International Journal on Robotics, Automation and Sciences

Optimizing Reviewer Assignment with Recommender Systems: Models, Related Work, and Evaluation

Ye-Xin Lim, Su-Cheng Haw* and Jayapradha J

***Abstract* –Peer reviewer assignment to academic articles is important in ensuring the quality and originality of academic publications. Traditional methods of selecting reviewers are generally plagued by inefficiency, reviewer burnout, and inconsistency between the subject of the manuscript and the reviewer area of expertise. In attempting to avoid such drawbacks, recommender systems have been explored as a means of solving the reviewer assignment problem. This article reviews the recommender system techniques in detail by reviewing their application in peer reviewer selection. Additionally, related works shall be examined for how different methods work, their strength and limitations, the dataset used by them, and evaluation metrics used in measuring system performance.**

Keywords—Recommender System, Hybrid-based, Peer Review, Evaluation Metric, Research Article, Reviewer Assignment.

I. INTRODUCTION

Journal management systems have made life easier for authors, reviewers, and editors by speeding up submissions, improving communication, and simplifying editorial tasks. Yet, even with these digital platforms, there are still major pain points when handling large volumes of manuscripts, balancing reviewer assignments, and keeping communication transparent among all parties. These inefficiencies can cause delays, uneven workloads, and confusion during the editorial process, ultimately affecting a journal's ability to maintain consistent quality and timely publications [1].

One area in which such limitations become even more apparent is the process of assigning peer reviewers. Editors often depend on their networks for manuscript allocation, and this results in reviewer fatigue, workload imbalances, and sometimes even mismatches between the article's content and a reviewer's expertise. Furthermore, the journal's capability of making sure that submission is assigned to the best reviewer may be compromised because, in the past, various systems for tracking reviewer performance and availability have lacked the resolution necessary to enable rapid and accurate assignments [2]. These have the potential to slow down the editorial cycle and lessen the quality of feedback to authors.

To address these challenges, hybrid-based filtering techniques (HB) recommender systems have emerged as a promising approach to enhance the assignment of peer reviewer process. HB systems combine multiple recommendation strategies, such as collaborative filtering techniques (CF) and content-based filtering techniques (CB), to provide more

*Corresponding Author email: sucheng@mmu.edu.my, ORCID: 0000-0002-7190-0837

Ye-Xin Lim is with Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia (e-mail: 1211104730@student.mmu.edu.my).

Su-Cheng Haw is with Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia (e-mail: sucheng@mmu.edu.my).

Jayapradha J is with Department of Computing Technologies, SRM Institute of Science and Technology, Tamil Nadu, India (email: jayapraj@srmist.edu.in).

International Journal on Robotics, Automation and Sciences (2025) 7, 2:56-76
https://doi.org/10.33093/ijoras.2025.7.2.6
Manuscript received:11 Mar 2025 | Revised: 30 Apr 2025 | Accepted: 13 May 2025 | Published: 30 Jul 2025

precise reviewer suggestions. By analyzing factors such as expertise, availability, and past performance, these systems enable journal editors to assign reviewers more effectively and equitably. This integration of advanced technology into journal management systems not only improves the efficiency of reviewer assignments but also ensures that the quality and fairness of peer reviews are maintained [3].

II. Recommender System

A. Overview of Recommender System

Recommender systems are now essential parts of modern digital platforms, providing users with personalized suggestions based on their analysis of needs, tastes, and use. They have deep applications in various fields, from business-to-consumer e-commerce [4] [5], healthcare [6] and entertainment to academic publishing. Academic recommender systems have also enabled decision-making activities, such as suggesting peer reviewers or appropriate journals for manuscript submission [3].

The recommender systems have gained much momentum in managing the wide diversity of multimedia content people face today. They provide a basic framework for users to find interesting and relevant content in various forms, ranging from audio and text to images and videos [7]. Complex algorithms analyzing user behaviours based on interactions, preferences, and past activities make personalized content delivery possible. Recent developments in machine learning and deep learning have increased both the importance and the accuracy of these recommendations. For example, [8] emphasize that incorporating social interactions and item relationships in a graph-based recommender system provides better context-aware recommendations, increasing user engagement and satisfaction.

With a modern increase in application, recommender systems show their role in enhancing healthcare by providing recommendations on treatments, medical resources, and lifestyle modifications. According to [9], structured data inputs such as case history and patient responses may facilitate the generation of personalized recommendations. Contextual and user-specific information enhances decision-making by bettering patient outcomes and proving the transformative potential of personalized technologies in health.

The primary function of a recommender system is to sort through large collections of data and present relevant information or options to the users. In academic applications, this means correlating article metadata—keyword, abstracts, topics—and against potential reviewers or journals with their work history and publication record. The recent advances in machine learning and deep learning also enhanced the accuracy and efficacy of such systems [10]. For example, CB analyzes manuscript features, while CF depends on reviewer feedback and refine the matching process.

In journal management systems, recommender systems make reviewer assignment easier with less human effort, more consistency and higher relevance of reviewer-paper pairings. This method not only reduces issues like reviewer fatigue and mismatches in expertise but also ensures timely and fair peer reviews [11]. By limiting the context to academic use cases, i.e., peer review assignment, this research aims to explore how various recommendation approaches can improve the quality and efficiency of scholarly publishing processes.

Apart from that, recommender systems are applicable in efficiently identifying the most appropriate journals for research submissions. [12] discuss the CB filtering approach, where the system matches a researcher with journals based on the analysis of attributes of articles such as keywords, topics, and abstracts. Using past publication data or users' preferences to enhance its accuracy in journal recommendations saves researchers a lot of time in selecting journals, thus having a higher chance of getting their article accepted in the first attempt.

B. Peer Review Challenges in Journal Management System

Peer review forms the very backbone of academic publishing to ensure quality, reliability, and scientific stringency in the findings presented. In so doing, the process involves an assessment by experts of the manuscripts submitted regarding originality, its methodology, and contribution in the field. The greatest aim of peer review in all research should be refinement for scholarship that leads to improved impact. Nevertheless, its weaknesses, despite the great importance of the peer review process, are numerous in its effectiveness and credibility.

One significant challenge is the deteriorating quality of the reviews. Sometimes understandably elated to learn about acceptance, authors were usually submerged under an overwhelming avalanche of reviewer comments that ran into dozens or hundreds in some cases. Furthermore, reviews sometimes prioritise minor problems—such as language and formatting over major criticisms of the research's scientific validity and methodological rigour [13]. Current solutions to this problem stress the need for a more selective approach such as weighing the reviewer's qualifications and expertise in the focus of the manuscript to invite reviewers. Those referees who may have a conflict of interest or are only able to review part of the work, for example, research methodology or statistical analyses, must clarify the scope of the review with the journal beforehand. This will keep the review focused and not let the quest for less important points ignore important scientific validity and methodological rigor feedback.

Another major problem is the inconsistency and subjectivity inherent in peer reviews. Reviewer recommendations generally are opposite: one reviewer may lavish praise on a manuscript, while another might castigate it. This might confuse editors and compound the decision-making process of authors. Biases and conflicts of interest may also compromise the objectivity of reviews. Social or cultural biases can insidiously affect judgments;

predatory journals—profit-oriented rather than interested in scholarly contributions—often have fake or cursory reviews that undermine trust in the process [14]. With all the challenges, some strategies that various journals can consider making their peer review processes better include using standardized review templates or scorecards to formalize assessments; offering continued training to reviewers for purposes of honing their expertise and raising awareness; providing constructive feedback as a way of ensuring self-assessment and accountability. Besides that, systematic use of ratings on the quality of reviews, acknowledgement of exemplary reviewers, and the application of double blinding reduces opportunities for bias. Lastly, inclusive measures ensure that the reviewing of diverse and junior reviewers get recruited to ensure greater dependability and fairness. In the final analysis, these measures strengthen the system by increasing rigour and credibility—two properties that strengthen peer review for consistency and subjectivity.

The difficulty in finding qualified reviewers poses another serious challenge for journal editors. While many academics are ready and willing to be on an editorial board, the type of reviewers sought for expert advice on a given article becomes very difficult for editors to balance. This is the timeliness in review processes versus ensuring competence among the reviewers [15]. These are often compounded by time pressures arising from stringent publication schedules, which leave editors with few options and open to the risk of either delays or inadequate evaluations. One can also contemplate how journals should work on giving reviewers more acknowledgement and incentives, such as certificates for their work on peer reviewing, acknowledging reviewers in public on the journal's homepage, or naming them in articles that have been published. In addition, to quicken the review process with an eye on the efficiency of the peer review system, there is the intention to introduce fast-track review methods. These innovations will provide more participation in reviewing and further smooth its efficiency without compromising quality.

Adding to the challenges in journal management systems is reviewer fatigue. Results from a 2018 Publons study have shown that reviewers are becoming harder to secure. In 2013, an average of 1.9 invitations had to be sent out to get one completed review, but this number increased to 2.4 in 2017 and is likely to increase further to 3.6 by 2025. The COVID-19 pandemic exacerbated this burden and limited the availability of academics for peer review [11]. These pressures have led to a few peer review models that challenge the traditional model: open peer review to increase transparency, post-publication peer review for sustained criticism, opulent peer review to provide monetary incentives, transportable peer review to increase the likelihood of the previous work used in different contexts, community peer review for increased collaboration, and cascading peer review where previously rejected articles are routed to other journals. These innovations aim to address reviewer fatigue, speeding it up and allowing for flexibility in its process.

Pressures to publish have resulted in fraud cases. Cases of fraudulent peer reviews, one such unethical practice, are designed by the authors to facilitate the eventual acceptance of an article in review by submitting fictitious reviews, often by themselves or associates, but employing fake email addresses [16]. Such activities make the entire scientific publishing process vulnerable to exploitation of a loophole in the management system of journals. New methods are being researched to detect and address these issues. The repetition of text patterns within reports of multiple reviewers may be discovered by examining certain datasets for referees' comments that are provided during peer review. While template use in and of itself is not a definitive marker of bad conduct, multiple referee accounts using the same templates as the article under review's author give much stronger indications of unethical behavior. The research thus shows the potential for the application of data-driven approaches toward the finding and mitigation of fraudulent practices in such a way as to strengthen the transparency and credibility of the peer-review system [17].

Addressing these vulnerabilities in the peer-review system requires robust technological solutions. Open Journal Systems (OJS), on the one hand, is an open-source, versatile, online system that helps the scholar manage submission workflows and peer review, thus enabling more open access to academic journals. OJS is confronting some challenges with the technical expertise of staff, potential security risks, and complex usability for users. Now, community-driven development has delivered updates, patches, and improvements in usability to overcome such concerns. For current solutions, its architecture is modular; it can thus tweak workflow and interface with external applications because of a large ecosystem of plugins, making the platform adaptable and successful for such an evolving academic need for communication [18].

EDAS is an extensively applied online conference management system. EDAS manages abstracts, peer reviews, registration, and payments, among other tasks. However, EDAS lacks hosting features or wide customization options, which would limit its flexibility for conferences with specific or sophisticated needs. EDAS aims to overcome these constraints by providing strong standard capabilities, including integrated registration and payment processing and a thorough peer-review system. These solutions help organizers to efficiently manage regular academic events despite the lack of advanced customizing and simplify their administrative burden [19].

Popular conference management tool EasyChair has certain limitations in customizing choices, which might not fit events with specific requirements or procedures. Apart from that, EasyChair does not have hosting services, but OpenConf and ConfBay let you access them. With its own strengths and limitations, EasyChair invests effort in providing ease to its users through the simplicity of interfaces and strong core functionality like article submission, peer review management, and support for virtual conferences. EasyChair allows for the effective running of academic events by incorporating tools such as automated review assignments, email notifications, and Virtual Conference Support (VCS), which enables smooth coordination for both in-person and hybrid conferences despite the lack of extensive customization and hosting capabilities [19].

III. RECOMMENDER SYSTEM TECHNIQUES

Recommender systems stand at the heart of personalized experiences in today's data-driven world. They study user preferences and interactions to suggest items relevant to users. These encompass CB filtering, CF filtering, and HB models; all offer different ways to identify patterns and improve the accuracy of recommendations [20]. Other advanced methods involve Semantic-based (SB) and Generative AI-based (GAI), enhancing recommendations with context and creativity. These systems find broad applications in e-commerce, health care, and entertainment, among other fields, where technique choice is crucial for effective and relevant recommendations. Each approach has its strengths; understanding them allows for tailored solutions to suit the specific needs of many applications.

Alongside the various techniques highlighted, the methodologies of CF, CB approach, and HB models are crucial in recommender systems concerning effectiveness in the education sector [21]. The recommender techniques identify relevant resources, courses, or material about an individual's needs or preferences; the presented learning experience must be personalized. As challenges related to recommender systems are involved, initial scarcity of data requires a well-thought-out balance of privacy concerns versus personalization. These have great potential for increasing engagement and thus improving learning outcomes. The accuracy of recommendations is influential in the choice itself and extends to the general learning process; therefore, the best method should be chosen regarding educational context and purposes.

The five broad categories into which most data filtering methods used within recommender systems include GAI, CB, HB, CF, and SB. Each has specific advantages and focuses on various dimensions of suggestion generation. In other words, CB filtering operates by analyzing the attributes of a certain item in relation to prior user preferences or actions. Technique(s) using this paradigm essentially depend on the essential characteristics from keywords and genres of the products that shape these suggestions [22]. On the other hand, recommendations in CF are developed based on analyzing the trends and similarities of similar users' behaviors; these come from interaction [23]. While CB filtering focuses on a single user's history, CF makes recommendations based on the preferences and actions of a large pool of users. HB filtering combines collaborative and CB methods to overcome weaknesses inherent in either method or improve the effectiveness of the recommendation process. SB filtering further advances this process by embedding contextual understanding, leading to better and more accurate recommendations. GAI filtering allows AI models to generate dynamic, personalized suggestions, improving recommender systems' effectiveness. Which filtering technique to choose depends on the application for which the recommendations must be practical and relevant. Figure 1 summarizes various filtering techniques used in recommender systems.

A. Content-Based Filtering Techniques

The aim of CB, as a personalized recommendation approach, is to understand each user's unique tastes and align them with the products' characteristics [22]. This approach relies on constructing a detailed user profile based on information about his past interactions, including things he/she enjoyed, has given high ratings on, or interacted favourably with. At the same time, items are described by specific attributes, like keywords, features, or tags, which allow structured identification of similarities between items. By comparing the user's profile with these attributes, CB algorithms recommend items that closely match those the user has previously enjoyed. CB does a great job at making precise and relevant suggestions since it only considers the information and preferences of the single user. It's a trusted method to develop good suggestions; it values user customization, and the results are usually flawless.

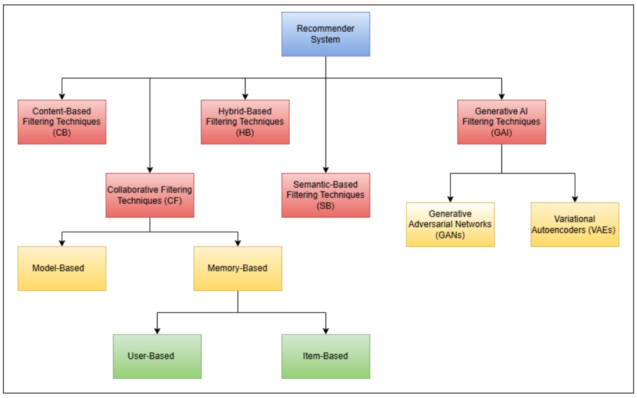

FIGURE 1. Overview of recommender system filtering techniques.

Besides, there is a wide application in real-world scenarios, where different domains use CB to do personalized recommendations. Music recommendation systems suggest songs that blend with the user's tastes by analyzing attributes such as genres, artists, tempos, or style of music. For instance, if the user likes some rock band, the system will most probably suggest other rock bands or songs with similar stylistic features or which belong to the same genre. In e-commerce, CB is usually used to recommend a product the user may prefer by checking the parameters of the products, such as brand, category, colour, and price range [24]. Additionally, customers who often purchase exercise machines may be recommended auxiliary merchandise, such as yoga mats, dumbbells, or resistance bands. In this approach, products are profiled on unique features—like genre for movies, authors for books, and technical specifications for electronic devices—and matched with the user's profile to make highly relevant and personalized recommendations.

Additionally, the use of complex algorithms for CB approaches helps to identify similarities of things and thus to suggest items in a personalized manner. Among the commonly used, one can find the TF-IDF algorithm that weights keywords in two ways: based on its occurrence in the single item and based on their rareness in the dataset, as combined by Term Frequency–Inverse Document Frequency (TF–IDF) according to the work of [25]. It does so by giving more importance to unique and relevant features of documents and dampening the effects of common terms and, therefore, less discriminative. Once items have been vectorized using TF–IDF scores, a cosine similarity method is used to measure similarity among those vectors. Cosine similarity computes the angle between vectors' cosine and indicates how similar an item is compared to a user's profile or other objects. CB thus makes the suggestions accurate and highly relevant, as TF-IDF effectively extracts the features and cosine similarity determines the similarity. These recommendations also fit well with the user's taste and history of historical interaction.

B. Collaborative Filtering Techniques

CF is a popular approach in recommendation system design that has been proven to predict user preferences. It recommends items that fit personal tastes using substantial historical data on user behaviour, such as ratings, interactions, and purchase histories [26]. Unlike CB filtering, which requires an investigation of item properties, CF detects commonalities between people or objects based on common interactions. When enough interaction data is available, CF finds similar persons or items using similarity algorithms to give personalized recommendations. It correctly gives appropriate recommendations by utilization of group user behavior, thus leveraging collective action [23].

Moreover, comments on user-item are quite important because CF involves complete user interactions with objects. Most comments, which help identify user preferences, are derived mostly from implicit data, including clicks, likes, and consumption patterns. For example, item-level implicit feedback tells customers about their preferences and behavior while linking them with the products they are most involved with. Such comments are particularly useful in systems where user interactions are high since they enable CF systems to spot trends and similarities for personalized effective recommendations [27]. A deep assessment of such input will enable the CF system to respond to dynamic changes in user behavior to ensure relevance and effectiveness.

Beyond traditional recommendation systems [28], CF has demonstrated great adaptability by being applied in specialist domains including Web service Quality of Service (QoS) prediction. It projects the user's own QoS levels based on historical data in particular contexts. CF's success comes from its several techniques, generally classified as HB, memory-based, and model-based. In contrast, while model-based methods depend on sophisticated algorithms that explore hidden patterns within the interaction matrix, memory-based methods calculate similarity directly from data. Combining the best of these techniques, HB models melt them together, thereby increasing a certain method's final accuracy and adaptability. This makes CF flexible for rather different user needs and application domains and strong for a broad spectrum of recommendation jobs.

Generally, the CF methods can be classified into two types: model-based CF and memory-based CF. Each approach uses different ways to generate recommendations based on the user-item interaction data.

Model-Based Collaborative Filtering Technique: Model-based CF techniques aim to identify patterns from historical data to develop predictive models that will estimate unknown ratings or preferences. These tactics use varied algorithms to reveal hidden connections between a user and a product. The common model-based techniques are clustering, matrix factorization, support vector machines, and stochastic gradient descent [29]. Each one gives different ways to estimate and predict user behaviours; hence, model-based CF is powerful for providing personalized recommendations in complex, large data sets. By making underlying data structure more interpretable, the model-based strategy provides more accurate and scalable recommendations.

In model-based CF, factorization-based models are widely utilised to predict unknown ratings. The models decompose the user-item rating matrix into two smaller matrices representing latent user and item factors. Matrix Factorisation is the predominant technique in this category, particularly effective for addressing sparse data characterized by numerous missing user-item ratings. The model predicts user ratings for unengaged items by representing users and items through latent factors. Predictions are derived by computing the dot product of the respective user and item latent factor vectors [30]. This method is essential to model-based CF as it facilitates the creation of personalized recommendations derived from patterns identified in historical data.

Model-based techniques under CF include latent factor methods representing user preference and item characteristics in reduced dimensional space and factorization-based models. These find hidden patterns in user-item interaction, which explicit ratings cannot be obtained. Latent factor methods are very effective when combined with clustering techniques that divide users or items into groups based on similar characteristics [31]. The clusters enhance recommendations toward increased accordance with the user preferences.

Memory-Based Collaborative Filtering Technique: Memory-based CF techniques utilize the complete user-item rating dataset to produce predictions. These systems utilize statistical methods to identify a cohort of users, termed neighbours, who exhibit preferences analogous to the target user. The fundamental premise is that users who have previously consented will likely re-consent. Two main approaches to memory-based CF exist: user-based and item-based. User-based CF produces predictions by examining the target user's preferences compared to those of analogous users. Item-based CF emphasizes identifying similar items through analysing user rating histories [32]. The two methods integrate neighbour preferences through different algorithms to produce a prediction for the active user, which will be examined in the following sections.

According to [33] user-based CF starts by calculating the similarities between the active user and other users. Common similarity measures used in CF include cosine similarity, adjusted cosine similarity, and the Pearson correlation coefficient. Neighbours are selected based on their high similarity to the active user. The system uses ratings from neighbouring users for an item to compute a weighted average; using similarities as weights yields a predicted rating. The recommender systematically sorts all items by their predicted ratings in order and then recommends the Top N items.

Item-based CF centers around the computation of item similarities. The basic idea is that items similar in terms of user ratings are likely to share similar features. The similarity measures from user-based CF are applied to estimate the relationships between items. Once the neighbors of a target item have been identified, the system computes a weighted average of ratings, taking these similarities as weights to predict the rating of the target item. Like user-based CF, scalability may pose a problem when the number of items is large or when the item catalogue undergoes frequent changes [33].

C. Hybrid-Based Filtering Techniques

HB has represented a sophisticated recommender system methodology that combines all the merits of collaborative and CB filtering to diminish their respective limitations [34]. The basis of CF is founded on similar user preferences, whereas the CB filtering technique bases its basis on the user's previous interactions. Both methods face challenges, noticeably the cold-start problem resulting from insufficient data on new users or items to improve recommendations. HB systems meet these challenges by integrating multiple strategies to improve personalisation, manage data sparsity, and enhance robustness and inclusiveness of recommendations [35].

According to [24], HB relies on seven integrative techniques at its core for incorporating multiple recommendation methods, each to handle weaknesses of individual methods and improve the overall suggestion performance.

Weighted Hybridization: It is a method that quantitatively integrates scores of different suggestion elements by giving distinct weights to each. The weights express the importance or reliability of each approach and are used to calculate the final suggestion.

Switching Hybridization: Switching involves selection and execution of a recommender component, depending on the set of defined parameters or criteria.

Mixed Hybridization: In the mixed approach, it simply combines suggestions received through several other techniques into its results. It blends several types of recommendation content into one interface that produces improved coverage and increases variety for better results for each user.

Feature Combination: This approach fuses features extracted from different knowledge bases to build a unified recommendation framework. The quality of recommendations can be enhanced by fusing different sets of features. This may provide a deeper understanding of users' preferences.

Feature Augmentation: Feature augmentation deals with the generation of additional features or refinement in some of the existing ones. Those more refined or newly developed features have become the input for a fancier recommendation algorithm. Improvement of input data by adding more relevant information greatly enhanced the quality of recommendations.

Cascade Hybridization: Cascade systems make recommendations based on a set hierarchy of priorities. The main approach is tried first; other approaches are used only to break ties or refine the recommendations. This approach ensures that decision-making is orderly and efficient.

Meta-level Hybridization: Hybridization at the meta-level involves using one recommendation technique in modeling, which acts as the input for another recommendation technique. Combining them in a series allows the system to use the advantages of each technique and makes the recommendation mechanism finer and more effective.

These hybridization strategies allow HB recommender systems to overcome the limitations of individual methodologies. Thus, they provide recommendations that are both accurate and diverse, as well as dependable.

D. Semantic-Based Filtering Technique

Recently, SB filtering techniques have been found to be one of the leading approaches, and they have significantly improved the accuracy and diversity of recommendations. Main challenges, such as data sparsity or the cold-start problem, were addressed by applying reasoning techniques from the Semantic Web. Compared to the traditional syntactic-based systems that usually don't consider key information in the recommendation processes, SB methods take advantage of those deeper relationships among entities and use them to make more significant suggestions. These systems provide recommendations that align more closely with users' preferences and interests by revealing implicit semantic connections and transcending superficial similarities [36].

Ontology adds value to SB filtering techniques by providing a broad knowledge base that describes structured relationships among entities or concepts in a specific domain. This overall conceptual scheme then enables a variety of semantic relations, setting it apart from more simplistic representation approaches, like keyword-based methods. Recent improvements in SB recommender systems have refined the similar estimates made by conventional CB and CF techniques by applying ontologies and semantic reasoning, leading to better recommendations [37].

Further developing the function of ontology in SB systems, recent methodologies are improvements to existing methods for detecting spam. Most recent research suggests SB methods, which identify concealed improvement patterns in textual data. Latent semantic indexing is one such technique that proposes a latent space representation generated through singular value decomposition to create latent spaces that represent the actual underlying semantics of the documents more precisely. Similarly, advanced models such as LDA and labeled-LDA have also been applied to build semantic representations for spam detection. The proposed method leverages semantic ontologies to extract dominant topics from text messages. These consistently outperform traditional models such as the BoW framework, which again underlines the strength of SB classification in handling complex tasks such as spam detection [38].

E. Generative AI Filtering Techniques

In recent years, remarkable advancements have been observed in GAI, with new capabilities transforming industries and redefining technological possibilities. Among innovative areas were recommender systems, in which research focused on how GAI might augment or supplant traditional components. Deployment of these advancements in real-world systems, especially e-commerce, is problematic because of the complexity of industrial recommender systems. These systems require an end-to-end integration of AI models, infrastructure, operational workflows, and business considerations. That makes the holistic approach very relevant to the adoption [39].

According to [40], GAI made creative arts, computer vision, and natural language processing evolve by allowing them to create artificial data which accurately showed the real-life situations. Their adaptability extended the area of application that included various sectors like healthcare, entertainment, and financial services, the outcome of which was innovative applications: text generation, image synthesis, music composition, conversational AI, and much more. All of this adds up to a very fruitful success with great promise for education, applied data science, and various other disciplines; thus, the GAI will serve as the catalyst.

Building on its transformative potential, GAI employs powerful models like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), which excel at creating high-quality, novel content. These models enhance the reliability of generated data and provide some of the privileged advantages of traditional methods applied in different recommendation systems. The GANs and VAEs learn from available data or samples and generate innovative content; hence, they are crucial for developing the efficiency and effectiveness of recommender systems [40].

Generative Adversarial Networks (GANs): GANs consist of a generator and a discriminator that compete for improvement. They define a collection of roles that can be used to create recommendation systems. IRGAN and other interaction-based algorithms generate significant training data by relying on negative discoveries. GANs can come up with user options or interactions that supplement training datasets and model learning skills. GANs can also generate whole pages or a list of recommendations, meaning they tend to be pretty good at providing full-page suggestions [41].

Variational Autoencoders (VAEs): Variational autoencoders have found applications in recommender systems for issues regarding the shortcomings of traditional CF, such as the sparsity of rating matrices and a cold-start problem that makes it challenging to generate recommendations for new users owing to a lack of previously available data. From another perspective, VAEs may apply differential privacy at an individual user level to strongly protect privacy, offering one secure way of managing sensitive user data. These non-linear models, now driven by neural networks, can show complex patterns hidden in the data. Variational autoencoders transform high-dimensional inputs into a lower-dimensional "latent space" that represents the essential structure of the data in a probabilistic manner. VAEs are a very successful approach to improving the performance of recommender systems since, by sampling from this latent space, they can generate outputs like the original data, as pointed out by [42].

Choosing the best approach for a recommender system requires deep knowledge of its advantages, disadvantages, and underlying mechanisms. This section considers some methods and weighs up their pros and cons, as highlighted in Table 1. This table is useful in determining which approach best fits the data properties and the intended forecast results.

TABLE 1. Comparison of recommender techniques.

Advantages	Limitations
Content-Based (CB):	
- It creates each user's unique profile by going through each user's preference and generates recommendations based solely on information from the currently viewed user. - These systems are particularly effective at recommending new and unrated items, helping new users increase their options and improving their experience of exploring. - Does not suffer from cold-start problem	- Requires a good number of item features. - Developing qualities for items within areas can be a complicated endeavor. - Often promotes uniform items, resulting in a problem of overspecialization.
Collaborative Filtering (CF):	

- Embeddings can be learned automatically, without the need for manual engineering features. - Relies on user behavior and preferences. It allows for item suggestions without comprehensive product content information. - Be able to scale well on large datasets since the number of users and items becomes large in a recommendation space, and it will need only a pattern of user interactions to make recommendations.	- Suffer from data sparsity. - The cold-start problem in CF arises when there is inadequate user data to produce precise recommendations. - The system often cannot embed an item that was not encountered during training, rendering it difficult to query the model with that item.
Hybrid-Based (HB):	
- Employed to surmount the constraints of CB, and CF methodologies. - Yields more robust and personalized suggestions for users. - HB systems demonstrate flexibility to different domains and data types, allowing for the integration of several sources of data.	- Requires a substantial database to maintain updated data metrics. - High computational complexity.
Semantic-Based (SB):	
- Provide highly personalized recommendations by understanding user intent and preferences while looking at semantically related objects and offering a range of recommendations. - A range of recommendations. - Suggest new items based on their attributes and relationships, effectively overcoming the cold-start problem for items.	- Rely heavily on detailed and accurate metadata or ontologies, which can be time-consuming and expensive to develop and sustain. - An excessive focus on similar semantics may lead to too narrow recommendations, decrease variety, and possibly prevent users from finding new kinds of products.
Generative-AI (GAI):	
- Infer patterns even with limited interaction data, addressing sparsity issues. - Handle multimodal data, such as combining text, images, and audio, to deliver richer and more versatile recommendations.	- Requires substantial computational power and memory for training and inference. - Requires advanced expertise and resources, making them challenging for smaller organizations.

To better understand the real-world applicability and performance of different recommendation methods, a summary of some recent research papers is presented. Table 2 provides a general comparison of various recommender system techniques, including content-based, collaborative filtering, hybrid, sentiment-based, and generative. A concise summary of the main methodology, strengths and weaknesses, datasets, and performance metrics is included in each item.

The comparative analysis demonstrates that content-based recommender systems are appropriate for learning user tastes from item features such as course materials, movie features, or restaurant features. Similarity metrics such as cosine similarity, TF-IDF, and deep belief networks (DBNs) are the standard methods used in matching similar items to like users. The models prove useful when the histories of users are brief, or one needs relevance to particular domains. But they have the tendency towards issues like cold-start problem and over-specialization, where the same kind of items are continuously given to the users without having sufficient novelty and diversity.

Collaborative filtering techniques, especially those based on matrix factorization and deep learning, are effective in domains with high user-item interaction data. They predict using collective user behavior and are widely applied in employment, film, and product recommendation. Traditional collaborative filtering models are hindered by scalability and data sparsity problems, which are addressed in current studies by neural models, graph models, and user clustering such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN). These advancements increase precision and generalization but still depend on the availability of user ratings.

Comparative analysis also points to the newer trend of hybrid, sentiment-sensitive, and generative approaches. Hybrid models use content and collaborative signals to boost personalization, lower cold-start, and increase diversity. Sentiment-based systems use opinion mining to make a better sense of user opinions and attain higher precision and F1-scores. Lastly, generative approaches-autoencoders and large language models (LLMs) are a cutting-edge phenomenon in recommendation studies. They imitate user preference or infer hidden attributes to generate more context-driven and realistic recommendations. Generally, outcomes of Table 2 reveal that although traditional methods still dominate space, newer systems that incorporate deep learning and sentiment analysis perform better across domains.

IV. Evaluation Metrics

Evaluation metrics are important entities in defining quantitatively a machine learning model or an algorithm's performance regarding how it may achieve or attempt to obtain any aim, target, task, goal, or challenge in question. In evaluation metrics, models or algorithms are compared for a particular action in estimation and identification of where improvement must be sought over time. The evaluation metrics range widely depending on the task or goal that is aimed at by a model or algorithm. Some of the common ones include accuracy, precision, recall, F1 score. Each of these metrics has strengths and weaknesses, and it is necessary to select the correct evaluation metric considering the task or goal of the model or algorithm. Besides, in many cases, the use of multiple evaluation metrics allows for a better understanding of the model's performance. In conclusion, evaluation metrics are among the important building blocks in developing, testing, and improving machine learning models and algorithms. For this review, the evaluation metrics are accuracy, precision, recall, F1 score.

TABLE 2. Comparative summary of recommender system techniques based on strengths, weaknesses, datasets, and evaluation metrics

References & Recommender System Techniques	Strengths & Weaknesses	Dataset	Evaluation Metrics
Mokarrama, Khatun and Arefin. [43] Content-Based (CB)	Strengths: - Considers academic and non-academic factors (GPA, fees, university ranking, Google ratings). - Real-time recommendation using normalized data. - It is easy for students to map preferences to university data. Weaknesses: - Cold-start problem if user has no minimal data. - Not easily scalable to new users/universities without updating scraped data.	Real-world dataset of 97 departments from 15 private universities in Bangladesh; features: SSC/HSC GPA, tuition fees, world ranking, Google rating.	The model achieved accuracy (89.05%), recall (95.85%), F1-score (92.32%), specificity (48%), and balanced accuracy (71.93%) with good ability in recommending helpful universities but less accurate for discarding irrelevant ones.
Sridhar, Latha and Dhanasekaran. [44] Content-Based (CB)	Strengths: - Builds detailed user profiles from Facebook, with high personalization. - Independent of other users' data, minimizing cold-start for user-based systems. Weaknesses: - Content-based only, with potential for over-specialization and sparse diversity. - Computationally heavy and could fail to scale for real-time deployment.	Facebook user profile dataset (565 users) and MovieLens 100K dataset (943 users, 1,682 movies, 100,000 ratings).	The system achieved Mean Absolute Error (MAE) of 0.716 and Root Mean Square Error (RMSE) of 0.915. The system achieved precision of 97.35% and recall of 96.60%, which shows very high precision and strong match between user interest and recommended films.
Aljunid and DH. [10] Collaborative Filtering (CF)	Strengths: - Resolves sparsity issue inherent in collaborative filtering systems. - Learns abstract user-item interactions via multilayer neural networks. Weaknesses: - Solely based on user-item rating history, without any room for new items or users (cold-start). - Costly to train models, especially with big datasets.	MovieLens 100K and MovieLens 1M datasets (ratings on 1,682 and 3,952 movies by 943 and 6,041 users, respectively).	RMSE was employed to measure the system. The Deep Learning Collaborative Recommender System (DLCRS) proposed in this research performed better than baseline approaches such as Singular Value Decomposition (SVD), User/Movie Average, Cosine Similarity, and standard Matrix Factorization (MF) based on RMSE with scores 0.917 (MovieLens 100K) and 0.903 (MovieLens 1M).
Mishra and Rathi. [45] Collaborative Filtering (CF)	Strengths: - Recommender system is focused on personalized job suggestions using user profiles, behavior, and resumes. - Combines collaborative filtering with graph-based and hybrid models to ensure increased scalability and accuracy. - Tested on actual job portal ecosystems like LinkedIn and Work4. Weaknesses:	Real-world information of Work4, CareerBuilder, and LinkedIn Job Ecosystem. Includes user profiles, resumes, behavior, and job description. Four data sets taken into consideration: random, feedback, candidate-based, and aggregated (total 26,669 instances).	Evaluated on Precision, Recall, F1-score, Maximum Likelihood Estimation (MLE), Pointwise Mutual Information (PMI), and Scalability. Linear SVM did best with up to 9% accuracy on aggregated data, beating Cosine Similarity (8.5%). Graph-Based Approaches (GBA) performed best on different job datasets.

	- Cold-start and sparsity issues are still partially unresolved Does not possess standardization across data sets and evaluation schemes.		
Chalkiadakis et al. [46] Hybrid-Based (HB)	Strengths: - Employs a light-weight Bayesian elicitation procedure to quickly construct user profiles. - Combines two types of semantic similarity: hierarchy-based (XWP) and non-hierarchy-based (WEJS). - Resolves cold-start issues common in tourism applications. Weaknesses: - Efficiency on elder age groups marginally lower. Validation on real-users still pending (evaluation on synthetic users).	Real-world dataset of 430 Points of Interest (POIs) from Agios Nikolaos, Crete, Greece and tourist preference data collected from 150 real visitors and extrapolated to 600 synthetic users across 6 age groups.	Performance of the system was assessed via Precision, Recall, and Upper Bound Recall (UB). Best performing version (4 elicitation slates × 6 images per slate) resulted in average precision up to 89.3% over age groups and recall reaching 85% upper bound recall. Precision was well preserved ($\geqslant$ 84%) under most conditions and ensured the utility of the hybrid system in terms of recommending relevant POIs. The CB module alone (with WEJS + XWP) obtained an average precision of 79.1%. The hybrid system was better than individual CB models, particularly in personalization and diversity of recommendation.
Vahidi Farashah et al. [47] Hybrid-Based (HB)	Strengths: - Solves cold-start problem using hybrid similarity and link prediction. - Combines Deep Neural Network (DNN) and DBSCAN clustering to obtain effective user clustering. - Suggests improved Pro-FriendLink algorithm to obtain social link-based recommendations. Weaknesses: - Higher processing time compared to simple models. - Higher complexity to limit deployment to low-weight systems.	MovieLens 2013 dataset: 1,000,209 ratings, 3,900 movies, 6,040 users (each rated $\geqslant$ 20 movies); includes demographic info and social links.	The system was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), precision, and accuracy. With 100 users, the model achieved MAE = 0.35 and RMSE = 0.59; with 900 users, MAE = 0.73 and RMSE = 0.95. In classification comparisons, it achieved precision of 98.92% and accuracy of 93.9%, outperforming Decision Trees, SVM, K-Nearest Neighbors (KNN), and Random Forest in both precision and recommendation accuracy.
Asani, Nejad and Sadri. [48] Sentiment-Based (SB)	Strengths: - Retrieves user food preferences directly from comments by applying sentiment analysis. - Does semantic clustering (WordNet + Wu–Palmer) for boosting the extraction precision of user food preferences. - Contextual: location, time, and availability of restaurants are taken into consideration as well. Weaknesses: - Executed differently under different complexity of sentences and qualities of user review. - Was trained on publicly scraped data of low generalizability.	100 users' restaurant reviews from TripAdvisor (Jan–Oct 2018). Reviews from Jan–Jun used for training, Jul–Oct for evaluation.	The model was trained on Precision, Recall, and F1-score at Top-1, Top-3, and Top-5 levels. At Top-1, the model was about 73% accurate, 69% recalled, and its F1-score was 71%. At Top-3, precision was approximately 86%, recall was 82%, and the F1-score was 84%. At Top-5, the model was best with 92.8% accuracy, 88% recall, and a 90% F1-score. These results surpassed POST-VIA360's and Buon Appetito's, particularly for F1-score and Top-5 accuracy, with the best results being obtained from Wu–Palmer clustering using Cosine similarity.
Karabila et al. [49] Sentiment-Based (SB)	Strengths: - Leverages explicit ratings and review sentiment for recommendation personalization. - Bidirectional Long Short-Term Memory (Bi-LSTM)	Amazon Kindle Book Reviews (12,000 reviews) and Amazon Digital Music Reviews (64,000 reviews), including user ratings and text reviews	The system was evaluated by accuracy, F1-score, and AUC on the sentiment model, and MAE/RMSE on recommendations. The Bi-LSTM model achieved 93% accuracy, 94% F1-score, and

	enhances sentiment prediction by learning left and right contexts. - Trialled on two product categories to check stability. Weaknesses: - Sentiment-based item-based CF illustrates poorer performance as α-values grow higher. - Binary sentiment (positive/negative) constrains nuanced feedback capture.		92% AUC on Kindle data; 94% accuracy, 96% F1-score, and 78% AUC on music data. When integrated with collaborative filtering, the user-based CF with sentiment (α = 0.7) reduced MAE/RMSE from 2.30/2.63 to 1.12/1.30 (Kindle) and 2.18/2.66 to 1.15/1.28 (Music). Item-based CF with sentiment (α = 0.3) reduced MAE/RMSE to 1.18/1.35 (Kindle) and 1.32/1.46 (Music) from 2.36/2.73 and 1.96/2.55, respectively, beating SVM-based baselines.
Binti Mohd Romzi et al. [39] Generative-AI (GAI)	Strengths: - Addressing data sparsity and cold-start using generative modelling. - Learning implicit user/item attributes for product recommendations. - Interleaving similarity scoring for best-N recommendations. Weaknesses: - Train on one single dataset and measure (MAE). - Comparisons are missing with other generative or even classical models.	Amazon Consumer Reviews dataset from Kaggle (Datafiniti Product Database)	Evaluation was done using MAE on both training and test set. The model achieved a MAE of 0.2327, indicating high precision in predicting user ratings. Additionally, loss curves showed decreasing error without overfitting, and similarity scores were used to rank top-N product recommendations. Though precision/recall was not provided, the low MAE and generalization to test data are evidence of the effectiveness of the autoencoder model in generating good recommendations.
Zhang et al. [50] Generative-AI (GAI)	Strengths: - Mimics actual human-like choice, behavior, and emotions. - Permits offline and feedback-based testing of recommender strategies. - Couples both social traits with emotion reasoning. Weaknesses: - LLM hallucination occasionally decreases simulation accuracy. - Limited action space and dependency on rich item descriptions. - Real user deployment yet to be tested.	Real-world datasets: MovieLens-1M, Steam, Amazon-Book.	Evaluation was done on two fronts: Agent alignment and Recommender performance. Agents achieved 65% accuracy and 75% recall at detecting relevant items, with precision and F1-score deteriorating with more distractors due to LLM hallucination.

Accuracy: Accuracy is one of the most basic model evaluation metrics for model evaluation and represents the proportion of correctly predicted instances against their total number. It is an appropriate metric within classification models that represent the counts of true positive and true negatives against the sum of the overall dataset. This provides a straightforward indicator of the overall effectiveness of the model; therefore, it is very popular for binary or multi-class classification tasks. Formula of accuracy is shown Equation (1) below.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{Total number of cases}) \tag{1}$$

Where:

- TP = True Positives
- TN = True Negatives

Precision: Precision represents the accuracy of the positive predictions of the model and is thus one of the most important metrics for model evaluation in a classification context. It can be computed by dividing the number of true positives by the sum of true positives and false positives. Precision proves to be useful in cases when the cost associated with false positives is high and provides insight into a model's ability to flag only the relevant instances as positive, hence reliable in its positive predictions. The formula for precision is given below as Equation (2):

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \tag{2}$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives

Recall: Recall, or sensitivity, is one of the most important metrics when it comes to assessing the performance of classification models regarding how well a model can identify all the relevant instances in any given dataset. It is calculated as the ratio of true positives to the sum of true positives and false negatives; the latter represents positive instances that have been predicted to be negative by the model. Recall is very important in scenarios were missing out on positive instances has critical implications-such as, for instance, medical diagnostics or fraud detection-which is why the emphasis has gone on the model's capability to capture as many positives as possible. The formula for recall is given below as Equation (3):

$$\mathrm{Recall} = \mathrm{TP} / (\mathrm{TP} + \mathrm{FN}) \qquad (3)$$

Where:

- TP = True Positives
- TN = True Negatives
- FN = False Negatives

F1 Score: The F1 Score is one of the critical metrics concerning binary classification model evaluation, mainly when classes are imbalanced. This is the harmonic mean of Precision and Recall; hence, a balanced measure would consider both the precision of the model-that is, how many of the selected items are relevant-and the recall, which refers to how many relevant items are selected. The F1 Score helps balance cases where the model could be optimized for either Precision or Recall to the point that it gives misleading conclusions about its performance. It's particularly useful in cases where the cost of false positives and false negatives is very high. The formula for F1 Score is given below as Equation (4):

$$\mathrm{F1\ Score} = 2 * (\mathrm{Precision} * \mathrm{Recall}) / (\mathrm{Precision} + \mathrm{Recall}) \qquad (4)$$

V. Related Work

This section will further elaborate on the related research works on the assignment of reviewers to articles.

Protasiewicz et al. [51] proposed a content-based recommender system to facilitate improvement in the assignment of reviewers to research proposals for addressing the issue of manual selection of proper experts in high-volume national funding programs. TF-IDF weighting, cosine similarity, and keyword-based full-text search are used to compute the semantic similarity of proposal content and reviewer expertise profiles. The system generates reviewer profiles based on data excavated from Microsoft Academic Search, with author disambiguation and keyword extraction for matching them appropriately. In testing, they utilized a real-world dataset of over 4,000 reviewer profiles and approximately 1,000 submitted proposals from the Polish National Center for Research and Development. Its performance was measured by traditional information retrieval metrics: precision (78.4%), recall (85.6%), and F1-score (81.8%), with high agreement with human-assigned reviewers. Its key advantages are its scalability, automation of the process of finding reviewers, and robustness of the matching quality without requiring reviewer self-declarations. Though, as a limitation, the authors cite the need to improve keyword extraction and increase the coverage and quality of reviewer profiles, especially in areas where publication data is sparse highlighted as an area for future development.

Kameko [52] proposed a heuristic algorithm that can efficiently and effectively assign reviewers to articles by addressing key challenges in workload balancing, ensuring alignment of expertise, and improving computational efficiency. Unlike traditional approaches, the algorithm achieves near-optimal assignment accuracy, which is 98-99% compared to maximum-weighted matching algorithms, at the same time guaranteeing a lower time complexity of $\Theta(n^2)$, hence scalable for large-scale conferences. This is the list of features in the algorithm: uniform distribution of articles among reviewers, iterative and interactive execution, and the ability to ensure at least one competent reviewer per article when resources are available. The algorithm works by using a similarity matrix showing the competence of reviewers in respect to specific articles. First, it sorts articles with respect to similarity scores against reviewers and iteratively performs assignments of reviewers such that a balanced load is maintained on the reviewers. Dynamic updates of the similarity scores of a article to prefer reviewers who have fewer assignments pending, avoiding over-assignment of reviewers. Extensive experiments are carried out with simulated datasets, as well as real data for nine conferences of the CompSysTech series for 2010-2018. These methods are compared in terms of assignment accuracy, computational efficiency, etc. The Hungarian algorithm was also considered as a benchmark method. The results indicated that the algorithm was highly accurate and had higher computational efficiency by a big margin than the greedy and brute-force approaches. Advantages are scalability, the ability to handle real-world constraints like limited resources, and support for interactive adjustments. Limitations are that major issues in multiple assignments of the same article to different reviewers across several iterations may affect subsequent assignment qualities. Future work refines these processes and further optimizes the algorithm for greater applicability.

Besides, Hoang et al. [53] proposed a holistic framework for solving the problem of assigning reviewers in an automated manner through selection and assigning reviewers for research proposals or articles. The system is intended to bring more accuracy and fairness into the process of assigning reviewers concerning relevance of topic, balancing workload, and avoidance conflict of interest. It contains three major modules: data collection, reviewer identification, and group prediction. Advanced profiling and similarity metrics are integrated into the framework to provide better quality in reviewer recommendations, along with diversity criteria to form well-rounded reviewer groups. It involves three major steps: data collection, reviewer identification, and group prediction. The data collection module creates a comprehensive database of scientist profiles by collecting from open-source databases like DBLP and ResearchGate and unstructured web data. The assessment of quality will be indicated by the h-index, publication count, and citation metrics, whereas relevance will be determined through topic similarity analysis utilizing methods such as LDA and co-citation analysis. The Group Prediction Module allows for diverse reviewer groups with respect to affiliations and co-authorships. Assignment considers constraints like conflict of interest, workload limits, and required number of reviewers. The system was evaluated on a dataset of 479 articles along with over 1,000 reviewers, which were collected from the DBLP computer science bibliography. The system was assessed against Normalized Discounted Cumulative Gain (NDCG) and precision. Results have shown significant improvements over baseline models with higher accuracy and better diversity in reviewer recommendations. The advantages include handling large datasets, creating a profile automatically, and the effective integration of quality and diversity metrics. The limitations of the system are that it cannot weigh the reputation of reviewers or adapt to non-bibliographic contexts. Incorporating metrics related to the reputation of reviewers and refining diversity calculations in future work would further develop the framework.

In addition, Tan et al. [54] proposed Word and Semantic-based Iterative Model (WSIM) to solve RAP, which is a crucial issue in academic peer review. This problem requires the appropriate reviewers to be chosen for any given article. In addition, previous methods seldom considered two important factors, which include incomplete reviewer data and the interference of non-article-related articles. It improves the calculation of similarities between reviewers and articles by incorporating word and semantic features through topic modelling and language modelling, thus overcoming such challenges. The model integrates an iterative framework into its structure that helps in minimizing irrelevant article influences with a view to improving recommendation accuracy. The proposed model operates along three dimensions: feature extraction, ranking-based similarity calculation, and iterative refinement. First, the semantic features are extracted by applying LDA to the topic distributions, while the word features are modeled by a boosted language model (LM). Specifically, a ranking-based measure, NDCG, is used instead of an exact probability to give more importance to qualitative relevance for reducing overfitting to incomplete data. The iterative model fuses semantic and word similarities together to update the weight of highly relevant articles for accurate reviewer assignments. It evaluated WSIM on two datasets: one manually constructed, including 400 reviewers and 100 manuscripts, and a larger one derived from arXiv with 1,885 reviewers and 685 articles. Among all metrics compared, including precision, recall, F1 score, MAP, NDCG, and bpref, WSIM has outperformed the compared seven methods in every case while improving by at least 2.5% regarding top 20 recommendation accuracy. Advantages of the proposed model include coping with incomplete data, lesser interference of irrelevant articles, and interpretative relevance rankings. Limitations include real-world testing; it may face scalability challenges if applied on a higher scale. Future research will work on improving the effectiveness of the model and test its applicability in different academic settings.

Moreover, Pradhan et al. [55] proposed an integrated approach for the reviewer recommendation problem in academia with three interrelated layers: Topic Network, Citation Network, and Reviewer Network. Called Topic Network, Citation Network, and Reviewer Network (TCRRec). The Topic Network applies LDA to find key topics from the article abstracts and titles, assigning higher importance to recent publications to capture the current interest of reviewers. The Citation Network creates the graph of article relationships by bibliographic coupling and co-citation, thus enabling the system to consider both direct and indirect topic connections. Finally, the Reviewer Network evaluates the reviewers' expertise and authority given by, for example, h-index, and the links between co-authors to recommend the most suitable ones efficiently with the use of RWR. Figure 2 shows the architecture of TCRRec.

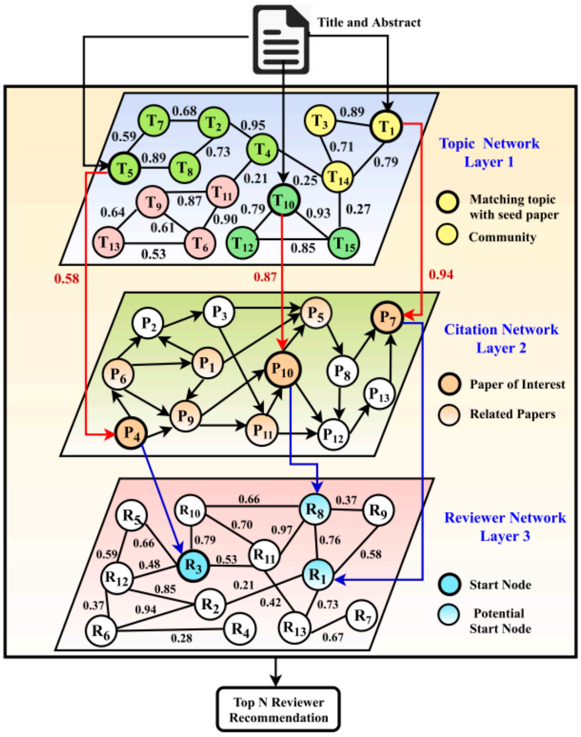

FIGURE 2. The architecture of TCRRec [49].

This work was experimented on using NIPS and AMiner datasets and showed improved performances with respect to precision, authority, and diversity compared to previous methods. The effectiveness of TCRRec is evidenced from different evaluation metrics such as Precision@k, Mean Reciprocal Rank, and nDCG@k. Other main advantages of this approach include addressing the cold-start problem and reduction of computational load by pre-building networks. Authors have proposed enhancements in future work, for example, incorporating more dynamic interests of reviewers and deeper semantic analyses that will further enhance reviewer-article matching.

Apart from that, Kreutz and Schenkel [56] proposed RevASIDE which is an automated reviewer recommendation system designed to solve the complex problem of assigning complementary sets of reviewers for articles from a fixed candidate pool. It targets challenges in traditional reviewer assignments, such as manual effort, subjective decision-making, and lack of diversity. This article targets the recommendation of sets of reviewers that balance expertise, authority, diversity, interest, and seniority while avoiding conflicts of interest for comprehensive and unbiased reviews. This approach avoids manually defined reviewer profiles or keywords, hence providing a scalable and objective solution. This is a two-step process wherein expert search identifies suitable reviewers by comparing articles to reviewer profiles for semantic and topical similarity using document embeddings such as TF-IDF, BERT, and LDA; the system assembles the reviewer sets which optimize the five key dimensions using a scoring mechanism. Conflicts and collaborations among reviewers are considered to maintain objectivity. For experiments, three newly introduced collections were used: MOL'17, BTW'17, and ECIR'17, which include articles and reviewer pools of different conferences. These datasets offer a wide range of metadata and textual representations regarding the profiles of reviewers and articles for evaluation. Metrics such as Precision@10, Mean Average Precision, and NDCG showed the superior performance of RevASIDE compared to baseline methods. Indeed, the system generated quality reviewer sets, which both quantitative and qualitative assessments confirmed. The main advantages of RevASIDE are its holistic consideration of several dimensions, scalability, and ability to operate with no manual inputs, like reviewer bidding. Limitations of the current system include further optimizations that need to be applied for very large-scale venues, while more advanced document representation techniques will be integrated and were pointed out as future work.

Beyond that, Liu, Wang, and Zhu [57] proposed a method for reviewer recommendation that, if applied, would better the efficiency and accuracy in matching reviewers to scientific research proposals. To solve the problem of traditional keyword-based systems, which often have serious sparsity and cannot understand the context, RRM represents both proposals and reviewer expertise using advanced word embedding techniques in low-dimensional continuous vectors. These capture semantic and syntactic relationships, thus enabling more accurate similarity measurements. The approach also fuses ranking techniques in combining multi-source similarity metrics for robust reviewer recommendations that could guide funding agencies such as the NSFC. The approach includes four major steps: profiling, representation, ranking, and evaluation. In general, profiling gathers data from candidate reviewers in the form of publications and project proposals to construct knowledge representation models. These models are then trained under the Word2vec framework to generate semantic embeddings of proposals and reviewer data. In the ranking phase, three different fusion methods-Average, SumRank, and CombRKP-are used to combine the similarities between proposals and reviewers into a ranked list of reviewers for each proposal. Then, tests were conducted on the system using NSFC data that focused on the Science and Technology Management and Policy division, G0404. The set included 12 proposals, 136 reviewers, 982 publications, and 167 projects. Below figure 3 show the framework of proposed reviewer recommendation method.

Among the metrics applied are Precision, Strict-Precision, and Recall. In these, RRM was compared with several baselines such as Research Analytics Framework (RAF) and Apache Lucene, outperforming the traditional method by 17% in Precision and 15% in Strict-Precision.

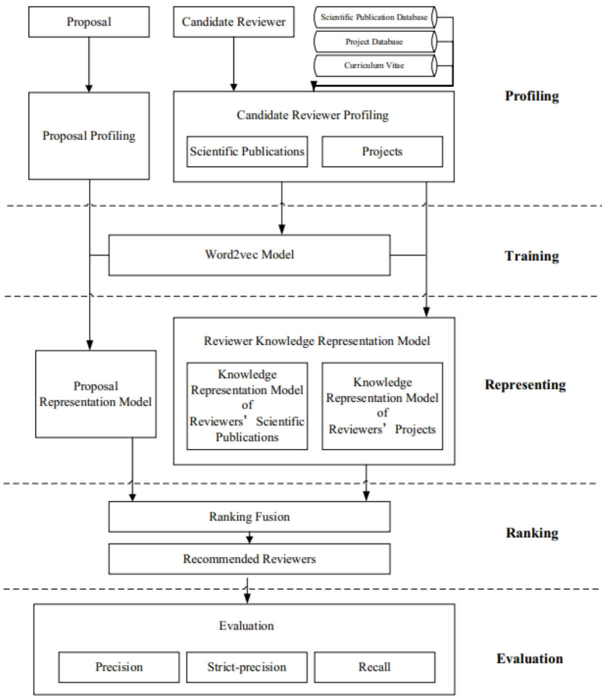

FIGURE 3. Framework of reviewer recommendation method by Liu, Wang, and Zhu. [51].

Advantages are that RRM handles multi-source data, integrates semantic relationships, and is adaptable to different domains of research. On the other hand, some of the limitations are a standard dataset for validation does not exist and diversity of reviewers, which is one of the big challenges. Further work involves dataset enlargement, improvements of diversity metrics, and enabling mechanisms for a broader selection of reviewers.

Additionally, Yong, Yao and Zhao [58] proposed a feature crossing-based solution to the reviewer assignment problem (RAP) for peer review with a personalized reviewer recommendation model, TRPRM. Their method is topic-topic feature crossing for deriving interactions among various research interests with a tree-based topic search space for increasing reviewer-paper matching accuracy. TRPRM integrates Attentive Factorization Machine (AFM) and a neural network for learning the most appropriate feature combinations at a semantic level. The article utilized a real-world dataset from the Dissertation Knowledge Discovery System (DKDS) of the University of Chinese Academy of Sciences, containing over 44,000 reviewer assignment records. During the evaluation of their model, they used AUC, Accuracy, Precision, and F1-score as baseline metrics, along with an additional novelty metric (Potential-Similarity) for accuracy and diversity guarantee in reviewer selection. The method proposed performed improved accuracy and novelty compared to baseline approaches such as DeepFM, AFM, and traditional Factorization Machines (FM), for reviewer diversity enhancement and preventing overdependence on past assignments. However, their future work is limited in the sense that it needs the expansion of the specialized vocabulary for improved system performance, and potential limits in handling cross-disciplinary reviewer recommendations. Future work can be accomplished in the addition of deep semantic learning techniques for improved topic understanding and reviewer assignment flexibility.

Next, Azad et al. [59] proposed a reviewer recommender system to provide improved article reviewer assignments based on a novel threshold similarity detection approach. This would, therefore, help the already present efficiency of the system-automated or even human assignment, which results in poor quality due to low calculation of accurate similarities, or utilizing criticisms provided by reviewers. Apart from ensuring that the number of articles a reviewer is assigned remains modest, and vice-versa, reviewer recommender system will strive for the maximum average confidence score of high resemblance between the experience of a selected reviewer and the contents of the document. The proposed general methodology of the reviewer recommender system approach was divided into four major stages: keyphrase extraction, pre-processing, data gathering, and reviewer suggestion. It involves the collection of the publication history, articles reviewed before, and confidence score taken from the records at the conference or from Google Scholar. Then the TF-IDF method ranks, with respect to key relevance for articles and expertizing reviewers. Cosine and Jaccard's similarity algorithms will finally determine those; a threshold would decide on the similarity level of the best suitable reviewer that is defined. Selection techniques involved in the Reviewer Recommendation phase of this approach are Random Walk and Hybrid Random Brute-force for allocating the reviewers while preserving the prescribed constraints. Figure 4 below shows the overview of the proposed system.

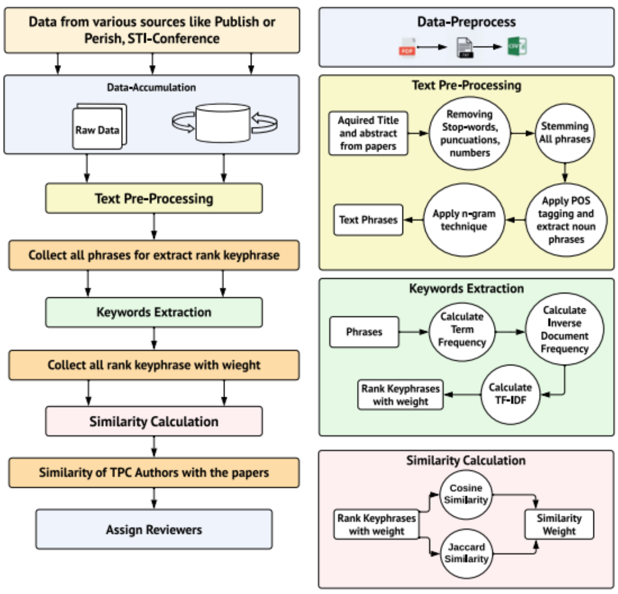

FIGURE 4. The architecture of method by Azad et al. [53].

Experiments were conducted on a dataset prepared from the 3rd International Conference on Sustainable Technologies for Industry 4.0, STI-2021, containing 67 articles, 57 reviewers, and corresponding confidence scores. The metrics of accuracy and precision showed that the share of suitable assignments for the reviewer recommender system was very high, workloads were well-balanced. The immediate advantages are that it can ensure a high-quality review by setting thresholds based on confidence, be scalable for larger datasets, and be flexible to real-world constraints. The limitation includes dependencies on high-quality data sources and probably high computational overhead while calculating similarities. Future work will include fine-tuning the process of discovering a similarity threshold and expanding the dataset for broader applicability.

In addition, Huang et al. [60] proposed a multilayer network diffusion-based model for reviewer recommendation, based on multilayer network diffusion and demanding no textual data. These typically rely on text-based information; incomprehensive or missing much of the time, hence inefficient. In the approach presented here, one considers a two-layer network of scholars and articles interconnected by means of co-authorship among scholars and bibliographic coupling among articles. It proposes a recommendation algorithm based on the process of diffusion; the model efficiently spreads out resources through these networks to find suitable reviewers based on research proximity and collaboration. The model is evaluated on data gathered from the NetSci-X 2018 conference, consisting of 70 reviewers, 2,804 scholars, and 24,197 articles. The key metrics used to evaluate the performance are three in total: recall, hit rate (HR), and ranking score (RS). The proposed approach outperformed all benchmark approaches, including text-based models and traditional network-based algorithms, with over 7.62% improvements in recall, 5.66% in HR, and 47.53% in RS. Its benefits are that it does not need to rely on the existence of textual data, scalable, and enhances accuracy for reviewer matching. However, limitations include the lack of consideration for real-world constraints such as conflicts of interest and reliance on simulated peer-review relations instead of real-world data. Future work will involve incorporating real peer-review data and addressing the cold-start problem for new articles.

Furthermore, Liao et al. [61] proposed a new model of Graph Neural Networks (GNN) to enable an academic reviewer recommendation task. The proposed method gives an effective way out on the critical challenge faced within the sparse and ambiguous reviewer-submission interaction with inability to reliably suppose any unobserved interaction as a negative sample. The proposed pseudo Neg-Label strategy in the RevGNN task promotes this contrastive graph learning. This efficiently solves the problem of false negatives and would reflect reviewer preference more precisely. It combines behavior encoding on the bipartite graph with semantic knowledge encoding leveraging the pre-trained scientific language model to enable accurate and scalable reviewer recommendations. RevGNN features a two-step encoding: First, it decouples the submission-scholar graph into behavior and knowledge representations using as decoupling GNNs and domain-specific language models, respectively. The second stage refines embeddings using a contrastive learning mechanism where negative samples are found by clustering-based pseudo-labeling to reduce false negatives. The proposed method will be evaluated on three benchmark datasets: Frontiers-4k, Frontiers-8k, and NIPS. Metrics used are Recall, NDCG, HR, and Precision. Among them, RevGNN outperformed the baseline models constantly and achieved as high as 34.16% improvements in Recall@20 compared to the traditional methods, while doing well in sparse graph settings. The advantages of RevGNN are that it can process sparse graphs efficiently, perform the task of reviewer identification rather well, and encode both behavioral and knowledge information. The drawbacks identified include dependence on sufficient computational resources and proper structuring of datasets. Future work will involve deploying it in real-world scenarios across different academic domains, while refining clustering methods to improve scalability.

VI. DISCUSSION AND WAY FORWARD

This section provides an overview of the related works that have been presented above in the section. Table 3 shows a high-level overview of the available research, such as major findings, datasets used, and evaluation metrics employed in recommender systems for peer reviewer assignment.

The current research landscape regarding the assignment of reviewers in peer review systems is characterized by significant evolution toward automating the selection process [62], [63]. Recent studies showcase an increased application of data-driven methodologies aimed at boosting effectiveness and equity in reviewer assignments, particularly through hybrid recommender systems. These systems incorporate various recommendation techniques, which include CB filtering, CF, and machine learning algorithms, all designed to improve the match between reviewers and articles.

Although automated reviewer assignment systems are more efficient, there are some ethical issues that need to be addressed. One of them is fairness, since systems that learn from experience might reinforce unfairness by repeatedly suggesting the same reviewers and thus potentially suppress early-career researchers. In addition, conflicts of interest may also arise when the system assigns reviewers with close institutional or personal relationships to the authors, particularly in the absence of human oversight. Furthermore, algorithmic transparency tends to be poor, especially where deep learning models are used, and hence uninterpretable and un-auditable assignments cannot be made. To enable proper deployment, the future systems will have to have mechanisms for conflict detection, promote ethical reviewer selection, and prefer explainability of algorithms.

TABLE 3. Summary of prior related research works.

References & Titles	Findings & Datasets	Evaluation Metrics
Protasiewicz et al. [51] A recommender system of reviewers and experts in reviewing problems	Finding: The article suggests a content-based recommender system to achieve automation in assigning reviewers to research proposals. The system offsets the inefficiency and subjectivity of manual reviewer assignment through semantic similarity between reviewer background and proposal content. Dataset: The system was evaluated on a real-world dataset of 4,000 reviewer profiles and 1,000 proposals of the National Center for Research and Development in Poland and the publication data were obtained from Microsoft Academic Search.	The system was assessed based on standard information retrieval measures. Precision = 78.4%, Recall = 85.6%, F1-score = 81.8%, which measures high agreement with expert judgments.
Kalmukov. [52] An Algorithm for Automatic Assignment of Reviewers to Papers	Finding: The article proposes a hybrid-based recommender system for automated reviewer assignment of articles in peer review systems. The method fuses heuristic algorithms, bipartite graph matching, and similarity-based techniques for improving reviewer-article allocation precision. The method preserves workload balance as well as expertise-based matching. Dataset: The algorithm is tested on real conference datasets, including nine years of conference data from CompSysTech (2010-2018).	The proposed method holds a 98-99% success rate when compared to the Hungarian algorithm but provides computational complexity reduction by a factor of Θ (n^2 time complexity). Testing is conducted on extensive simulations, statistical testing (ANOVA, regression), and comparisons with existing heuristic and brute-force methods.
Hoang et al. [53] Decision Support System for Solving Reviewer Assignment Problem	Finding: The article proposes a hybrid-based recommender system along with a decision support framework to allocate reviewers. The system has three significant modules: data collection, reviewer discovery, and group prediction of the reviewers. The system applies machine learning-based topic modeling and word embedding methods to analyze reviewer skills and allocate reviewers to articles. The system attempts to optimize the reviewer-	The system is evaluated on Normalized Discounted Cumulative Gain (nDCG) and precision. The system outperforms baseline models through improved reviewer-article assignment accuracy. Experimental results show that the proposed model achieves improved relevance and diversity in reviewer selection without conflict of interest.

	article assignment with expertise, relevance, and diversity to limit biases. Dataset: The approach is evaluated on the DBLP computer science bibliography dataset, which has 479 articles and more than 1000 reviewers.	
Tan et al. [54] Improved Reviewer Assignment Based on Both Word and Semantic Features	Finding: This study proposes a word and semantic-based iterative model (WSIM) to improve reviewer-article matching by addressing the incompleteness of reviewer data and interference from non-relevant articles. The model enhances similarity calculations using an improved language model (LM) and latent Dirichlet allocation (LDA) for feature extraction. The method incorporates Normalized Discounted Cumulative Gain (NDCG) for ranking-based similarity and employs an iterative model to refine assignments. Dataset: The approach is validated on two real-world datasets: (1) A computer science publication dataset with 400 reviewers and 100 articles, and (2) A large arXiv dataset with 1,885 reviewers and 685 articles.	This model is evaluated by precision, recall, F1-score, MAP, NDCG, and bpref. Experimental results show that WSIM improves recommendation accuracy at least 2.5% on top 20 compared to baseline models. The iterative model greatly reduces interference from out-of-relevance articles and makes more accurate reviewer-article assignments.
Pradhan et al. [55] A Proactive Decision Support System for Reviewer Recommendation in Academia	Finding: This article proposes a hybrid-based reviewer recommender system, TCRRec. The approach integrates Topic Network, Citation Network, and Reviewer Network to enhance reviewer-article matching. The system addresses issues like cold-start, data sparsity, and diversity in reviewer selection. The method applies Random Walk with Restart (RWR) for optimizing reviewer ranking and machine learning, social network analysis, and natural language processing (NLP) for discovering reviewer expertise. Dataset: The method is validated using AMiner and NIPS datasets, which consist of a massive repository of research articles, author metadata, and citation networks.	This model is evaluated with Precision@k, Mean Reciprocal Rank (MRR), Normalized Discounted Cumulative Gain (nDCG@k), authority, expertise, diversity, and coverage. The experiment confirms that TCRRec outperforms baseline recommendation methods, with improved reviewer-article relevance while alleviating the biases and conflicts of interest.
Kreutz and Schenkel. [56] RevASIDE: Assignment of Suitable Reviewer Sets for Publications from Fixed Candidate Pools	Finding: This article presents RevASIDE, a reviewer set recommendation model that suggests the right reviewer sets from a pre-defined candidate pool without manually defined reviewer profiles. The model considers five primary factors: expertise, authority, diversity, interest, and seniority in reviewing tasks. The reviewer-article matching is improved by integrating topic modeling (LDA, Doc2Vec, BERT) and vector-based representation of research profiles. The system also includes seniority balancing to achieve the optimal mix of junior and senior reviewers. Dataset: The framework is evaluated on three actual datasets of MOL'17, BTW'17, and ECIR'17 conferences on accepted articles and program committee members.	The system is evaluated on precision@k, MAP (Mean Average Precision), and Normalized Discounted Cumulative Gain (nDCG@k). The experiments demonstrate that RevASIDE significantly outperforms the traditional expert search practices, providing a more balanced and precise reviewer-article matching and diverse opinions in the review process.
Liu, Wang, and Zhu. [57] Reviewer recommendation method for scientifc research proposals: a case for NSFC	Finding: A reviewer recommendation method (RRM) for reviewing scientific research proposals, namely for the National Natural Science Foundation of China (NSFC), is introduced in this research. The approach applies word embedding techniques (Word2Vec) to construct vector-based word representations, capturing semantic and syntactic relationships. The model integrates reviewer and proposal knowledge representation models and employs ranking fusion techniques (Average, SumRank, CombRKP) to acquire ranked reviewer recommendations. Dataset: The approach is evaluated on actual NSFC proposal data, which includes 276,617 project proposals received by NSFC in 2020 and 45,656 funded proposals.	The model is checked by Precision, Strict-Precision, and Recall. The outcomes of the experiments show that accuracy is boosted by 17% by RRM while Strict-Precision is boosted by 15% over baseline methods. The approach provides useful solutions for improving reviewer assignment effectiveness at granting agencies.
Yong, Yao and Zhao. [58] Beyond Accuracy: A Feature Crossing Method for Chinese Thesis Reviewer Recommendation	Finding: The article proposes a feature crossing-based approach for enhancing reviewer assignment in peer review and offers a personalized recommendation model TRPRM. The method applies topic-topic feature crossing to express interactions among different research	The model is assessed by AUC, Accuracy, Precision, and F1-score for assignment performance measurement. In addition, Potential-Similarity (P@k) is also suggested as a novelty score to balance accuracy against reviewer heterogeneity. The model outperforms baselines such as DeepFM, Attentive Factorization Machine

	interests and a tree-based topic search space to support reviewer-paper matching. Dataset: The model is tested with a real-world dataset from the Dissertation Knowledge Discovery System (DKDS) of the University of Chinese Academy of Sciences, involving 44,533 reviewer assignment records across multiple disciplines.	(AFM), and traditional Factorization Machines (FM).
Azad et al. [59] A Reviewer Recommender System for Scientific Articles Using a New Similarity Threshold Discovery Technique	Finding: This article proposes a reviewer recommender system that uses a new similarity threshold discovery technique to improve reviewer selection for scientific documents. The technique attempts to maximize the average reviewer confidence score subject to constraints like a finite number of reviewers per document and a finite number of documents per reviewer. Cosine similarity and Jaccard similarity are employed to estimate reviewer-article affinity and optimize choices based on probability-based threshold discovery. Dataset: The evaluation was performed for a new dataset which was derived based on the 3rd International Conference on Sustainable Technologies for Industry 4.0 (STI—2021) reviewer lists, articles, and confidence scores. Additional reviewer expertise data were also extracted from Google Scholar, IEEE Xplore, and Scopus.	The system is assessed on Precision, Recall, and Confidence Score Analysis. Results are that higher similarity scores are positively correlated with reviewer confidence, and the proposed similarity threshold mechanism improves accuracy by 17% compared to traditional reviewer selection approaches. The approach ensures enhanced quality reviews as reviewers with higher expertise matching are selected.
Huang et al. [60] A Multilayer Network Diffusion-Based Model for Reviewer Recommendation	Finding: A reviewer recommendation algorithm is proposed in this study, which works based on a diffusion process in a multilayer network, preventing dependency on textual information. Scholar-article relationships, reviewer collaborations, and bibliographic coupling between articles are used by the model for better reviewer-article matching. Contrary to classical text mining-based techniques, in this approach, a diffusion process in a two-layer network is utilized, picking up the significance of reviewers to submissions based on scholarly collaborations as well as citation relationships. Dataset: It is tested on peer-review performance from a conference on science consisting of 70 reviewers and 2,804 researchers. The article data are drawn from Scopus in the form of 24,197 articles and bibliographic citations between them.	The model is ranked on the basis of Recall, Hit Rate (HR), and Ranking Score (RS). From the results, it is seen that the diffusion-based model is superior to basic machine learning, graph random walk, and matrix factorization-based models with improved recall by 7.62%, HR by 5.66%, and ranking score by 47.53%. The study establishes that network-based diffusion models are more accurate than text-based methods such as TF-IDF, LDA, and BERT embeddings.
Liao et al. [61] RevGNN: Negative Sampling Enhanced Contrastive Graph Learning for Academic Reviewer Recommendation	Finding: In this article, RevGNN, a graph neural network (GNN) model with added negative sampling strategies, is suggested to improve reviewer recommendation in academia. The model remedies the false negative sampling caused by unseen reviewer-submission matching. RevGNN employs contrastive graph learning (GCL) with Pseudo Neg-Label strategy for optimizing negative sample selection for improved reviewer-article matching. Both reviewer preference and submission semantic understanding are well represented using a two-stage encoder in the method. Dataset: The model is evaluated on three real-world datasets: (1) Frontiers-4k, (2) Frontiers-8k, and (3) NIPS (Neural Information Processing Systems) reviewer dataset, which contains over 10,000 reviewers and 20,000 submissions.	The model is compared in terms of Recall@20, Precision@20, Hit Rate@20, and Normalized Discounted Cumulative Gain (NDCG@20). Experimental results show that RevGNN outperforms state-of-the-art baselines by 32.19% in terms of Recall@20 over RecVAE and 34.88% in terms of Precision@20. The method significantly enhances reviewer selection accuracy and mitigates biases created by virtue of false negative sampling.

Recently, more of what's called hybrid systems that combine different types of recommendation methods have been used [64] Due to the integration of strengths found in various methods, hybrid systems are realized to lessen drawbacks associated with pure or standalone methodologies, hence increasing accuracy in matching [65].

On top of this, blockchain technology may prove to be a solution in making peer review more transparent and honest, possibly through better tracking of the contributions made by the reviewers and also reducing fraudulent activities [66], [67].

There has been growing concern about the fairness in assigning reviewers. Present-day studies aim at reducing bias in the selection process to ensure that diverse views are included in peer review [68], [69].

VII. CONCLUSION AND FUTURE WORK

In this review, we reviewed hybrid-based reviewer recommendation systems and how they have an advantage over traditional approaches using multiple techniques such as graph neural networks, knowledge graphs, and machine learning to improve reviewer-article matching. From our review, we noticed that hybrid models achieve success in addressing cold-start issues, data sparsity, and bias, leading to more accurate and fair assignments.

As future work, we plan to implement a hybrid-based recommender system on a peer reviewer assignment dataset to further evaluate its performance and effectiveness in real-world academic settings.

ACKNOWLEDGMENT

We thank the reviewers and editor of the journal for their guidance in improving the quality of our article.

FUNDING STATEMENT

There are no funding agencies supporting the research work.

AUTHOR CONTRIBUTIONS

Ye-Xin Lim: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation;

Su-Cheng Haw: Project Administration, Supervision, Writing – Review & Editing;

Jayapradha J: Writing – Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS

This research did not involve human participants, animal subjects, or sensitive personal data, and therefore did not require ethical approval.

REFERENCES

[1] R. Muscarella and L. Poorter, "Ten simple rules for managing communications with a large number of coauthors," *PLoS Computational Biology*, vol. 18, no. 6, p. e1010185, 2022. DOI: https://doi.org/10.1371/journal.pcbi.1010185

[2] R. Smith, "Peer review: A flawed process at the heart of science and journals," *Journal of the Royal Society of Medicine*, vol. 99, no. 4, pp. 178–182, 2006. DOI: https://doi.org/10.1258/jrsm.99.4.178

[3] A. O. Ogundea, M. O. Odim, O. O. Olaniyan, T. O. Ojewumi, A. O. Oguntunde, M. A. Fayemiwo, T. A. Olowookere, and T. H. Bolanle., "The design of a hybrid model-based journal recommendation system," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 6, pp. 35–45, 2020. DOI: https://doi.org/10.25046/aj0506139

[4] F.T. Abdul Hussien, A.M.S. Rahma and H.B. Abdul Wahab, "Recommendation Systems for E-commerce Systems: An Overview," *Journal of Physics: Conference Series*, vol. 1897, no. 1, p. 012024, 2021. DOI: https://doi.org/10.1088/1742-6596/1897/1/012024

[5] T.-E. Tai, S.-C. Haw, W.-E. Kong and K.-W. Ng, "Performance Evaluation of Machine Learning Techniques on Resolution Time Prediction in Helpdesk Support System," *International Journal on Robotics, Automation and Sciences*, vol. 6, no. 2, pp. 59–68, 2024. DOI: https://doi.org/10.33093/ijoras.2024.6.2.9

[6] S.-C. Haw, J. Jayaram, E.A. Anaam and H.A. Santoso, "Exploring Recommender Systems in Healthcare: A Review on Methods, Applications and Evaluations," *International Journal on Robotics, Automation and Sciences*, vol. 6, no. 2, pp. 6–15, 2024. DOI: https://doi.org/10.33093/ijoras.2024.6.2.2

[7] Y. Deldjoo, M. Schedl, P. Cremonesi and G. Pasi, "Recommender Systems Leveraging Multimedia Content," *ACM Computing Surveys*, vol. 53, no. 5, p. 110, 2020. DOI: https://doi.org/10.1145/3407190

[8] S. Wang, L. Hu, Y. Wang, X. He, Q. Z. Sheng, M. A. Orgun, L. Cao, F. Ricci, and P. S. Yu., "Graph Learning based Recommender Systems: A Review," *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 4493–4500, 2021. DOI: https://doi.org/10.24963/ijcai.2021/630

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

[10] M.F. Aljunid and M. Dh, "An Efficient Deep Learning Approach for Collaborative Filtering Recommender System," *Procedia Computer Science*, vol. 171, pp. 1054–1061, 2020. DOI: https://doi.org/10.1016/j.procs.2020.04.090

[11] P. Gonzalez, G.S. Wilson and A.J. Purvis, "Peer review in academic publishing: Challenges in achieving the gold standard," *Journal of University Teaching and Learning Practice*, vol. 19, no. 5, p. 1, 2022. DOI: https://doi.org/10.53761/1.19.5.1

[12] Z. Gu, Y. Cai, S. Wang, M. Li, J. Qiu, S. Su, X. Du, and Z. Tian., "Adversarial attacks on content-based filtering journal recommender systems," *Computers, Materials & Continua*, vol. 64, no. 3, pp. 1755–1770, 2020. DOI: https://doi.org/10.32604/cmc.2020.010739

[13] R.J. DiDomenico, W.L. Baker and S.T. Haines, "Improving peer review: What reviewers can do," *American Journal of Health-System Pharmacy*, vol. 74, no. 24, pp. 2080–2084, 2017. DOI: https://doi.org/10.2146/ajhp170190

[14] S.T. Haines, W.L. Baker and R.J. DiDomenico, "Improving peer review: What journals can do," *American Journal of Health-System Pharmacy*, vol. 74, no. 24, pp. 2086–2090, 2017. DOI: https://doi.org/10.2146/ajhp170188

[15] A. Kumar and S.N. Ahmed, "Challenges faced in the peer review system in open access journals," *Asian Journal of Medical Sciences*, vol. 13, no. 12, pp. 1–5, 2022. DOI: https://doi.org/10.3126/ajms.v13i12.49361

[16] T.C.M. Bakker and J.F.A. Traniello, "Peer-review reciprocity and commitment to manuscript evaluation," *Behavioral Ecology and Sociobiology*, vol. 73, no. 3, p. 38, 2019. DOI: https://doi.org/10.1007/s00265-019-2647-2

[17] A. Day, "Exploratory analysis of text duplication in peer-review reveals peer-review fraud and paper mills," *Scientometrics*, vol. 127, no. 10, pp. 6183–6200, 2022. DOI: https://doi.org/10.1007/s11192-022-04504-5

[18] B. Tabatadze, "Technological Aspects of Open Journal Systems (OJS)," *Journal of Technical Science and Technologies*, vol. 8, no. 1, pp. 23–29, 2024. DOI: https://doi.org/10.31578/jtst.v8i1.151

[19] W.H.W. Ishak, F.M. Yamin, M.F.M. Mohsin and M.F. Mansor, "A Comparative Review of Conference Management Systems," *Journal of Technology and Operations Management*, vol. 18, no. 2, pp. 87–93, 2023. DOI: https://doi.org/10.32890/jtom2023.18.2.7

[20] S. Zhang, L. Yao, A. Sun and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Computing Surveys*, vol. 52, no. 1, p. 5, 2019. DOI: https://doi.org/10.1145/3285029

[21] S. Dwivedi and V.S.K. Roshni, "Recommender system for big data in education," *Proceedings of the 5th National Conference on E-Learning and E-Learning Technologies (ELELTECH 2017)*, pp. 1–5, 2017.

DOI: https://doi.org/10.1109/ELELTECH.2017.8074993

[22] G. Geetha, M. Safa, C. Fancy and D. Saranya, "A Hybrid Approach using Collaborative Filtering and Content Based Filtering for Recommender System," *Journal of Physics: Conference Series*, vol. 1000, no. 1, p. 012101, 2018. DOI: https://doi.org/10.1088/1742-6596/1000/1/012101

[23] M. Srifi, A. Oussous, A.A. Lahcen and S. Mouline, "Recommender systems based on collaborative filtering using review texts—A survey," *Information*, vol. 11, no. 6, p. 317, 2020. DOI: https://doi.org/10.3390/INFO11060317

[24] R. Widayanti, M.H.R. Chakim, C. Lukita, U. Rahardja and N. Lutfiani, "Improving Recommender Systems using Hybrid Techniques of Collaborative Filtering and Content-Based Filtering," *Journal of Applied Data Sciences*, vol. 4, no. 3, pp. 196–207, 2023. DOI: https://doi.org/10.47738/jads.v4i3.115

[25] Y. Afoudi, M. Lazaar and M. Al Achhab, "Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network," *Simulation Modelling Practice and Theory*, vol. 113, p. 102375, 2021. DOI: https://doi.org/10.1016/j.simpat.2021.102375

[26] J. Wei, J. He, K. Chen, Y. Zhou and Z. Tang, "Collaborative filtering and deep learning based recommendation system for cold start items," *Expert Systems with Applications*, vol. 69, pp. 29–39, 2017. DOI: https://doi.org/10.1016/j.eswa.2016.09.040

[27] J. Chen, H. Zhang, X. He, L. Nie, W. Liu and T.S. Chua, "Attentive collaborative filtering: Multimedia recommendation with item- and component-level attention," *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2017)*, pp. 335–344, 2017. DOI: https://doi.org/10.1145/3077136.3080797

[28] Z. Zheng, X. Li, M. Tang, F. Xie and M. R. Lyu, "Web Service QoS Prediction via Collaborative Filtering: A Survey," *IEEE Transactions on Services Computing*, vol. 15, no. 4, pp. 2094–2109, 2022. DOI: https://doi.org/10.1109/TSC.2020.2995571

[29] G. George and A.M. Lal, "Hy-MOM: Hybrid Recommender System Framework Using Memory-Based and Model-Based Collaborative Filtering Framework," *Cybernetics and Information Technologies*, vol. 22, no. 1, pp. 100–115, 2022. DOI: https://doi.org/10.2478/cait-2022-0009

[30] A. Pujahari and D.S. Sisodia, "Model-based collaborative filtering for recommender systems: An empirical survey," *Proceedings of the 1st International Conference on Power, Control and Computing Technologies (ICPC2T 2020)*, pp. 251–255, 2020. DOI:https://doi.org/10.1109/ICPC2T48082.2020.9071454

[31] G. Chornous, I. Nikolskyi, M. Wyszyński, G. Kharlamova and P. Stolarczyk, "A hybrid user-item-based collaborative filtering model for e-commerce recommendations," *Journal of International Studies*, vol. 14, no. 4, pp. 157–170, 2021. DOI: https://doi.org/10.14254/2071-8330.2021/14-4/11

[32] P.B. Thorat, R.M. Goudar and S. Barve, "Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System," *International Journal of Computer Applications*, vol. 110, no. 4, pp. 31–36, 2015. DOI: https://doi.org/10.5120/19308-0760

[33] Z. Yang, B. Wu, K. Zheng, X. Wang and L. Lei, "A survey of collaborative filtering-based recommender systems for mobile internet applications," *IEEE Access*, vol. 4, pp. 3273–3287, 2016. DOI: https://doi.org/10.1109/ACCESS.2016.2573314

[34] M. Riyahi and M.K. Sohrabi, "Providing effective recommendations in discussion groups using a new hybrid recommender system based on implicit ratings and semantic similarity," *Electronic Commerce Research and Applications*, vol. 40, p. 100938, 2020. DOI: https://doi.org/10.1016/j.elerap.2020.100938

[35] S. Sharma, V. Rana, and M. Malhotra, "Automatic recommendation system based on hybrid filtering algorithm," *Education and Information Technologies*, vol. 27, no. 2, pp. 1523–1538, 2022. DOI: https://doi.org/10.1007/s10639-021-10643-8

[36] F.S. Gohari, H. Haghighi and F.S. Aliee, "A semantic-enhanced trust based recommender system using ant colony optimization," *Applied Intelligence*, vol. 46, no. 2, pp. 380–398, 2017. DOI: https://doi.org/10.1007/s10489-016-0830-y

[37] M. Al-Hassan, H. Lu and J. Lu, "A semantic enhanced hybrid recommendation approach: A case study of e-Government tourism service recommendation system," *Decision Support Systems*, vol. 72, pp. 97–109, 2015. DOI: https://doi.org/10.1016/j.dss.2015.02.001

[38] N. Saidani, K. Adi and M.S. Allili, "A semantic-based classification approach for an enhanced spam detection," *Computers & Security*, vol. 94, p. 101716, 2020. DOI: https://doi.org/10.1016/j.cose.2020.101716

[39] N.A.N.B.M. Romzi, S.-C. Haw, W.-E. Kong, H.A. Santoso and G.-K. Tong, "Generative AI Recommender System in E-Commerce," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 14, no. 6, pp. 1823–1835, 2024. DOI: https://doi.org/10.18517/ijaseit.14.6.10509

[40] M.O. Ayemowa, R. Ibrahim and M.M. Khan, "Analysis of Recommender System Using Generative Artificial Intelligence: A Systematic Literature Review," *IEEE Access*, vol. 12, pp. 87742–87766, 2024. DOI: https://doi.org/10.1109/ACCESS.2024.3416962

[41] Y. Deldjoo, Z. He, J. McAuley, A. Korikov, S. Sanner, A. Ramisa, R. Vidal, M. Sathiamoorthy, A. Kasirzadeh, and S. Milano., "A Review of Modern Recommender Systems Using Generative Models (Gen-RecSys)," *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 6448–6458, 2024. DOI: https://doi.org/10.1145/3637528.3671474

[42] C.P. Ling Tung, S.-C. Haw, W.-E. Kong and P. Naveen, "Movie Recommender System Based on Generative AI," *Proceedings of the 2024 International Symposium on Parallel Computing and Distributed Systems (PCDS)*, pp. 1–7, 2024. DOI:https://doi.org/10.1109/PCDS61776.2024.10743897

[43] M.J. Mokarrama, S. Khatun and M.S. Arefin, "A content-based recommender system for choosing universities," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 28, no. 4, pp. 2070–2083, 2020. DOI: https://doi.org/10.3906/ELK-1911-37

[44] S. Sridhar, D. Dhanasekaran and G.C. P. Latha, "Content-Based Movie Recommendation System Using MBO with DBN," *Intelligent Automation and Soft Computing*, vol. 35, no. 3, pp. 3031–3043, 2023. DOI: https://doi.org/10.32604/iasc.2023.030361

[45] R. Mishra and S. Rathi, "Efficient and Scalable Job Recommender System Using Collaborative Filtering," *Lecture Notes in Electrical Engineering*, vol. 635, pp. 931–939, 2020. DOI: https://doi.org/10.1007/978-981-15-1420-3_91

[46] G. Chalkiadakis, I. Ziogas, M. Koutsmanis, E. Streviniotis, C. Panagiotakis and H. Papadakis, "A Novel Hybrid Recommender System for the Tourism Domain," *Algorithms*, vol. 16, no. 4, p. 215, 2023. DOI: https://doi.org/10.3390/a16040215

[47] M.V. Farashah, A. Etebarian, R. Azmi and R.E. Dastjerdi, "A hybrid recommender system based-on link prediction for movie baskets analysis," *Journal of Big Data*, vol. 8, no. 1, p. 10, 2021. DOI: https://doi.org/10.1186/s40537-021-00422-0

[48] E. Asani, H. Vahdat-Nejad and J. Sadri, "Restaurant recommender system based on sentiment analysis," *Machine Learning with Applications*, vol. 6, p. 100114, 2021. DOI: https://doi.org/10.1016/j.mlwa.2021.100114

[49] I. Karabila, N. Darraz, A. El-Ansari, N. Alami and M. El Mallahi, "Enhancing Collaborative Filtering-Based Recommender System Using Sentiment Analysis," *Future Internet*, vol. 15, no. 7, p. 235, 2023. DOI: https://doi.org/10.3390/fi15070235

[50] A. Zhang, Y. Chen, L. Sheng, X. Wang and T.-S. Chua, "On Generative Agents in Recommendation," *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2024)*, pp. 1807–1817, 2024. DOI: https://doi.org/10.1145/3626772.3657844

[51] J. Protasiewicz, W. Pedrycz, M. Kozłowski, S. Dadas, T. Stanisławek, A. Kopacz, and M. Gałężewska., "A recommender system of reviewers and experts in reviewing problems," *Knowledge-Based Systems*, vol. 106, pp. 164–178, 2016.

DOI: https://doi.org/10.1016/j.knosys.2016.05.041
[52] Y. Kalmukov, "An algorithm for automatic assignment of reviewers to papers," *Scientometrics*, vol. 124, no. 3, pp. 1957–1974, 2020. DOI: https://doi.org/10.1007/s11192-020-03519-0
[53] D.T. Hoang, N.T. Nguyen, B. Collins and D. Hwang, "Decision Support System for Solving Reviewer Assignment Problem," *Cybernetics and Systems*, vol. 52, no. 5, pp. 393–411, 2021. DOI: https://doi.org/10.1080/01969722.2020.1871227
[54] S. Tan, Z. Duan, S. Zhao, J. Chen and Y. Zhang, "Improved reviewer assignment based on both word and semantic features," *Information Retrieval Journal*, vol. 24, no. 3, pp. 323–349, 2021. DOI: https://doi.org/10.1007/s10791-021-09390-8
[55] T. Pradhan, S. Sahoo, U. Singh and S. Pal, "A proactive decision support system for reviewer recommendation in academia," *Expert Systems with Applications*, vol. 169, p. 114331, 2021. DOI: https://doi.org/10.1016/j.eswa.2020.114331
[56] C.K. Kreutz and R. Schenkel, "RevASIDE: Assignment of Suitable Reviewer Sets for Publications from Fixed Candidate Pools," *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM 2021)*, 2021. DOI: https://doi.org/10.1145/3487664.3487673
[57] X. Liu, X. Wang and D. Zhu, "Reviewer recommendation method for scientific research proposals: a case for NSFC," *Scientometrics*, vol. 127, no. 6, pp. 3285–3305, 2022. DOI: https://doi.org/10.1007/s11192-022-04389-4
[58] Y. Yong, Z. Yao and Y. Zhao, "Beyond Accuracy: A Feature Crossing Method for Chinese Thesis Reviewer Recommendation," *Proceedings of the 2021 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 1609–1614, 2021. DOI: https://doi.org/10.1109/SMC52423.2021.9658668
[59] S. Azad, M. A. Hoque, N. A. Rimon, M. M. S. Habib, M. Mahmud, M. S. Kaiser, and M. R. Karim., "A Reviewer Recommender System for Scientific Articles Using a New Similarity Threshold Discovery Technique," *Lecture Notes in Networks and Systems*, vol. 658, pp. 481–494, 2023. DOI: https://doi.org/10.1007/978-981-19-9483-8_42
[60] Y. Huang, S. Xu, S. Cai and L. Lü, "A multilayer network diffusion-based model for reviewer recommendation," *Chinese Physics B*, vol. 33, no. 3, p. 038901, 2024. DOI: https://doi.org/10.1088/1674-1056/ad181d
[61] W. Liao, Y. Zhu, Y. Li, Q. Zhang, Z. Ou, and X. Li, "RevGNN: Negative Sampling Enhanced Contrastive Graph Learning for Academic Reviewer Recommendation," *Proceedings of the 2024 ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024. DOI: https://doi.org/10.1145/3679200
[62] J.A. Drozdz and M.R. Ladomery, "The Peer Review Process: Past, Present, and Future," *British Journal of Biomedical Science*, vol. 81, p. 12054, 2024. DOI: https://doi.org/10.3389/bjbs.2024.12054
[63] B. Aczel, A.-S. Barwich, A. B. Diekman, A. Fishbach, R. L. Goldstone, P. Gomez, O. E. Gundersen, P. T. von Hippel, A. O. Holcombe, S. Lewandowsky, N. Nozari, F. Pestilli, and J. P. A. Ioannidis., "The present and future of peer review: Ideas, interventions, and evidence," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 122, no. 5, p. e2401232121, 2025. DOI: https://doi.org/10.1073/pnas.2401232121
[64] S.-A. Teh, S.-C. Haw and H. A. Santoso, "Hybrid-based Research Article Recommender System," *International Journal of Membrane Science and Technology*, vol. 10, no. 2, pp. 1587–1606, 2023. DOI: https://doi.org/10.15379/ijmst.v10i2.1830
[65] Z.-T. Yap, S.-C. Haw and N.E.B. Ruslan, "Hybrid-based food recommender system utilizing KNN and SVD approaches," *Cogent Engineering*, vol. 11, no. 1, p. 2436125, 2024. DOI: https://doi.org/10.1080/23311916.2024.2436125
[66] C.H. Morales-Alarcón, E. Bodero-Poveda, H.M. Villa-Yánez, and P.A. Buñay-Guisñan, "Blockchain and Its Application in the Peer Review of Scientific Works: A Systematic Review," *Publications*, vol. 12, no. 4, p. 40, 2024. DOI: https://doi.org/10.3390/publications12040040
[67] P. Shrestha, S. Bhunia, A. Carvalho, C. Anderson and G. Lee, "A Case Study on Blockchain-based Anonymous Reviewer Incentive Token (BARIT)," *Proceedings of the 2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1564–1571, 2024. DOI:https://doi.org/10.1109/SMC54092.2024.10831118
[68] A.C. Ribeiro, A. Sizo and L.P. Reis, "Investigating the reviewer assignment problem: A systematic literature review," *Journal of Information Science*, 2023. DOI: https://doi.org/10.1177/01655515231176668
[69] K. Kousha and M. Thelwall, "Artificial intelligence to support publishing and peer review: A summary and review," *Learned Publishing*, vol. 37, no. 1, pp. 4–12, 2024. DOI: https://doi.org/10.1002/leap.1570