

## Article

# Bias Reduction News Recommendation System

Shaina Raza 

Department of Computer Science, Toronto Metropolitan University, 350 Victoria St,  
Toronto, ON M5B 2K3, Canada; shaina.raza@torontomu.ca

**Abstract:** News recommender systems (NRS) are crucial for helping users navigate the vast amount of content available online. However, traditional NRS often suffer from biases that lead to a narrow and unfair distribution of exposure across news items. In this paper, we propose a novel approach, the Contextual-Dual Bias Reduction Recommendation System (C-DBRRS), which leverages Long Short-Term Memory (LSTM) networks optimized with a multi-objective function to balance accuracy and diversity. We conducted experiments on two real-world news recommendation datasets and the results indicate that our approach outperforms the baseline methods, and achieves higher accuracy while promoting a fair and balanced distribution of recommendations. This work contributes to the development of a fair and responsible recommendation system.

**Keywords:** BIAS; recommender systems; news; diversity; large language models

## 1. Introduction

The proliferation of digital media and information has led to an exponential increase in the availability of news content. However, this abundance of information poses challenges for users in selecting relevant and high-quality content. In response, news recommender systems (NRS) [1] have become vital components of online news platforms. These systems provide personalized recommendations to users based on their behavior, preferences, and interactions [2]. The primary goal of NRS is to enhance user engagement and retention by delivering tailored news consumption experiences [3].

We present an example of a NRS, in Figure 1, which considers User A, who frequently reads news articles related to environmental issues and sustainability. An NRS that focuses solely on accuracy would continuously recommend articles related to the environment and sustainability, based on User A's past behavior. This can lead to an 'echo chamber' effect [1], where the user is exposed only to news and opinions that align with their existing beliefs and interests. Conversely, the NRS prioritizing fairness might recommend a broader range of topics, including politics, economics, technology, and social issues.

In this study, we define 'fairness' in the context of NRS as a balance between personalized recommendations and promoting content "diversity". This approach aims to broaden users' exposure to a variety of ideas and perspectives, though it might sometimes lead to less tailored recommendations, potentially affecting user engagement and satisfaction [4]. NRS often exhibit bias by prioritizing content that aligns with a user's existing preferences, thus reinforcing their current beliefs and interests [5,6]. To counter this, fairness in NRS is about offering a diverse and representative range of content. We introduce a novel method that seeks to strike a balance between personalization and diversity, with the goal of optimizing both the relevance of recommendations and the breadth of content exposure.

State-of-the-art NRS predominantly concentrate on enhancing recommendation accuracy [2,7], which is commendable. However, there is a growing recognition of the need to balance accuracy with fairness [7–9]. In NRS, a focus solely on accuracy may inadvertently lead to limited content exposure, potentially creating echo chambers [10]. Conversely, a system prioritizing fairness might offer a diverse range of content but with recommendations that are less aligned with individual user preferences [11]. Hence, there is an imperative



**Citation:** Raza, S. Bias Reduction News Recommendation System. *Digital* **2024**, *4*, 92–103. <https://doi.org/10.3390/digital4010003>

Academic Editor: Phivos Mylonas

Received: 1 November 2023

Revised: 10 December 2023

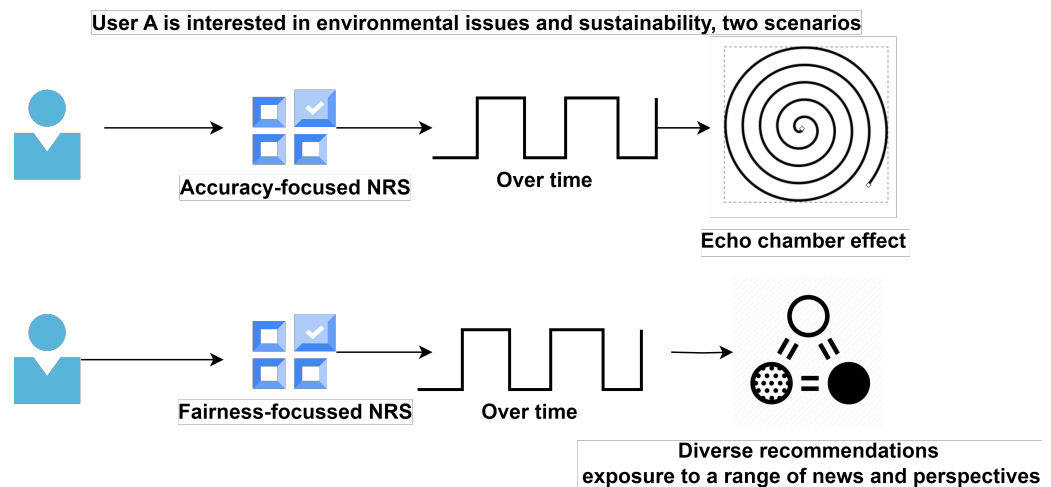
Accepted: 25 December 2023

Published: 28 December 2023



**Copyright:** © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

need for strategies that simultaneously achieve both accuracy and fairness, aiming for a more inclusive news consumption experience.



**Figure 1.** User A news recommendation dilemma: balancing accuracy and fairness (diversity aspect).

This study introduces a novel approach to news recommendation that optimizes both accuracy and fairness. Fairness is a wide subject, encompassing various dimensions and interpretations depending on the context. In the realm of information systems and algorithms [4,12], fairness often involves considerations like equal representation, unbiased treatment of different groups, and equitable distribution of resources or opportunities. It is about ensuring that systems and decisions do not favor one group or individual over another unjustly and that they reflect a balanced and inclusive approach. This is particularly crucial in areas such as news recommendations, hiring practices, and financial services, where the impact of unfairness can be significant on both individuals and society at large.

In this work, we particularly refer to ‘diversity’ for fairness, ensuring that the recommendations cater to a wide range of interests and perspectives [13]. We propose a multi-objective optimization strategy for an NRS. Our contributions include the development of an algorithm that balances multiple factors, such as user preferences, content diversity, and fairness in exposure to different news sources. This approach not only enhances the relevance of recommendations but also promotes a diverse and inclusive news consumption experience. Our aim is to create a more holistic and responsible NRS.

We introduce the “Contextual Dual Bias Reduction Recommendation System (C-DBRRS)”, which ensures fairness in item and exposure aspects of news recommendation. Built on LSTM networks and optimized with a multi-objective function, C-DBRRS harmonizes accuracy and diversity in news recommendations. The hyperparameter  $\lambda$  allows for a tunable balance between relevance (precision, recall, NDCG) and fairness (Gini coefficient). This is a novel approach, as it recognizes the trade-offs between providing relevant recommendations to users and ensuring a fair representation of items.

## 2. Literature Overview

News recommender systems (NRS) have been a widely researched topic in recent years due to the increasing amount of online news content and the need for personalized recommendations [1]. Traditional NRS are designed to provide users with personalized news articles based on their past behavior, preferences, and interests [14]. Collaborative filtering and content-based filtering are two common approaches used in recommender systems [15]. Collaborative filtering recommends news articles to a user based on the preferences of similar users, whereas content-based filtering recommends articles based on the content of the articles and the user’s past behavior.

Accuracy is a critical factor in the performance of recommender systems [1]. The accuracy of a recommender system refers to its ability to recommend items that are relevant

and of interest to the user [3]. Commonly used metrics for evaluating the accuracy of recommender systems include the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the F1-score [16]. Several approaches have been proposed to improve the accuracy of recommender systems, such as incorporating additional contextual information [17], or using hybrid recommendation algorithms that combine the strengths of both collaborative filtering and content-based filtering [18].

Bias in recommender systems can manifest in various ways [8,9,19,20], including popularity bias, where the system tends to recommend popular items, leading to a lack of exposure for less popular or niche items [1]. Additionally, there might be biases related to gender, race, or other demographic factors, leading to discriminatory recommendations [21]. Researchers have proposed various fairness metrics to evaluate the fairness of recommender systems. These include statistical parity, which requires that the recommendations are independent of a protected attribute (e.g., gender, race), and disparate impact, which measures the ratio of positive outcomes for the protected group to the positive outcomes for the non-protected group [22,23]. A related approach [24] proposed a fairness-aware ranking algorithm that considers both the utility of the recommendations to the user and the fairness towards the items. Another approach is to re-rank the recommendations generated by a standard recommender system to improve fairness [20,23].

Fairness in recommender systems has gained attention due to the potential biases that can arise from the recommendation algorithms. Studies have shown that recommender systems can inadvertently reinforce existing biases in the data, leading to unfair recommendations for certain groups of users [21]. For example, a study [25] showed that a music recommender system was biased towards popular artists, leading to less exposure for less popular or niche artists. Several approaches have been proposed to address fairness in recommender systems, such as re-ranking the recommendations [22] or modifying the recommendation algorithm to incorporate fairness constraints [20].

Recent advancements in NRS emphasize a balance between accuracy and fairness, particularly in terms of diversity [5]. A novel multi-objective optimization strategy has been proposed to refine recommender system models. For instance, ref. [26] discusses an innovative algorithm utilizing multi-objective optimization. Similarly, ref. [13] proposes a framework to optimize recommender systems, emphasizing fairness across multiple stakeholders. Moreover, the use of multi-objective optimization for recommending online learning resources is effectively demonstrated [27]. Finally, ref. [28] highlights the application of big data in enhancing recommendation systems through multi-objective optimization.

There is often a trade-off between fairness and accuracy in recommender systems [29]. Improving fairness in the recommendations may lead to a decrease in accuracy, and vice versa. For example, a study [24] showed that incorporating fairness constraints into the recommendation algorithm led to a decrease in recommendation accuracy. Similarly, a study [30] showed that re-ranking the recommendations to improve fairness led to a decrease in accuracy. Therefore, it is important to carefully consider the trade-off between fairness and accuracy when designing and evaluating recommender systems.

### 3. Materials and Methods

#### 3.1. Data

In our study, we utilized two real-world news recommendation datasets, namely MIND-small and Outbrain Click Prediction, which offer a comprehensive view of user interactions and preferences in news consumption.

**MIND-small [2]:** Derived from the larger MIND dataset, the MIND-small dataset, curated by Microsoft team, captures the interactions of 50,000 users on Microsoft News over a one-month period. For our study, we focused on this subset, analyzing their interactions with news articles, and associated metadata such as titles, categories, and abstracts.

**Outbrain Click Prediction [31]:** Sourced from a Kaggle competition hosted by Outbrain, this dataset provides an extensive record of user page views and clicks across various

publisher sites in the United States over a span of 14 days. It offers valuable insights into user behaviors regarding displayed and clicked ads.

Our data preparation methodology involves several key stages to ensure the datasets were optimally configured for our recommender system. We started with fundamental data-cleaning procedures to enhance the quality and reliability of our datasets. This involved the removal of duplicate records and the addressing of missing or incomplete data through exclusion criteria. To enrich our analysis, we extracted several critical features from the datasets:

- Textual Embeddings (BERT): we converted the text content of news articles into numerical vector representations using the BERT (base-uncased) model to encapsulate their semantic content.
- Topic Modeling (LDA): articles were categorized into specific genres or themes using Latent Dirichlet Allocation (LDA).
- Sentiment Analysis (VADER): we employed the VADER [32] tool to analyze the emotional tone of articles, classifying them as positive, negative, or neutral.

Post feature extraction, we standardized the scale of numerical features to ensure uniformity. Subsequently, we integrated detailed user interaction data, including clicks, views, and duration of engagement with each article, with the article features. This integration facilitated the creation of comprehensive user–article interaction profiles.

For effective processing using Long Short-Term Memory (LSTM) networks, we structured our data in the following manner:

- Unified Input Vectors: each user–article interaction was represented as a unified vector, consolidating user behavioral data with the extracted article content features.
- Time Series Formation: we structured the data into time series to capture the temporal dynamics of user interactions, a crucial aspect of LSTM processing.

The final stage in our data preparation was the division of the dataset into distinct sets for training, validation, and testing, following the 80-10-10 scheme in chronological temporal order. We combined and structured data from the MIND-small and Outbrain Click Prediction datasets to capture the temporal dynamics of user interactions with news articles. The following elements constitute our data structure:

- User and Article Identifiers: unique IDs for users and articles to track interactions.
- Interaction Timestamps: capture the timing of each interaction, crucial for time series analysis.
- Interaction Types: categorized as clicks, views, and engagement duration.
- Content Features: textual embeddings, topic categories, and sentiment scores for articles from MIND-small.
- Sequential Interaction History: chronological sequence of user interactions, vital for learning user behavior patterns over time.

### 3.2. Contextual Dual Bias Reduction Recommendation System

In this section, we introduce the Contextual Dual Bias Reduction Recommendation System (C-DBRRS) algorithm, which is an advanced LSTM-based algorithm tailored for news recommendation. C-DBRRS is designed to balance content relevance and fairness by mitigating item and exposure biases while adapting to dynamic user interactions and news features. The notations used in the equations of this section are in Table 1 and the algorithm is given in Algorithm 1.

**Table 1.** Notation table for C-DBRRS.

Symbol	Description
$x = (x_1, x_2, \dots, x_t)$	Sequence of input data representing user interactions and news features
$t$ ine $i_t$	Time step in the input sequence Input gate at time step $t$
$f_t$ ine $\tilde{c}_t$	Forget gate at time step $t$ Candidate values for cell state update at time step $t$
$c_t$	Cell state at time step $t$
$o_t$ ine $h_t$	Output gate at time step $t$ Hidden state at time step $t$
$W_{xi}, W_{xf}, W_{xc}, W_{xo}$	Weight matrices connecting input to gates
$W_{hi}, W_{hf}, W_{hc}, W_{ho}$	Weight matrices connecting hidden state to gates
$b_i, b_f, b_c, b_o$	Bias terms for gates
$\alpha, \beta, \gamma$	Hyperparameters for balancing loss terms
$\lambda$	Hyperparameter for tuning fairness in recommendations
$L_{acc}$	Loss term for accuracy (mean squared error)
$L_{item}, L_{exp}$	Loss terms for item bias and exposure bias

The C-DBRRS employs a Long Short-Term Memory (LSTM) network to process sequences of input data  $x = (x_1, x_2, \dots, x_t)$ , integrating user interactions with news content features. The LSTM updates its internal state at each time step  $t$  through the following mechanisms:

- Input Gate controls how much new information flows into the cell state:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (1)$$

- Forget Gate determines the information to be removed from the cell state:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (2)$$

- Cell State Update generates new candidate values for updating the cell state:

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (4)$$

- Output Gate outputs the next hidden state reflecting the processed information:

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

To manage the trade-off between relevance and fairness in recommendations, the system employs a hyperparameter  $\lambda$  (described below). Relevance is assessed using metrics like Precision, Recall, and NDCG, whereas fairness is evaluated through the Gini coefficient. The optimization objective is formulated as:

$$\text{Minimize } L = \alpha L_{acc} + \lambda(\beta L_{item} + \gamma L_{exp}) \quad (7)$$

In this optimization objective:

- $L_{acc}$  (Accuracy Loss) is typically the mean squared error (MSE) between the predicted and actual user interactions. It measures how accurately the system predicts user preferences based on their interaction history and content features.
- $L_{item}$  (Item Bias Loss) aims to reduce the bias towards frequently recommended items. It is computed by measuring the deviation of the item distribution in the recommendations from a desired distribution, such as a uniform distribution.
- $L_{exp}$  (Exposure Bias Loss) is designed to ensure that all items receive a fair amount of exposure in the recommendations. This is measured as the variance in the number of times different items are recommended, penalizing the model when certain items are consistently under-represented.
- The hyperparameters  $\alpha, \beta, \gamma$  are used to balance these different aspects of the loss function. They are typically determined through experimentation and tuning, based on the specific characteristics of the data.

---

**Algorithm 1** Training Procedure for the Contextual Dual Bias Reduction Recommendation System (C-DBRRS)

---

```

1: Initialize LSTM model parameters and hyperparameters  $\alpha, \beta, \gamma, \lambda$ .
2: for each epoch do
3:   for each user do
4:     Input: Sequence of user interaction history and news content features.
5:     Process: Forward pass through LSTM.
6:     Compute Loss: Calculate  $L_{acc}, L_{item}, L_{exp}$ .
7:     Optimize: Update model parameters using Stochastic Gradient Descent (SGD).
8:   end for
9: end for

```

---

## 4. Experimental Setup

### 4.1. Baseline Methods

Our evaluation of the proposed C-DBRRS includes comparisons with a range of established recommendation methods, each offering unique strengths:

- Popularity-based Recommendation (POP): this method ranks news articles based on their overall popularity, measured by the total number of user clicks.
- Content-based Recommendation (CB): this method suggests articles to users by aligning the content of articles with their past preferences.
- Collaborative Filtering (CF): this method utilizes user behavior patterns, recommending items favored by similar users.
- Matrix Factorization (MF) [3]: this method decomposes the user-item interaction matrix into lower-dimensional latent factors for inferring user interests.
- Neural Collaborative Filtering (NCF) [33]: this method combines neural network architectures with collaborative filtering to enhance recommendation accuracy.
- BERT4Rec [34]: this model employs the Bidirectional Encoder Representations from Transformers (BERT) architecture, specifically designed for sequential recommendation. It captures complex item interaction patterns and user preferences from sequential data. We used the BERT-base-uncased model.

### 4.2. Evaluation Metrics

To assess the performance of our model against the baselines, we employed several key metrics:

- Precision@K measures the proportion of relevant articles in the top-K recommendations, reflecting accuracy.
- Recall@K indicates the fraction of relevant articles captured in the top-K recommendations, highlighting the model's retrieval ability.
- Normalized Discounted Cumulative Gain (NDCG)@K assesses ranking quality, prioritizing the placement of relevant articles higher in the recommendation list.

- Gini Index evaluates the fairness of recommendation distribution, with lower values indicating more equitable distribution across items.

We consider the value of top@k as 5 ( $k = 5$ ) following standard works in recommender systems theory [16].

#### 4.3. Settings and Hyperparameters

We temporally split the datasets into training, validation, and testing sets with a ratio of 80:10:10. The hyperparameters of the models were tuned on the validation set. All experiments were conducted on a machine with an Intel Xeon processor, 32GB RAM, and an Nvidia GeForce GTX 1080 Ti GPU. The set of hyperparameters is given in Table 2.

**Table 2.** Hyperparameters and ranges.

Hyperparameter	Description	Values/Range
Learning Rate	Step size for updating weights	0.001, 0.01, 0.1
Batch Size	Samples processed before update	32, 64, 128
Num. of Epochs	Passes through entire dataset	10, 20, 30
LSTM Units	Number of LSTM units	50, 100, 150
Dropout Rate	Fraction of units to drop	0.2, 0.5
$\alpha$	Weight for accuracy loss	0.3, 0.5, 0.7
$\beta$	Weight for item bias loss	0.1, 0.3, 0.5
$\gamma$	Weight for exposure bias loss	0.2, 0.4, 0.6, default is 0.5

In our C-DBRRS, various hyperparameters are carefully tuned for optimal performance, as shown in Table 2. Hyperparameters  $\alpha, \beta, \gamma$  are crucial for weighting different components of the loss function, controlling how the model balances prediction accuracy, item bias, and exposure bias, respectively. In particular,  $\lambda$  serves as a key parameter for overall balancing between relevance and fairness in the recommendation output, in line with the system's optimization objective.

## 5. Results

### 5.1. Overall Results

Table 3 illustrates the performance of various recommendation methods on the MIND-small and Outbrain datasets, respectively, including our proposed C-DBRRS and the state-of-the-art models. The performances reported are averaged over five runs to ensure statistical reliability, with the standard deviation included to indicate performance variability.

The Table 3 compares the performance of different recommendation methods on two datasets, MIND-small and Outbrain, using metrics such as Precision@5, Recall@5, NDCG@5, and the Gini Index. In both datasets, the C-DBRRS method demonstrates higher performance, achieving the highest scores in Precision@5, Recall@5, and NDCG@5, alongside the lowest Gini Index. In this study, we value a lower GINI index as it indicates a desirable level of diversity in recommendations. However, we aim to achieve this without resorting to entirely random recommendations. Our goal is to strike a balance between diversity and relevance, ensuring that the recommendations are diverse yet still meaningful and aligned with user interests.

Other methods like POP, CB, CF, MF, NCF, and BERT4Rec show lower performance compared to C-DBRRS. The improvement in performance metrics from simpler methods like POP to more advanced ones like C-DBRRS highlights the efficacy of sophisticated recommendation systems. Particularly, the lower Gini Index in methods like C-DBRRS highlights their capability in ensuring a more equitable distribution of recommendations (not too highly diverse).

**Table 3.** Performance of different methods averaged over 5 runs. Bold indicates the best score. Higher accuracy scores (precision, recall, NDCG) and a lower GINI index are indicative of higher performance.

Dataset	Method	Precision@5	Recall@5	NDCG@5	Gini Index
MIND-small	POP	0.35 ± 0.05	0.25 ± 0.04	0.30 ± 0.05	0.45 ± 0.06
	CB	0.40 ± 0.05	0.30 ± 0.04	0.35 ± 0.05	0.40 ± 0.06
	CF	0.45 ± 0.05	0.35 ± 0.04	0.40 ± 0.05	0.35 ± 0.06
	MF	0.50 ± 0.05	0.40 ± 0.04	0.45 ± 0.05	0.30 ± 0.06
	NCF	0.55 ± 0.05	0.45 ± 0.04	0.50 ± 0.05	0.25 ± 0.06
	BERT4Rec	0.57 ± 0.05	0.47 ± 0.04	0.52 ± 0.05	0.22 ± 0.06
	C-DBRRS	<b>0.65 ± 0.04</b>	<b>0.55 ± 0.04</b>	<b>0.60 ± 0.04</b>	<b>0.18 ± 0.05</b>
Outbrain	POP	0.32 ± 0.05	0.24 ± 0.04	0.29 ± 0.05	0.43 ± 0.06
	CB	0.38 ± 0.05	0.29 ± 0.04	0.33 ± 0.05	0.39 ± 0.06
	CF	0.42 ± 0.05	0.33 ± 0.04	0.38 ± 0.05	0.34 ± 0.06
	MF	0.48 ± 0.05	0.37 ± 0.04	0.42 ± 0.05	0.30 ± 0.06
	NCF	0.52 ± 0.05	0.42 ± 0.04	0.47 ± 0.05	0.26 ± 0.06
	BERT4Rec	0.54 ± 0.05	0.44 ± 0.04	0.49 ± 0.05	0.23 ± 0.06
	C-DBRRS	<b>0.62 ± 0.04</b>	<b>0.52 ± 0.04</b>	<b>0.57 ± 0.04</b>	<b>0.19 ± 0.05</b>

Overall, these results suggest a clear advantage of advanced methods like C-DBRRS in enhancing both the accuracy and fairness of recommendations in News Recommendation Systems. Because of the comparable patterns found in both datasets and the superior performance of our model on the MIND dataset, we will present the results of our subsequent experiments using the MIND dataset.

### 5.2. Recommendation Distribution across Different News Categories

To assess the fairness of the C-DBRRS's recommendation distribution across different news categories on the MIND dataset, we employed the Gini coefficient as a measure of inequality—a Gini coefficient of 0 expresses perfect equality, and a Gini coefficient of 1 implies maximal inequality among values. We first calculated the Gini coefficients for each category in the baseline model, where no fairness constraints were applied. Subsequently, we integrated the fairness constraints into the C-DBRRS model and recalculated the Gini coefficients.

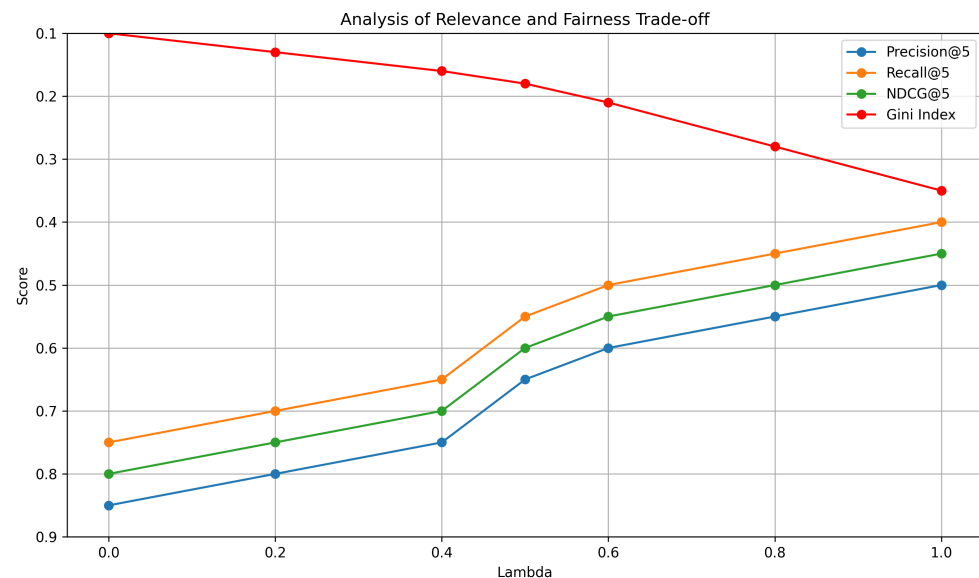
As evident from Table 4, all categories exhibit a significant reduction in the Gini coefficient, indicating a more equitable distribution of news recommendations. Specifically, the 'Environment' category showed the most considerable improvement, with a Gini coefficient reduction from 0.65 to 0.30, followed by the 'Arts' category, which reduced from 0.60 to 0.28. The 'Politics' and 'Sports' categories also observed notable improvements. The reduction in the Gini coefficients across all news categories indicates that the C-DBRRS model successfully addresses the challenges of item bias and exposure bias, leading to a more equitable and diverse set of recommendations.

**Table 4.** Gini coefficients for news categories.

Category	Gini (before)	Gini (after)
Politics	0.55	0.25
Sports	0.45	0.23
Arts	0.60	0.28
Science	0.50	0.26
Environment	0.65	0.30

### 5.3. Analysis of Relevance and Fairness Trade-Off

The results presented in Figure 2 illustrate the trade-off between relevance and fairness for the C-DBRRS model. Specifically, as the value of  $\lambda$  increases from 0 to 1, the model places more emphasis on fairness, leading to a decrease in relevance as measured by Precision@5, Recall@5, and NDCG@5. For example, when  $\lambda$  is 0 (meaning the model only considers relevance), the Precision@5, Recall@5, and NDCG@5 are highest. However, when  $\lambda$  is increased to 1 (meaning the model only considers fairness), these values decrease. This indicates that there is a trade-off between achieving high relevance and high fairness, as improving fairness leads to a decrease in relevance.



**Figure 2.** Analysis of Relevance and Fairness Trade-off in C-DBRRS Model. The parameter  $\lambda$  (Lambda) is used to balance the trade-off between relevance (accuracy) and fairness (diversity). The table includes relevance measures for Precision, Recall, Normalized Discounted Cumulative Gain (NDCG), and fairness measures through Gini Index at a cutoff of 5.

The Gini Index, a measure of inequality, improves (decreases) as  $\lambda$  increases, indicating that the recommendations are becoming more fair. For example, when  $\lambda$  is 0, the Gini Index is 0.10, whereas it decreases to 0.18 when  $\lambda$  is 0.5 and increases further when  $\lambda$  is 1. This shows that the C-DBRRS model is effective at improving the fairness of the recommendations while maintaining a reasonable level of relevance. Researchers and practitioners using this model will need to carefully select the value of  $\lambda$  to balance the trade-off between relevance and fairness based on their specific application and requirements.

## 6. Discussion

### 6.1. Practical and Theoretical Impact

In this paper, we presented a novel approach optimized with a multi-objective function to balance the accuracy and diversity aspects of fairness. The C-DBRRS model, which forms the foundation of our approach, leverages the capability of LSTM networks to capture temporal patterns in users' interactions with items, thereby providing more accurate and personalized recommendations. By predicting the next items a user is likely to interact with and ranking them based on their predicted interaction probabilities, our model can recommend the top-ranked items to the user.

Our multi-objective optimization goal is a key contribution to this work. This approach integrates multiple bias-related objectives, namely item bias and exposure bias, in addition to accuracy. By formulating this as a multi-objective optimization problem and minimizing the Gini coefficients of the distribution of recommended items and exposure, we encourage a more equal and unbiased distribution of recommendations and exposure.

### 6.2. Limitations

Our approach has some limitations. First, the computational complexity of the model may be high due to the use of LSTM networks and the need to compute Gini coefficients for user–item interactions. This may limit the scalability of our approach to very large datasets. Second, the performance of our approach may be sensitive to the choice of hyperparameters, and determining the optimal values may require extensive grid search. Furthermore, our approach assumes that the temporal dynamics of user–item interactions are important for making recommendations. This assumption may not hold for all types of items or users. Moreover, whereas our focus is on finding biases, it is essential to acknowledge that fairness in recommendation systems is beyond diversity. Future studies should delve into nuanced facets of fairness, including equitable item representation and user recommendations.

Future work could explore alternative optimization algorithms, different types of recurrent neural networks, and the applicability of our approach to other types of recommendation problems.

### 6.3. Recommender Systems Fairness in the Era of Large Language Models

The emergence of Large Language Models (LLMs) has brought a new dimension to the fairness of recommender systems. A recent survey [35] highlights the importance of integrating fairness-aware strategies in these systems, focusing on countering potential biases and promoting equality. LLMs, with their advanced deep learning architectures and extensive training on diverse datasets, excel in identifying and predicting a wide array of user preferences and behaviors. This capability is important in counteracting the ‘echo chamber’ effect prevalent in recommender systems by offering a varied range of content. Such diversity in content exposes users to a broader spectrum of topics and perspectives, thereby promoting a more balanced consumption of information, whereas personalization is essential for user satisfaction, LLMs in recommender systems can aptly balance it with the need for recommendation diversity.

However, the utilization of LLMs in recommender systems is accompanied by challenges. Ensuring user privacy, managing biased training data, and maintaining transparency in the recommendation processes are critical considerations. Additionally, there is an ethical imperative to avoid manipulative practices in these systems. LLMs, with their extensive knowledge and understanding, are adept at delivering recommendations that are not only precise but also encompass a wide spectrum of content.

The integration of LLMs into recommender systems necessitates the development of advanced strategies to accommodate diverse user preferences and diminish biases [36]. The need for fairness testing in these systems to ensure equitable recommendations is emphasized in related research [37]. Moreover, various evaluation approaches and assurance strategies are proposed to uphold fairness in recommender systems [38]. The significance of privacy-preserving mechanisms in LLM-based recommender systems is also required, underlining the interplay between fairness and privacy [39]. All these approaches should align with the broader goal of fairness in recommender systems, ensuring that users are presented with a balanced mix of familiar and novel content, thus avoiding the creation of echo chambers.

## 7. Conclusions

In this paper, we presented the C-DBRRS that formulates the optimization goal as a multi-objective problem to encourage a more equal and unbiased distribution of recommendations and exposure. Our experiments on two real-world datasets demonstrated that our approach outperforms state-of-the-art methods in terms of accuracy, fairness, and balance. Additionally, recommendation distribution across different news categories confirmed the effectiveness of our approach in addressing item bias and exposure bias, leading to a more equitable and diverse set of recommendations. This is a crucial step towards developing more fair and responsible recommendation systems. Future work could explore alternative optimization algorithms, different types of recurrent neural networks, LLMs and the

applicability of our approach to other types of recommendation problems. Additionally, it would be interesting to investigate the impact of our approach on user satisfaction and engagement in a real-world setting.

**Funding:** This research has received no funding.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The data are publicly available at MIND-small [2] and Outbrain Click Prediction at [31].

**Conflicts of Interest:** The author declares no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

NRS	News Recommender Systems
LSTM	Long Short-Term Memory
C-DBRRS	Contextual Dual Bias Reduction Recommendation System
BERT	Bidirectional Encoder Representations from Transformers
POP	Popularity-Based Recommendation
CB	Content-Based Recommendation
CF	Collaborative Filtering
MF	Matrix Factorization
NCF	Neural Collaborative Filtering
NDCG	Normalized Discounted Cumulative Gain

## References

1. Raza, S.; Ding, C. News recommender system: A review of recent progress, challenges, and opportunities. *Artif. Intell. Rev.* **2022**, *55*, 749–800. [\[CrossRef\]](#)
2. Wu, F.; Qiao, Y.; Chen, J.H.; Wu, C.; Qi, T.; Lian, J.; Liu, D.; Xie, X.; Gao, J.; Wu, W.; et al. Mind: A large-scale dataset for news recommendation. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 3597–3606.
3. Raza, S.; Ding, C. News recommender system considering temporal dynamics and news taxonomy. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 920–929.
4. Wang, Y.; Ma, W.; Zhang, M.; Liu, Y.; Ma, S. A survey on the fairness of recommender systems. *ACM Trans. Inf. Syst.* **2023**, *41*, 1–43. [\[CrossRef\]](#)
5. Raza, S.; Ding, C. Deep neural network to tradeoff between accuracy and diversity in a news recommender system. In Proceedings of the 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 15–18 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 5246–5256.
6. Raza, S.; Garg, M.; Reji, D.J.; Bashir, S.R.; Ding, C. Nbias: A natural language processing framework for BIAS identification in text. *Expert Syst. Appl.* **2024**, *237*, 121542. [\[CrossRef\]](#)
7. Zheng, G.; Zhang, F.; Zheng, Z.; Xiang, Y.; Yuan, N.J.; Xie, X.; Li, Z. DRN: A deep reinforcement learning framework for news recommendation. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; pp. 167–176.
8. Wang, X.; Wang, W.H. Providing Item-side Individual Fairness for Deep Recommender Systems. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, 21–24 June 2022; pp. 117–127.
9. Wu, H.; Mitra, B.; Ma, C.; Diaz, F.; Liu, X. Joint multisided exposure fairness for recommendation. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022; pp. 703–714.
10. Helberger, N. On the democratic role of news recommenders. In *Algorithms, Automation, and News*; Routledge: London, UK, 2021; pp. 14–33.
11. Dwork, C.; Hardt, M.; Pitassi, T.; Reingold, O.; Zemel, R. Fairness through awareness. In Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, Cambridge MA, USA, 8–10 January 2012; pp. 214–226.
12. Dolata, M.; Feuerriegel, S.; Schwabe, G. A sociotechnical view of algorithmic fairness. *Inf. Syst. J.* **2022**, *32*, 754–818. [\[CrossRef\]](#)
13. Wu, H.; Ma, C.; Mitra, B.; Diaz, F.; Liu, X. A multi-objective optimization framework for multi-stakeholder fairness-aware recommendation. *ACM Trans. Inf. Syst.* **2022**, *41*, 1–29. [\[CrossRef\]](#)
14. Li, L.; Wang, D.D.; Zhu, S.Z.; Li, T. Personalized news recommendation: A review and an experimental investigation. *J. Comput. Sci. Technol.* **2011**, *26*, 754–766. [\[CrossRef\]](#)
15. Devooght, R.; Bersini, H. Collaborative filtering with recurrent neural networks. *arXiv* **2016**, arXiv:1608.07400.

16. Herlocker, J.L.; Konstan, J.A.; Terveen, L.G.; Riedl, J.T. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst. (TOIS)* **2004**, *22*, 5–53. [CrossRef]
17. Adomavicius, G.; Tuzhilin, A. Context-aware recommender systems. In *Recommender Systems Handbook*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 217–253.
18. Burke, R. Hybrid recommender systems: Survey and experiments. *User Model. User-Adapt. Interact.* **2002**, *12*, 331–370. [CrossRef]
19. Fu, Z.; Xian, Y.; Gao, R.; Zhao, J.; Huang, Q.; Ge, Y.; Xu, S.; Geng, S.; Shah, C.; Zhang, Y.; et al. Fairness-aware explainable recommendation over knowledge graphs. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, 25–30 July 2020; pp. 69–78.
20. Beutel, A.; Chen, J.; Doshi, T.; Qian, H.; Wei, L.; Wu, Y.; Heldt, L.; Zhao, Z.; Hong, L.; Chi, E.H.; et al. Fairness in recommendation ranking through pairwise comparisons. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2212–2220.
21. Yao, S.; Huang, B. Beyond parity: Fairness objectives for collaborative filtering. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
22. Zehlike, M.; Bonchi, F.; Castillo, C.; Hajian, S.; Megahed, M.; Baeza-Yates, R. FA\*IR: A fair top-k ranking algorithm. In Proceedings of the 2017 ACM Conference on Information and Knowledge Management, Singapore, 6–10 November 2017; pp. 1569–1578.
23. Farnadi, G.; Kouki, P.; Thompson, S.K.; Srinivasan, S.; Getoor, L. A fairness-aware hybrid recommender system. *arXiv* **2018**, arXiv:1809.09030.
24. Sonboli, N.; Smith, J.J.; Cabral Berenfus, F.; Burke, R.; Fiesler, C. Fairness and transparency in recommendation: The users perspective. In Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, Utrecht, The Netherlands, 21–25 June 2021; pp. 274–279.
25. Mehrotra, R.; McInerney, J.; Bouchard, H.; Lalmas, M.; Diaz, F. Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; pp. 2243–2251.
26. Cui, L.; Ou, P.; Fu, X.; Wen, Z.; Lu, N. A Novel Multi-objective Evolutionary Algorithm for Recommendation Systems. *J. Parallel Distrib. Comput.* **2017**, *103*, 53–63. [CrossRef]
27. Li, H.; Zhong, Z.; Shi, J.; Li, H.; Zhang, Y. Multi-objective optimization-based recommendation for massive online learning resources. *IEEE Sens. J.* **2021**, *21*, 25274–25281. [CrossRef]
28. Xu, C. A big-data oriented recommendation method based on multi-objective optimization. *Knowl.-Based Syst.* **2019**, *177*, 11–21. [CrossRef]
29. Raza, S.; Ding, C. A Regularized Model to Trade-off Between Accuracy and Diversity in a News Recommender System. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; pp. 551–560. [CrossRef]
30. Wu, Y.; Xie, R.; Zhu, Y.; Zhuang, F.; Xiang, A.; Zhang, X.; Lin, L.; He, Q. Selective fairness in recommendation via prompts. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022; pp. 2657–2662.
31. Outbrain Click Prediction. 2023. Available online: <https://www.kaggle.com/competitions/outbrain-click-prediction/data> (accessed on 4 September 2023).
32. Hutto, C.; Gilbert, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media, Ann Arbor, MI, USA, 1–4 June 2014; Volume 8, pp. 216–225.
33. He, X.; Liao, L.; Zhang, H.; Nie, L.; Hu, X.; Chua, T.S. Neural collaborative filtering. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 May 2017; pp. 173–182.
34. Sun, F.; Liu, J.; Wu, J.; Pei, C.; Lin, X.; Ou, W.; Jiang, P. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 1441–1450.
35. Pan, S.; Erfahrungen, P.K. A survey on fairness-aware recommender systems. *Inf. Fusion* **2023**. [CrossRef]
36. Fan, W.; Zhao, Z.; Li, J.; Liu, Y.; Mei, X.; Wang, Y.; Wen, Z.; Wang, F.; Zhao, X.; Tang, J.; et al. Recommender Systems in the Era of Large Language Models (LLMs). *arXiv* **2023**, arXiv:2307.02046.
37. Guo, H. Fairness Testing for Recommender Systems. In Proceedings of the International Symposium on Software Testing and Analysis, Seattle, WA, USA, 17–21 July 2023. [CrossRef]
38. Wu, Y.; Cao, J.; Xu, G. Fairness in Recommender Systems: Evaluation Approaches and Assurance Strategies. *ACM Trans. Knowl. Discov. Data* **2023**, *18*, 1–37. [CrossRef]
39. Carranza, A.G.; Farahani, R.; Ponomareva, N.; Kurakin, A.; Jagielski, M.; Nasr, M. Privacy-Preserving Recommender Systems with Synthetic Query Generation using Differentially Private Large Language Models. *arXiv* **2023**, arXiv:2305.05973.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.