior, satisfaction, and retention over extended periods. While industrial A/B testing studies [39], [42] have provided initial evidence of long-term benefits, comprehensive academic investigations of these effects remain limited. Future research should prioritize longitudinal user studies that track the evolution of user preferences, engagement patterns, and satisfaction as functions of exposure to serendipitous recommendations.

Furthermore, serendipity evaluation should be integrated within multi-objective optimization frameworks that explicitly model trade-offs with other system objectives such as accuracy, diversity, and fairness. Rather than treating accuracy-serendipity trade-offs as limitations, evaluation methodologies should characterize and optimize these relationships to provide actionable insights for system designers. This approach would enable more sophisticated performance characterizations that better reflect the complex objectives of real-world RS.

6 Real-World Initiatives: RecSys Challenge

Recent research and practical developments have revealed the limitations of accuracy-centric approaches [68], [69], as they do not reflect long-term user satisfaction. Therefore, recent ACM RecSys Challenges have incorporated beyond-accuracy metrics, including serendipity, novelty, and diversity [68]–[70]. RecSys Challenge 2024 included serendipity as one of their beyond-accuracy metrics [68], while RecSys Challenge 2025 did not score serendipity explicitly but retained a weighted score combining novelty and diversity with accuracy [69]. The novelty and diversity components capture unfamiliar yet relevant items, which embodies the core idea of serendipity. Together, these challenges illustrate how industry companies are researching user-centric metrics to enhance overall recommendation performance. Both challenges provide open, production-scale datasets that help researchers test their serendipitous RS approaches, bridging the gap between academic research and industry practice.

References

- [1] Z. Fu *et al.*, “The art of asking: Prompting large language models for serendipity recommendations,” in *Proceedings of the 2024 ACM SIGIR International Conference on Theory of Information Retrieval*, Washington DC USA: ACM, Aug. 2, 2024, pp. 157–166, ISBN: 979-8-4007-0681-3. DOI: [10.1145/3664190.3672521](https://doi.org/10.1145/3664190.3672521). URL: <https://dl.acm.org/doi/10.1145/3664190.3672521>.
- [2] H. Zhao *et al.*, “How TikTok leads users to flow experience: Investigating the effects of technology affordances with user experience level and video length as moderators,” *Internet Research*, vol. 33, no. 2, pp. 820–849, Jul. 6, 2022, Publisher: Emerald Publishing Limited, ISSN: 1066-2243. DOI: [10.1108/INTR-08-2021-0595](https://doi.org/10.1108/INTR-08-2021-0595). URL: <https://www.emerald.com/insight/content/doi/10.1108/intr-08-2021-0595/full/html>.
- [3] X. Li *et al.*, “Directional and explainable serendipity recommendation,” in *Proceedings of The Web Conference 2020*, Taipei Taiwan: ACM, Apr. 20, 2020, pp. 122–132, ISBN: 978-1-4503-7023-3. DOI: [10.1145/3366423.3380100](https://doi.org/10.1145/3366423.3380100). URL: <https://dl.acm.org/doi/10.1145/3366423.3380100>.
- [4] R. J. Ziarani *et al.*, “Deep neural network approach for a serendipity-oriented recommendation system,” *Expert Systems with Applications*, vol. 185, p. 115660, Dec. 2021, ISSN: 09574174. DOI: [10.1016/j.eswa.2021.115660](https://doi.org/10.1016/j.eswa.2021.115660). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0957417421010514>.
- [5] Z. Fu *et al.*, “Deep learning models for serendipity recommendations: A survey and new perspectives,” *ACM Comput. Surv.*, vol. 56, no. 1, pp. 19:1–19:26, Aug. 26, 2023, ISSN: 0360-0300. DOI: [10.1145/3605145](https://doi.org/10.1145/3605145). URL: <https://dl.acm.org/doi/10.1145/3605145>.
- [6] D. Kotkov *et al.*, “The dark matter of serendipity in recommender systems,” in *Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval*, Sheffield United Kingdom: ACM, Mar. 10, 2024, pp. 108–118, ISBN: 979-8-4007-0434-5. DOI: [10.1145/3627508.3638342](https://doi.org/10.1145/3627508.3638342). URL: <https://dl.acm.org/doi/10.1145/3627508.3638342>.
- [7] E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Group, The, Apr. 2011, 304 pp., ISBN: 978-1-59420-300-8.
- [8] J. L. Herlocker *et al.*, “Evaluating collaborative filtering recommender systems,” *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 5–53, Jan. 2004, ISSN: 1046-8188, 1558-2868. DOI: [10.1145/963770.963772](https://doi.org/10.1145/963770.963772). URL: <https://dl.acm.org/doi/10.1145/963770.963772>.
- [9] L. Iaquinta *et al.*, “Introducing serendipity in a content-based recommender system,” in *2008 Eighth International Conference on Hybrid Intelligent Systems*, Sep. 2008, pp. 168–173. DOI: [10.1109/HIS.2008.25](https://doi.org/10.1109/HIS.2008.25). URL: <https://ieeexplore.ieee.org/document/4626624/>.
- [10] M. Ge *et al.*, “Beyond accuracy: Evaluating recommender systems by coverage and serendipity,” in *Proceedings of the fourth ACM conference on Recommender systems*, Barcelona Spain: ACM, Sep. 26, 2010, pp. 257–260, ISBN: 978-1-60558-906-0. DOI: [10.1145/1864708.1864761](https://doi.org/10.1145/1864708.1864761). URL: <https://dl.acm.org/doi/10.1145/1864708.1864761>.
- [11] Q. Lu *et al.*, “Serendipitous personalized ranking for top-n recommendation,” in *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, vol. 1, Dec. 2012, pp. 258–265. DOI: [10.1109/WI-IAT.2012.135](https://doi.org/10.1109/WI-IAT.2012.135). URL: <https://ieeexplore.ieee.org/document/6511894/>.
- [12] M. Manca *et al.*, “Behavioral data mining to produce novel and serendipitous friend recommendations in a social bookmarking system,” *Information Systems Frontiers*, vol. 20, no. 4, pp. 825–839, Oct. 26, 2015, ISSN: 1572-9419. DOI: [10.1007/s10796-015-9600-3](https://doi.org/10.1007/s10796-015-9600-3). URL: <https://doi.org/10.1007/s10796-015-9600-3>.
- [13] M. Kaminskas *et al.*, “Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems,” *ACM Transactions on Interactive Intelligent Systems*, vol. 7, no. 1, pp. 1–42, Mar. 31, 2017, ISSN: 2160-6455, 2160-6463. DOI: [10.1145/2926720](https://doi.org/10.1145/2926720). URL: <https://dl.acm.org/doi/10.1145/2926720>.
- [14] D. Kotkov *et al.*, “A survey of serendipity in recommender systems,” *Knowledge-Based Systems*, vol. 111, pp. 180–192, Nov. 2016, ISSN: 09507051. DOI: [10.1016/j.knosys.2016.08.014](https://doi.org/10.1016/j.knosys.2016.08.014). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0950705116302763>.
- [15] F. Paiva *et al.*, “A serendipity-based approach to enhance particle swarm optimization using scout particles,” *IEEE Latin America Transactions*, vol. 15, no. 6, pp. 1101–1112, Jun. 2017, ISSN: 1548-0992. DOI: [10.1109/TLA.2017.7932698](https://doi.org/10.1109/TLA.2017.7932698). URL: <https://ieeexplore.ieee.org/document/7932698/>.

- [16] A. H. Afridi, "Visualizing serendipitous recommendations in user controlled recommender system for learning," *Procedia Computer Science*, vol. 141, pp. 496–502, 2018, ISSN: 18770509. DOI: [10.1016/j.procs.2018.10.136](https://doi.org/10.1016/j.procs.2018.10.136). URL: <https://linkinghub.elsevier.com/retrieve/pii/S187705091831785X>.
- [17] Y. Yang *et al.*, "Improving existing collaborative filtering recommendations via serendipity-based algorithm," *IEEE Transactions on Multimedia*, vol. 20, no. 7, pp. 1888–1900, Jul. 2018, ISSN: 1941-0077. DOI: [10.1109/TMM.2017.2779043](https://doi.org/10.1109/TMM.2017.2779043). URL: <https://ieeexplore.ieee.org/document/8125125/>.
- [18] Y. Xu *et al.*, "Neural serendipity recommendation: Exploring the balance between accuracy and novelty with sparse explicit feedback," *ACM Transactions on Knowledge Discovery from Data*, vol. 14, no. 4, pp. 1–25, Aug. 31, 2020, ISSN: 1556-4681, 1556-472X. DOI: [10.1145/3396607](https://doi.org/10.1145/3396607). URL: <https://dl.acm.org/doi/10.1145/3396607>.
- [19] C. Nishioka *et al.*, "Research paper recommender system with serendipity using tweets vs. diversification," in *Digital Libraries at the Crossroads of Digital Information for the Future*, A. Jatowt *et al.*, Eds., Cham: Springer International Publishing, 2019, pp. 63–70, ISBN: 978-3-030-34058-2. DOI: [10.1007/978-3-030-34058-2_7](https://doi.org/10.1007/978-3-030-34058-2_7).
- [20] S. A. Cerri *et al.*, "Serendipitous learning fostered by brain state assessment and collective wisdom," in *Brain Function Assessment in Learning*, C. Frasson *et al.*, Eds., Cham: Springer International Publishing, 2020, pp. 125–136, ISBN: 978-3-030-60735-7. DOI: [10.1007/978-3-030-60735-7_14](https://doi.org/10.1007/978-3-030-60735-7_14).
- [21] M. Zhang *et al.*, "SNPR: A serendipity-oriented next POI recommendation model," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Virtual Event Queensland Australia: ACM, Oct. 26, 2021, pp. 2568–2577, ISBN: 978-1-4503-8446-9. DOI: [10.1145/3459637.3482394](https://doi.org/10.1145/3459637.3482394). URL: <https://dl.acm.org/doi/10.1145/3459637.3482394>.
- [22] R. Shrivastava *et al.*, "On diverse and serendipitous item recommendation: A reinforced similarity and multi-objective optimization-based composite recommendation framework," in *Machine Intelligence Techniques for Data Analysis and Signal Processing*, D. S. Sisodia *et al.*, Eds., Singapore: Springer Nature, 2023, pp. 1–13, ISBN: 978-981-99-0085-5. DOI: [10.1007/978-981-99-0085-5_1](https://doi.org/10.1007/978-981-99-0085-5_1).
- [23] Y. Ping *et al.*, "Beyond accuracy measures: The effect of diversity, novelty and serendipity in recommender systems on user engagement," *Electronic Commerce Research*, Feb. 18, 2024, ISSN: 1572-9362. DOI: [10.1007/s10660-024-09813-w](https://doi.org/10.1007/s10660-024-09813-w). URL: <https://doi.org/10.1007/s10660-024-09813-w>.
- [24] G. Pandey *et al.*, "Recommending serendipitous items using transfer learning," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, ser. CIKM '18, New York, NY, USA: Association for Computing Machinery, Oct. 17, 2018, pp. 1771–1774, ISBN: 978-1-4503-6014-2. DOI: [10.1145/3269206.3269268](https://doi.org/10.1145/3269206.3269268). URL: <https://dl.acm.org/doi/10.1145/3269206.3269268>.
- [25] Y. Xu *et al.*, "GS-RS: A generative approach for alleviating cold start and filter bubbles in recommender systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 2, pp. 668–681, Feb. 2024, Conference Name: IEEE Transactions on Knowledge and Data Engineering, ISSN: 1558-2191. DOI: [10.1109/TKDE.2023.3290140](https://doi.org/10.1109/TKDE.2023.3290140). URL: <https://ieeexplore.ieee.org/document/10167850/?arnumber=10167850>.
- [26] R. J. Ziarani *et al.*, "Serendipity in recommender systems: A systematic literature review," *Journal of Computer Science and Technology*, vol. 36, no. 2, pp. 375–396, Apr. 1, 2021, ISSN: 1860-4749. DOI: [10.1007/s11390-020-0135-9](https://doi.org/10.1007/s11390-020-0135-9). URL: <https://doi.org/10.1007/s11390-020-0135-9>.
- [27] Y. Kim *et al.*, "Design of a serendipity-incorporated recommender system," *Electronics*, vol. 14, no. 4, p. 821, Jan. 2025, Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2079-9292. DOI: [10.3390/electronics14040821](https://doi.org/10.3390/electronics14040821). URL: <https://www.mdpi.com/2079-9292/14/4/821>.
- [28] Y. Kwon *et al.*, "Accurately or accidentally? recommendation agent and search experience in over-the-top (OTT) services," *Internet Research*, vol. 31, no. 2, pp. 562–586, Oct. 26, 2020, Publisher: Emerald Publishing Limited, ISSN: 1066-2243. DOI: [10.1108/INTR-03-2020-0127](https://doi.org/10.1108/INTR-03-2020-0127). URL: <https://www.emerald.com/insight/content/doi/10.1108/intr-03-2020-0127/full/html>.
- [29] N. Wang *et al.*, "How do item features and user characteristics affect users' perceptions of recommendation serendipity? a cross-domain analysis," *User Modeling and User-Adapted Interaction*, vol. 33, no. 3, pp. 727–765, Jul. 1, 2023, ISSN: 1573-1391. DOI: [10.1007/s11257-022-09350-x](https://doi.org/10.1007/s11257-022-09350-x). URL: <https://doi.org/10.1007/s11257-022-09350-x>.
- [30] N. Wang *et al.*, "The impacts of item features and user characteristics on users' perceived serendipity of recommendations," in *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and*

- Personalization*, Genoa Italy: ACM, Jul. 7, 2020, pp. 266–274, ISBN: 978-1-4503-6861-2. DOI: [10.1145/3340631.3394863](https://doi.org/10.1145/3340631.3394863). URL: <https://dl.acm.org/doi/10.1145/3340631.3394863>.
- [31] Z. Xu *et al.*, “A serendipitous recommendation system considering user curiosity,” in *Information Integration and Web Intelligence*, P. Delir Haghighi *et al.*, Eds., Cham: Springer Nature Switzerland, 2025, pp. 33–48, ISBN: 978-3-031-78093-6. DOI: [10.1007/978-3-031-78093-6_3](https://doi.org/10.1007/978-3-031-78093-6_3).
 - [32] Z. Fu *et al.*, “Modeling users’ curiosity in recommender systems,” *ACM Transactions on Knowledge Discovery from Data*, vol. 18, no. 1, pp. 1–23, Jan. 31, 2024, Publisher: Association for Computing Machinery (ACM), ISSN: 1556-4681, 1556-472X. DOI: [10.1145/3617598](https://doi.org/10.1145/3617598). URL: <https://dl.acm.org/doi/10.1145/3617598>.
 - [33] D. Kotkov *et al.*, “Rethinking serendipity in recommender systems,” in *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*, Austin TX USA: ACM, Mar. 19, 2023, pp. 383–387, ISBN: 979-8-4007-0035-4. DOI: [10.1145/3576840.3578310](https://doi.org/10.1145/3576840.3578310). URL: <https://dl.acm.org/doi/10.1145/3576840.3578310>.
 - [34] Z. Fu *et al.*, “A deep learning model for cross-domain serendipity recommendations,” *ACM Transactions on Recommender Systems*, vol. 3, no. 3, pp. 1–21, Sep. 30, 2025, ISSN: 2770-6699. DOI: [10.1145/3690654](https://doi.org/10.1145/3690654). URL: <https://dl.acm.org/doi/10.1145/3690654>.
 - [35] Z. A. Pardos *et al.*, “Designing for serendipity in a university course recommendation system,” in *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, ser. LAK ’20, New York, NY, USA: Association for Computing Machinery, Mar. 23, 2020, pp. 350–359, ISBN: 978-1-4503-7712-6. DOI: [10.1145/3375462.3375524](https://doi.org/10.1145/3375462.3375524). URL: <https://dl.acm.org/doi/10.1145/3375462.3375524>.
 - [36] B. Binst, “How to evaluate serendipity in recommender systems: The need for a serendipionnaire,” in *18th ACM Conference on Recommender Systems*, Bari Italy: ACM, Oct. 8, 2024, pp. 1335–1341, ISBN: 979-8-4007-0505-2. DOI: [10.1145/3640457.3688017](https://doi.org/10.1145/3640457.3688017). URL: <https://dl.acm.org/doi/10.1145/3640457.3688017>.
 - [37] Gyewon Jeon *et al.*, “Interactive feedback loop with counterfactual data modification for serendipity in a recommendation system,” *International Journal of Human-Computer Interaction*, vol. 40, no. 19, pp. 5585–5601, Oct. 1, 2024, Publisher: Taylor & Francis, ISSN: 1044-7318. DOI: [10.1080/10447318.2023.2238369](https://doi.org/10.1080/10447318.2023.2238369). URL: <https://doi.org/10.1080/10447318.2023.2238369>.
 - [38] T. Kamba, “SerendipitySeeker: A novel SNS viewer designed to broaden perspectives by encountering diverse information,” in *HCI International 2024 – Late Breaking Papers*, A. Coman *et al.*, Eds., Cham: Springer Nature Switzerland, 2025, pp. 190–200, ISBN: 978-3-031-76806-4. DOI: [10.1007/978-3-031-76806-4_15](https://doi.org/10.1007/978-3-031-76806-4_15).
 - [39] Y. Xi *et al.*, *Bursting filter bubble: Enhancing serendipity recommendations with aligned large language models*, Feb. 19, 2025. DOI: [10.48550/arXiv.2502.13539](https://doi.org/10.48550/arXiv.2502.13539). arXiv: [2502.13539\[cs\]](https://arxiv.org/abs/2502.13539). URL: <http://arxiv.org/abs/2502.13539>.
 - [40] J. Wang *et al.*, “Sparks of surprise: Multi-objective recommendations with hierarchical decision transformers for diversity, novelty, and serendipity,” in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, Boise ID USA: ACM, Oct. 21, 2024, pp. 2358–2368. DOI: [10.1145/3627673.3679533](https://doi.org/10.1145/3627673.3679533). URL: <https://dl.acm.org/doi/10.1145/3627673.3679533>.
 - [41] Y. Tokutake *et al.*, “Can large language models assess serendipity in recommender systems?” *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 28, no. 6, pp. 1263–1272, Nov. 2024, Publisher: Fuji Technology Press Ltd., ISSN: 1343-0130. DOI: [10.20965/jaciii.2024.p1263](https://doi.org/10.20965/jaciii.2024.p1263). URL: <https://cir.nii.ac.jp/crid/1390020762779347200>.
 - [42] Z. Wang *et al.*, “An industrial framework for personalized serendipitous recommendation in e-commerce,” in *Proceedings of the 17th ACM Conference on Recommender Systems*, Singapore Singapore: ACM, Sep. 14, 2023, pp. 1015–1018, ISBN: 979-8-4007-0241-9. DOI: [10.1145/3604915.3610234](https://doi.org/10.1145/3604915.3610234). URL: <https://dl.acm.org/doi/10.1145/3604915.3610234>.
 - [43] S. Z. U. Hassan *et al.*, “GCZRec: Generative collaborative zero-shot framework for cold start news recommendation,” *IEEE Access*, vol. 12, pp. 16 610–16 620, 2024, ISSN: 2169-3536. DOI: [10.1109/ACCESS.2024.3359053](https://doi.org/10.1109/ACCESS.2024.3359053). URL: <https://ieeexplore.ieee.org/document/10414986>.
 - [44] T. Hasan *et al.*, “Topic-level bayesian surprise and serendipity for recommender systems,” in *Proceedings of the 17th ACM Conference on Recommender Systems*, ser. RecSys ’23, New York, NY, USA: Association for Computing Machinery, Sep. 14, 2023, pp. 933–939, ISBN: 979-8-4007-0241-9. DOI: [10.1145/3604915.3608851](https://doi.org/10.1145/3604915.3608851). URL: <https://dl.acm.org/doi/10.1145/3604915.3608851>.

- [45] F. Ricci *et al.*, “Recommender systems: Techniques, applications, and challenges,” in *Recommender Systems Handbook*, F. Ricci *et al.*, Eds., New York, NY: Springer US, 2022, pp. 1–35, ISBN: 978-1-0716-2197-4. DOI: [10.1007/978-1-0716-2197-4_1](https://doi.org/10.1007/978-1-0716-2197-4_1). URL: https://doi.org/10.1007/978-1-0716-2197-4_1.
- [46] X. Su *et al.*, “A survey of collaborative filtering techniques,” *Advances in Artificial Intelligence*, vol. 2009, no. 1, p. 421 425, 2009, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1155/2009/421425>, ISSN: 1687-7489. DOI: [10.1155/2009/421425](https://doi.org/10.1155/2009/421425). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2009/421425>.
- [47] I. Coppens *et al.*, “Investigating different recommender algorithms in the domain of physical activity recommendations: A longitudinal between-subjects user study,” *User Modeling and User-Adapted Interaction*, vol. 35, no. 1, p. 6, Feb. 18, 2025, ISSN: 1573-1391. DOI: [10.1007/s11257-025-09427-3](https://doi.org/10.1007/s11257-025-09427-3). URL: <https://doi.org/10.1007/s11257-025-09427-3>.
- [48] D. Kotkov *et al.*, “How does serendipity affect diversity in recommender systems? a serendipity-oriented greedy algorithm,” *Computing*, vol. 102, no. 2, pp. 393–411, Feb. 1, 2020, ISSN: 1436-5057. DOI: [10.1007/s00607-018-0687-5](https://doi.org/10.1007/s00607-018-0687-5). URL: <https://doi.org/10.1007/s00607-018-0687-5>.
- [49] P. B. Thorat *et al.*, “Survey on collaborative filtering, content-based filtering and hybrid recommendation system,” *International Journal of Computer Applications*, vol. 110, no. 4, pp. 31–36, Jan. 16, 2015, Publisher: Foundation of Computer Science, ISSN: 0975-8887. DOI: [10.5120/19308-0760](https://doi.org/10.5120/19308-0760). URL: <http://research.ijcaonline.org/volume110/number4/pxc3900760.pdf>.
- [50] O. Stitini *et al.*, “An improved recommender system solution to mitigate the over-specialization problem using genetic algorithms,” *Electronics*, vol. 11, no. 2, p. 242, Jan. 2022, Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2079-9292. DOI: [10.3390/electronics11020242](https://doi.org/10.3390/electronics11020242). URL: <https://www.mdpi.com/2079-9292/11/2/242>.
- [51] S. Zhang *et al.*, “Deep learning based recommender system: A survey and new perspectives,” *ACM Computing Surveys*, vol. 52, no. 1, pp. 1–38, Jan. 31, 2020, Publisher: Association for Computing Machinery (ACM), ISSN: 0360-0300, 1557-7341. DOI: [10.1145/3285029](https://doi.org/10.1145/3285029). URL: <https://dl.acm.org/doi/10.1145/3285029>.
- [52] P. Li *et al.*, “PURS: Personalized unexpected recommender system for improving user satisfaction,” in *Fourteenth ACM Conference on Recommender Systems*, Virtual Event Brazil: ACM, Sep. 22, 2020, pp. 279–288, ISBN: 978-1-4503-7583-2. DOI: [10.1145/3383313.3412238](https://doi.org/10.1145/3383313.3412238). URL: <https://dl.acm.org/doi/10.1145/3383313.3412238>.
- [53] X. Chen *et al.*, “Cognitive-based knowledge learning framework for recommendation,” *Knowledge-Based Systems*, vol. 287, p. 111 446, Mar. 2024, Publisher: Elsevier BV, ISSN: 0950-7051. DOI: [10.1016/j.knsys.2024.111446](https://doi.org/10.1016/j.knsys.2024.111446). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0950705124000819>.
- [54] S. Inoue *et al.*, “Leveraging interactive evolutionary computation to induce serendipity in informal learning,” *Multimodal Technologies and Interaction*, vol. 8, no. 11, p. 103, Nov. 2024, Number: 11 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2414-4088. DOI: [10.3390/mti8110103](https://doi.org/10.3390/mti8110103). URL: <https://www.mdpi.com/2414-4088/8/11/103>.
- [55] H. I. Pohan *et al.*, “Recommender system using transformer model: A systematic literature review,” in *2022 1st International Conference on Information System & Information Technology (ICISIT)*, Jul. 2022, pp. 376–381. DOI: [10.1109/ICISIT54091.2022.9873070](https://doi.org/10.1109/ICISIT54091.2022.9873070). URL: <https://ieeexplore.ieee.org/document/9873070/>.
- [56] F. Sun *et al.*, “BERT4rec: Sequential recommendation with bidirectional encoder representations from transformer,” in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing China: ACM, Nov. 3, 2019, pp. 1441–1450, ISBN: 978-1-4503-6976-3. DOI: [10.1145/3357384.3357895](https://doi.org/10.1145/3357384.3357895). URL: <https://dl.acm.org/doi/10.1145/3357384.3357895>.
- [57] Z. Fu *et al.*, “Wisdom of crowds and fine-grained learning for serendipity recommendations,” in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Taipei Taiwan: ACM, Jul. 19, 2023, pp. 739–748, ISBN: 978-1-4503-9408-6. DOI: [10.1145/3539618.3591787](https://doi.org/10.1145/3539618.3591787). URL: <https://dl.acm.org/doi/10.1145/3539618.3591787>.
- [58] L. Li *et al.*, *Large language models for generative recommendation: A survey and visionary discussions*, Mar. 23, 2024. DOI: [10.48550/arXiv.2309.01157](https://doi.org/10.48550/arXiv.2309.01157). arXiv: [2309.01157\[cs\]](https://arxiv.org/abs/2309.01157). URL: <http://arxiv.org/abs/2309.01157>.
- [59] Y. Hou *et al.*, “Generative recommendation models: Progress and directions,” in *Companion Proceedings of the ACM on Web Conference 2025*, Sydney NSW Australia: ACM, May 8, 2025, pp. 13–16, ISBN:

- 979-8-4007-1331-6. DOI: [10.1145/3701716.3715856](https://doi.org/10.1145/3701716.3715856). URL: <https://dl.acm.org/doi/10.1145/3701716.3715856>.
- [60] F. Zhuang *et al.*, “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, ISSN: 1558-2256. DOI: [10.1109/JPROC.2020.3004555](https://doi.org/10.1109/JPROC.2020.3004555). URL: <https://ieeexplore.ieee.org/document/9134370/>.
 - [61] S. Boo *et al.*, “Serendipity into session-based recommendation: Focusing on unexpectedness, relevance, and usefulness of recommendations,” in *28th International Conference on Intelligent User Interfaces*, Sydney NSW Australia: ACM, Mar. 27, 2023, pp. 83–86. DOI: [10.1145/3581754.3584138](https://doi.org/10.1145/3581754.3584138). URL: <https://dl.acm.org/doi/10.1145/3581754.3584138>.
 - [62] D. Kotkov *et al.*, “Investigating serendipity in recommender systems based on real user feedback,” in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, Pau France: ACM, Apr. 9, 2018, pp. 1341–1350, ISBN: 978-1-4503-5191-1. DOI: [10.1145/3167132.3167276](https://doi.org/10.1145/3167132.3167276). URL: <https://dl.acm.org/doi/10.1145/3167132.3167276>.
 - [63] L. Chen *et al.*, “How serendipity improves user satisfaction with recommendations? a large-scale user evaluation,” in *The World Wide Web Conference*, San Francisco CA USA: ACM, May 13, 2019, pp. 240–250, ISBN: 978-1-4503-6674-8. DOI: [10.1145/3308558.3313469](https://doi.org/10.1145/3308558.3313469). URL: <https://dl.acm.org/doi/10.1145/3308558.3313469>.
 - [64] A. A. Deshmukh *et al.*, “A scalable clustering algorithm for serendipity in recommender systems,” in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, ISSN: 2375-9259, Nov. 2018, pp. 1279–1288. DOI: [10.1109/ICDMW.2018.00182](https://doi.org/10.1109/ICDMW.2018.00182). URL: <https://ieeexplore.ieee.org/document/8637463/>.
 - [65] Y. Ni *et al.*, “TUR: Utilizing temporal information to make unexpected e-commerce recommendations,” in *Web Information Systems Engineering – WISE 2022*, R. Chbeir *et al.*, Eds., Cham: Springer International Publishing, 2022, pp. 553–563, ISBN: 978-3-031-20891-1. DOI: [10.1007/978-3-031-20891-1_39](https://doi.org/10.1007/978-3-031-20891-1_39).
 - [66] X. Peng *et al.*, “CHESTNUT: Improve serendipity in movie recommendation by an information theory-based collaborative filtering approach,” in *Human Interface and the Management of Information. Interacting with Information*, S. Yamamoto *et al.*, Eds., Cham: Springer International Publishing, 2020, pp. 78–95, ISBN: 978-3-030-50017-7. DOI: [10.1007/978-3-030-50017-7_6](https://doi.org/10.1007/978-3-030-50017-7_6).
 - [67] A. Kim *et al.*, “Serendipity: Chance encounters in the marketplace enhance consumer satisfaction,” *Journal of Marketing*, vol. 85, no. 4, pp. 141–157, Jul. 1, 2021, Publisher: SAGE Publications Inc, ISSN: 0022-2429. DOI: [10.1177/00222429211000344](https://doi.org/10.1177/00222429211000344). URL: <https://doi.org/10.1177/00222429211000344>.
 - [68] J. Kruse *et al.*, “RecSys challenge 2024: Balancing accuracy and editorial values in news recommendations,” in *18th ACM Conference on Recommender Systems*, Bari Italy: ACM, Oct. 8, 2024, pp. 1195–1199, ISBN: 979-8-4007-0505-2. DOI: [10.1145/3640457.3687164](https://doi.org/10.1145/3640457.3687164). URL: <https://dl.acm.org/doi/10.1145/3640457.3687164>.
 - [69] Synerise. “RecSys challenge 2025 by synerise.” (2025), URL: <https://recsys.synerise.com/>.
 - [70] J. Kruse *et al.*, “EB-NeRD a large-scale dataset for news recommendation,” in *Proceedings of the Recommender Systems Challenge 2024*, Bari Italy: ACM, Oct. 14, 2024, pp. 1–11, ISBN: 979-8-4007-1127-5. DOI: [10.1145/3687151.3687152](https://doi.org/10.1145/3687151.3687152). URL: <https://dl.acm.org/doi/10.1145/3687151.3687152>.