Reading Between the Lines: A Study of Thematic Bias in Book Recommender Systems

Nityaa Kalra
nityaak5@gmail.com
Radboud University
Nijmegen, The Netherlands

Savvina Daniil
s.daniil@cwi.nl
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands

Abstract

Recommender systems help users discover new content, but can also reinforce existing biases, leading to unfair exposure and reduced diversity. This paper introduces and investigates thematic bias in book recommendations, defined as a disproportionate favoring or neglect of certain book themes. We adopt a multi-stage bias evaluation framework using the Book-Crossing dataset to evaluate thematic bias in recommendations and its impact on different user groups.

Our findings show that thematic bias originates from content imbalances and is amplified by user engagement patterns. By segmenting users based on their thematic preferences, we find that users with niche and long-tail interests receive less personalized recommendations, whereas users with diverse interests receive more consistent recommendations. These findings suggest that recommender systems should be carefully designed to accommodate a broader range of user interests. By contributing to the broader goal of responsible AI, this work also lays the groundwork for extending thematic bias analysis to other domains.

Keywords

Recommender Systems, Bias, Book Recommendation, Responsible AI

1 Introduction

Recommender systems have become central to how we engage with the vast digital ecosystems of today. Whether it is finding the next movie to watch, discovering new music, or selecting a book to read, these systems not only help us navigate overwhelming choices, but actively shape our consumption habits [11]. In the context of books, recommendation algorithms influence the types of narratives and ideas to which users are exposed. This exposure can lead to *psychological transportation*, a process known to influence individuals' attitudes, beliefs, and ultimately their interpretation of the world [12].

Sociologist Pierre Bourdieu's work [5] shows that cultural consumption is never neutral. Our preferences are shaped by underlying social structures, and in turn, they shape how we are perceived and judged by others. Further, Lawrence [17] argues that reading preferences, if left unchecked, can become barriers to a fair and democratic society. From this perspective, recommending books is not merely a technical task; it is an ethical responsibility of librarians to actively promote diverse content rather than simply catering to individual desires. This argument becomes even more important in the context of book recommender systems, which mediate our reading choices without human intervention.

However, these systems are not neutral. They often favor popular titles [20] or authors from dominant demographic groups [9], reinforcing existing cultural hierarchies. Although popularity and demographic bias have received increasing attention, this work turns to a less studied but equally important concern: **thematic bias**. Thematic bias refers to a content-level skew in which certain book themes are disproportionately favored, while others are overlooked. Unlike biases tied to user behavior or author identity, thematic bias emerges from the semantic and ideological framing of the content itself, reflecting what kinds of ideas and narratives are being amplified or marginalized. When algorithms amplify specific themes due to popularity dynamics or engagement metrics, they participate in shaping cultural discourse. This is visible in communities like BookTok and Bookstagram, where specific narratives often achieve viral dominance due to platform algorithms and influencer trends [3, 22]. Therefore, studying thematic bias is not simply about improving the accuracy of personalization. Instead, it is about recognizing how these systems might reproduce or even intensify cultural inequalities.

When discussing thematic bias, it is important to clarify what we mean by themes. Traditionally, they refer to the underlying ideas or messages conveyed by a work [23]. For example, a classic novel like *Pride and Prejudice* addresses multiple themes such as love, social class, and individual growth. However, in this work, we define themes as clusters of semantically related terms derived from a large collection of book descriptions using topic modeling techniques. While this approach may not capture the full depth or complexity of a literary theme, it provides a practical and scalable means to represent thematic content across thousands of titles. It is worth noting that thematic bias does not arise in isolation. It is often shaped by structural patterns within both the data and the algorithms. For instance, Daniil et al. [9] show how demographic bias stems from popularity bias. Similarly, thematic bias may result from the dominance of certain user preferences or widely consumed content.

In this paper, we investigate whether thematic biases emerge in book recommender systems and how user preferences interact with these biases using a multi-stage bias evaluation framework. We adopt a segmentation strategy inspired by intersectionality, a concept originally developed to understand how overlapping social identities create unique experiences of privilege or disadvantage [6, 8, 15]. By extending this idea to user modeling, we analyze how multiple dimensions of reading behavior, such as thematic diversity and propensity for popular items, combine to shape users' recommendation experiences. To guide this investigation, we pose the following research questions:

(1) To what extent does thematic bias exist in the input data used by book recommender systems?

(2) How do different recommendation algorithms amplify or mitigate thematic biases in their outputs?

(3) How do thematic biases affect different user groups, particularly in relation to their thematic diversity and propensity for popular items?

Our contributions include enriching the Book-Crossing dataset [30] with thematic metadata via topic modeling, proposing a theme-based bias analysis pipeline, and applying an intersectional segmentation of users based on their reading behaviors. To the best of our knowledge, this is the first work to analyze thematic bias using topic modeling in book recommender systems with an intersectional user-group fairness lens. To support transparency and reproducibility of our work, all code, data, and analysis notebooks are publicly available on GitHub.[1]

2 Related Work

2.1 Bias in Recommender Systems

While recommender systems enhance user experience by personalizing content, they are also susceptible to various biases that can lead to skewed, unfair, or homogenized recommendations. A widely documented phenomenon is **popularity bias**, where algorithms tend to favor already popular items. Abdollahpouri et al. [1] formalized this tendency within collaborative filtering (CF) algorithms. Further, Naghiaei et al. [20] demonstrated that these algorithms consistently under-served users with diverse or niche tastes, while performing well for bestseller-focused readers.

Recommendation outcomes are also shaped by deeper structural issues such as data-level and algorithmic biases. **Data-level bias** [7] arises from the nature of the user-item interaction logs that recommender systems rely on. These logs reflect only a small and non-random subset of user preferences. As a result, the data tends to overrepresent items that were already popular or frequently shown, regardless of their actual relevance. The issues in the input data are further compounded by **algorithmic bias** [14], particularly through feedback loops [19] which can reinforce existing interaction patterns, leading to an echo chamber effect [21].

Several studies have documented the demographic implications of bias in recommender systems. For instance, collaborative filtering models have been shown to reflect, and in some cases amplify gender distribution found in user profiles, thereby reinforcing existing disparities in representation [10]. Other work has shown that such models may also favor books by U.S. authors over those from other regions [9]. While Ziegler et al. [30] used the Book-Crossing dataset to examine topical diversity, their focus was on enhancing personalization rather than investigating thematic bias in recommendation systems. Initial steps to address these biases have been taken in adjacent domains such as music and movies, where researchers have proposed calibrated recommendation strategies to improve genre balance [25]. This work builds on that gap by examining whether recommendation algorithms favor certain literary themes over others, and how such biases affect different user groups.

2.2 Topic Modeling

Understanding thematic content of text is essential in many Natural Language Processing (NLP) and Information Retrieval tasks. One of the most widely used approaches is topic modeling [26], which aims to uncover the thematic structure in a collection of documents.

Traditional models such as Latent Dirichlet Allocation (LDA) [4] and Non-negative Matrix Factorization (NMF) [18] have been widely used to extract themes from textual data. These models represent each document as a mixture of topics, making them suitable for content-rich domains like books, where summaries and descriptions often convey multiple narrative dimensions. More recent advances in NLP have introduced embedding-based topic models that leverage pre-trained language models. For example, Top2Vec [2] and BERTopic [13] utilize word or sentence embeddings to represent documents in a semantic vector space, which are then clustered to identify coherent topics.

In the context of recommender systems, topic models have been used to derive content features that improve recommendation quality, especially when user-item interaction data is sparse [27]. However, the application of topic modeling in book recommendation research remains underdeveloped despite the richness of book metadata such as summaries, blurbs, and user-generated reviews. In this study, we utilized **BERTopic** [13] to extract fine-grained themes from book descriptions. These extracted themes formed the foundation of our analysis of **thematic bias** in recommendation outcomes.

3 Research Design

3.1 Dataset Preparation

For this study, we utilized the Book-Crossing dataset [30] which originates from a four-week crawl of the *BookCrossing*[2] website. The raw dataset consists of over 1 million book ratings. However, it is highly sparse and noisy. For example, many books appear under multiple ISBNs due to edition changes, author names follow inconsistent formats, and most books have received very few ratings.

To reduce noise and improve robustness, we applied preprocessing steps similar to Naghiaei et al. [20]. We removed users with fewer than 5 or more than 200 ratings, books with fewer than 5 ratings, and all implicit ratings (i.e., rating = 0). We also merged duplicates using a combination of title and author fields.

To support theme-based bias analysis, we enriched books with textual descriptions and genre labels from external sources. While genre labels are available in some cases, we chose not to rely on them exclusively due to their broadness and inconsistency. At times, they fail to capture the detailed thematic information needed for effective bias analysis. For instance, many books are simply labeled *Fiction* despite being clearly *Psychological Thrillers* or *Crime Thrillers*. Moreover, genre metadata is frequently missing for less popular titles, thereby limiting its coverage. To address these issues, we collected detailed book descriptions and later applied topic modeling to extract consistent thematic representations.

As shown in Figure 1, we first queried the Google Books API[3] using ISBN or title-author combination, and retrieved metadata for

approximately 2,000 books. Since Goodreads stopped issuing new API keys and discontinued public developer access in December 2020, we used a scraped version of the Goodreads dataset[4] for the remaining books. We also performed additional scraping through Amazon Books[5], Wikipedia, and DBpedia. Next, we cleaned book descriptions to remove outliers based on word count. Finally, books which could not be enriched via any of these methods were removed. Table 1 provides detailed statistics of the dataset, both before (without implicit ratings) and after pre-processing.
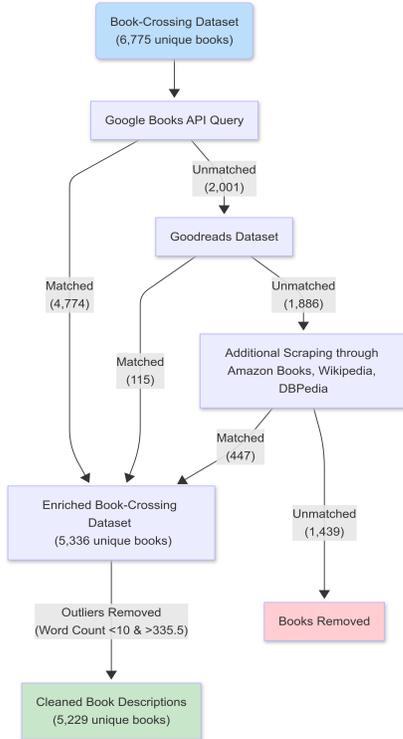

Figure 1: Metadata enrichment pipeline

Table 1: Dataset statistics before and after data enrichment

Metric	Raw Dataset	Post-Enrichment
Users	68,091	5,424
Books	149,836	5,229 (with Descriptions)
		4,984 (with Categories)
Interactions	383,842	70,622
Sparsity	99.90%	99.75%
Average Rating	7.63	7.84
Median Rating	8.00	8.00
Ratings/User	5.64	13.02
Ratings/Book	2.56	13.51

3.2 Thematic Modeling

To extract themes from the book descriptions, we used BERTopic [13], a state-of-the-art topic modeling method that uses contextual embeddings from transformer models.

We modeled 5,229 cleaned book descriptions with a minimum cluster size of 10 and cosine similarity threshold of 0.8 for topic merging. An initial run yielded 65 topics. After merging and pruning low-quality clusters, we retained 25 final topics and referred to them as *themes*. Next, we manually reviewed them based on the top 10 keywords, sample book descriptions, and the label created using BERTopic's inbuilt LLM functionality. Finally, labels were assigned to each of these themes to reflect the core narrative or subject matter in the book descriptions.

To validate the quality of discovered themes, we combined internal and external checks. Internally, we evaluated them through manual inspection, stability across multiple runs, and topic coherence scores (average = 0.52). Externally, we compared them against genre metadata and reading trends from the early 2000s. Common themes such as *Family Dynamics and Relationships*, *Crime Thriller* and *Detective Drama*, aligned well with bestseller patterns from *The New York Times* [28] and *Publishers Weekly* [29], while introspective themes such as *Spiritual and Self-Help* matched post-9/11 literary commentary [24]. This dual validation supported relevance of our themes for bias analysis.

3.3 Bias Analysis Framework

To investigate bias in book recommendation systems, we implemented a structured framework combining collaborative filtering (CF) algorithms with multi-stage bias evaluation.

We selected both memory-based and model-based approaches alongside two simple baselines. CF algorithms were chosen for their effectiveness in modeling user-item interactions and their known susceptibility to popularity bias [1, 9, 20]. Implementation was done using the Cornac library[6], with an 80-10-10 train-validation-test split and hyperparameter tuning based on NDCG@10. The algorithm choices were aligned with prior work on bias in book recommender systems [9, 20] to ensure comparability of our results. Table 2 lists all recommender algorithms used in our experiments.

Table 2: Recommender algorithms used in the study

Algorithm	Acronym	Approach
Random	Random	Baseline
Most Popular	MostPop	Baseline
Matrix Factorization	MF	Matrix Factorization
Weighted Matrix Factorization	WMF	Matrix Factorization
Non-negative Matrix Factorization	NMF	Matrix Factorization
Probabilistic Matrix Factorization	PMF	Matrix Factorization
Hierarchical Poisson Factorization	HPF	Matrix Factorization
Bayesian Personalized Ranking	BPR	Pairwise Ranking
User-based K-Nearest Neighbor	User-kNN	Nearest Neighbors
Neural Matrix Factorization	NeuMF	Neural Networks
Variational Autoencoders	VAE-CF	Neural Networks

We computed the top-10 book recommendations for each user using the selected algorithms. These recommendations were then

analyzed to assess three types of bias: **data bias**, **recommendation bias**, and **user group bias**.

To analyze user group bias, we first segmented users based on their propensity for popular books. Following the definition of popularity from [20], we identified three groups: **Mainstream readers**, whose reading history consisted of more than 70% popular books, **Mixed readers**, with 30–70% of their books from the popular set, and **Long-Tail readers**, with fewer than 30% of their books classified as popular. Next, we segmented users by their historical thematic diversity and Gini ratio. Users below the 25th percentile in theme count and above the 75th percentile in Gini were labeled as **Specialist readers**, those above the 75th percentile in theme count and below the 25th percentile in Gini as **Generalist readers**, and the rest as **Moderate readers**. We then compared the top-10 recommendations of each group with their historical preferences.

4 Results and Discussion

4.1 Bias in Data

We extracted **25 distinct themes** using BERTopic from book descriptions. The thematic distribution was highly imbalanced: about 20% of themes accounted for over 52% of unique books, revealing a clear imbalance in theme representation as seen in Figure 2.

We defined a popular subset of books using the approach of Naghiaei et al. [20] and analyzed this subset to see which themes were overrepresented among popular books, both by comparing theme distribution and by computing the average popularity ratio of books within each theme. Several themes such as *Crime Thrillers and Detective Drama*, *Family Dynamics and Relationships*, and *Vampire/ Paranormal Fantasy* showed clear overrepresentation among popular books. In contrast, themes such as *Cozy Mysteries* and *Space Adventures* were among the most underrepresented themes in the popular subset, despite of their moderate presence in the overall dataset. To determine if these differences were statistically meaningful, we conducted a Chi-squared test of independence. The results showed statistically significant disparities ($p < 0.05$) in the distribution for 8 out of 25 themes. Furthermore, themes like *Wizarding World* and *Holiday Classics*, despite of their moderate representation in the training dataset, had exceptionally high average popularity ratios (Figure 3), indicating engagement-driven amplification bias. This suggests that themes with fewer books can still exert disproportionate influence in recommendations if they attract high user engagement.

Therefore, in response to RQ1, thematic bias is clearly present in the input dataset used by book recommender systems and it manifests as a multi-layered phenomenon. It stems from the initial imbalance in theme representation and is reinforced by user engagement patterns. Together, these biases create a compounding disadvantage for underrepresented themes and set the stage for potential downstream disparities in recommendation outputs.

4.2 Bias in Recommendations

To analyze recommendation bias, we first compared predictive accuracy of different algorithms using various metrics such as precision, recall, F1 score, and NDCG at k=10. WMF consistently outperformed all models across these metrics. VAECF and NeuMF also showed strong performance, particularly in balancing ranking and

relevance. BPR and PF offered competitive performance, especially in Recall and F1, outperforming traditional factorization models. In contrast, user-KNN, MF and PMF performed poorly. Complete results across several other metrics are available in our Github repository.

Next, we analyzed how these models distributed themes in their **top-10** recommendations. Thematic recommendation percentages showed that *Vampire/Paranormal Fantasy*, *Family Dynamics and Relationships*, and *American Literature Classics* dominated the output of most algorithms, mirroring their prevalence in the input data. In contrast, themes such as *Cozy Mysteries*, *International Literature*, and *War and Historical Drama* remained nearly absent in recommendations. We calculated the exposure ratio to quantify the difference between the prevalence of a theme in the dataset and the recommendations.

$$Exposure\ Ratio = \frac{\%\ Theme\ in\ Recommendation}{\%\ Theme\ in\ Training\ Data}$$

Wizarding World consistently exhibited a high Exposure Ratio among all algorithms. In fact, other themes that were overrepresented were those which had a high average popularity ratio in the first place (Figure 3). This amplification highlights how themes with high engagement can be disproportionately promoted. We also evaluated item coverage across all algorithms. As expected, MostPop had the lowest coverage (0.19%), while WMF stood out with the highest coverage (91.11%), demonstrating its ability to explore a large portion of the item catalog. PMF showed moderate coverage (42.47%), but other models like NeuMF, PF, BPR, and VAECF covered less than 15% of the catalog, indicating a tendency to recommend from a small item pool and thereby limiting user exposure to diverse items.

To answer RQ2, we observe that most algorithms tend to amplify existing popularity biases. Models with high accuracy (WMF, VAECF) often overrepresent themes that were already popular in the input data. However, niche or underrepresented themes remain largely excluded, especially by models with low coverage and popularity-driven focus (MostPop, BPR). Even algorithms that achieve broader coverage (WMF) still contribute to thematic skewness by disproportionately favoring high-engagement content. Therefore, while predictive accuracy is not inherently tied to fairness, most algorithms do little to diversify exposure or correct imbalances in theme representation.

4.3 Bias Across User Groups

To analyze bias across different user groups, we first examined how recommendations varied based on user's propensity for popular items. A consistent trend was observed across all algorithms: **recommendations were heavily geared towards popular books, regardless of the user's historical preference**. This popularity bias was most pronounced among long-tail readers, whose preferences were overridden by mainstream recommendations. Interestingly, these users did not necessarily prefer niche themes, they simply engaged with less popular books within dominant themes, thereby challenging the assumption that long-tail users inherently seek niche content.
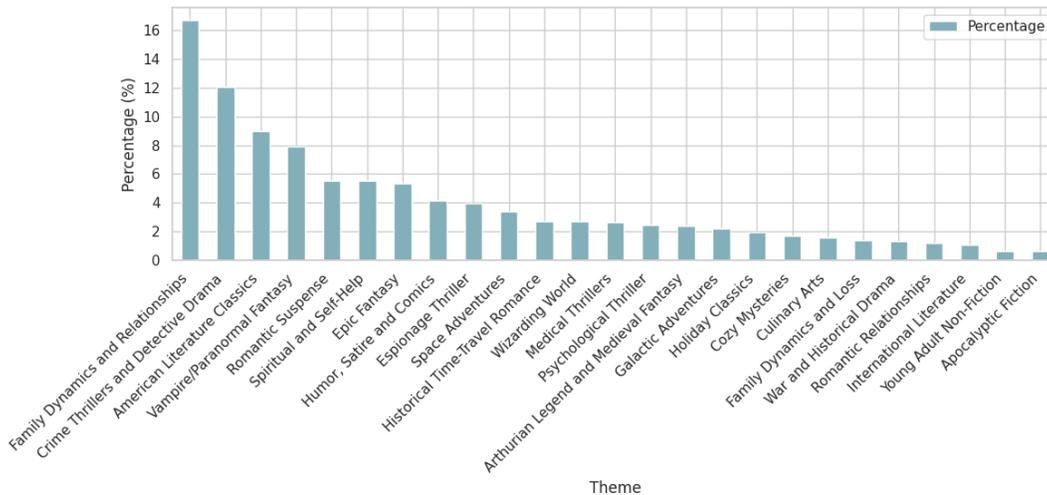
Figure 2: Distribution of themes among unique books in training dataset
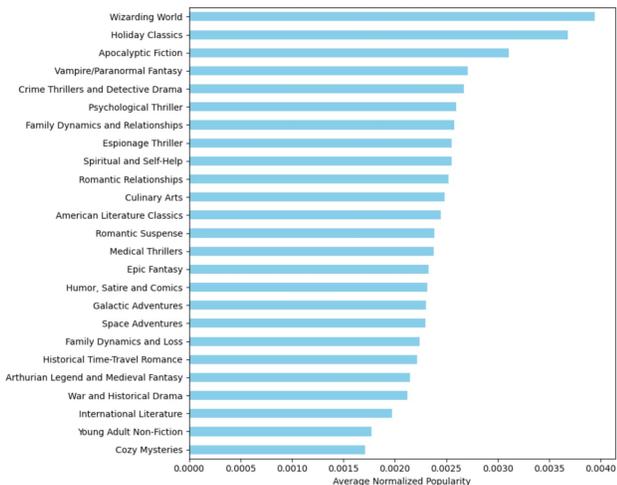

Figure 3: Average popularity ratio of themes

Next, we examined how recommendations differed for user groups defined by their historical thematic diversity. Specialist readers, with an average of 3.4 themes and a Gini index of 0.361, received more thematically diverse recommendations (6.8 themes, Gini reduced to 0.218), but this expansion often came with thematic displacement, as dominant themes in user profile shifted to *Vampire/Paranormal Fantasy* or *American Literature Classics*, depending on the algorithm. Generalist readers, with initially broad interests (10.2 themes, Gini 0.183), saw a reduction in diversity (to 7.2 themes) and were similarly steered towards dominant themes. Moderate readers showed mixed outcomes, some gained and others lost diversity. In summary, **despite starting from different levels of thematic diversity, both specialist and generalist readers were effectively normalized towards around 7 themes on average**, with a few themes being disproportionately represented.

This algorithmic homogenization raises concerns about long-term user engagement and discovery. The results are shown in Table 3.

By intersecting users' thematic preferences and propensity for popular items, we identified nine subgroups (Figure 5) and analyzed how their recommendations changed in terms of theme composition, popularity levels, and diversity. We observed that *specialist readers with long-tail reading preferences* were one of the most affected user groups. For these users, recommended books frequently diverged from their historical preferences, both in terms of dominant themes and popularity levels. In contrast, *moderate readers with mixed reading preferences* were least affected, with recommendation outputs remaining largely aligned with their past behavior.

Table 3: Average themes and gini ratio across user groups

User Group	Themes		Gini Ratio	
	Data	Rec.	Data	Rec.
Specialist	3.4	6.8	0.361	0.218
Moderate	6.8	7.1	0.226	0.204
Generalist	10.2	7.2	0.183	0.194

Therefore, to answer RQ3, thematic biases in recommendations impact users depending on their thematic diversity and propensity for popular items. We found a clear homogenizing effect where both specialist and generalist readers were steered towards a common range of themes. This homogenization was consistently biased towards a few dominant themes, displacing historical preferences. Our intersectional approach uncovered that users with both niche thematic preferences and long-tail consumption patterns experienced the highest degree of algorithmic skew towards popular items and dominant themes. This highlights that recommender systems can limit genuine discovery and reinforce echo chambers, especially for users with unique preferences.

Theme	BPR	MF	MostPop	NMF	NeuMF	PF	PMF	Random	UserKNN	VAECF	WMF
American Literature Classics	0.69	1.13	1.02	1.17	0.89	0.91	1.08	0.98	1.09	0.90	0.99
Apocalyptic Fiction	1.39	1.27	0.00	1.28	1.71	1.60	1.60	1.25	1.02	1.49	1.26
Arthurian Legend and Medieval Fantasy	0.81	0.85	0.00	0.94	0.84	0.49	0.87	0.86	0.68	0.52	0.85
Cozy Mysteries	0.13	0.53	0.00	0.70	0.18	0.04	0.59	0.69	0.66	0.16	0.69
Crime Thrillers and Detective Drama	1.19	0.83	0.00	0.58	1.34	1.11	0.93	1.07	0.90	1.22	1.07
Culinary Arts	0.66	0.81	0.00	0.96	0.69	0.98	0.86	1.00	0.74	0.66	0.96
Epic Fantasy	1.00	1.07	1.52	1.07	0.97	0.97	0.99	0.94	1.01	1.10	0.93
Espionage Thriller	1.00	0.76	0.00	0.66	0.89	1.24	0.85	1.03	0.82	0.87	1.02
Family Dynamics and Loss	0.64	0.85	0.00	0.76	0.41	0.59	0.78	0.90	0.80	0.28	0.92
Family Dynamics and Relationships	0.99	1.01	0.61	0.93	0.88	0.98	0.97	1.03	0.97	0.91	1.03
Galactic Adventures	1.01	1.04	0.00	1.04	0.91	0.95	1.07	0.93	0.92	1.16	0.92
Historical Time-Travel Romance	0.87	0.76	0.00	0.91	0.47	0.73	0.88	0.89	0.84	0.94	0.89
Holiday Classics	2.60	1.77	0.00	1.06	1.74	2.23	1.70	1.48	1.79	2.08	1.48
Humor, Satire and Comics	0.77	1.00	0.00	1.14	0.70	0.45	1.01	0.93	1.08	0.79	0.92
International Literature	0.53	0.76	6.13	1.08	0.88	0.65	0.71	0.79	0.73	0.73	0.79
Medical Thrillers	0.72	0.82	0.00	0.77	0.44	0.55	0.86	0.96	0.87	0.68	0.95
Psychological Thriller	1.08	0.84	0.00	0.95	0.26	0.53	0.95	1.04	1.03	0.98	1.07
Romantic Relationships	1.00	1.28	0.00	1.05	1.45	1.16	1.15	1.01	1.04	0.84	1.01
Romantic Suspense	1.03	0.71	0.00	0.72	0.98	0.79	0.83	0.96	0.80	0.90	0.95
Space Adventures	1.02	0.94	2.53	1.23	1.13	0.97	1.01	0.92	1.10	1.02	0.92
Spiritual and Self-Help	0.88	1.29	1.93	1.43	0.95	1.35	1.15	1.02	1.20	0.97	1.02
Vampire/Paranormal Fantasy	1.42	1.22	4.77	1.34	1.64	1.46	1.19	1.09	1.24	1.60	1.09
War and Historical Drama	0.30	1.01	0.00	1.23	0.00	0.61	0.92	0.85	0.85	0.39	0.86
Wizarding World	3.46	2.51	5.49	2.54	4.50	4.04	2.11	1.58	2.32	3.04	1.61
Young Adult Non-Fiction	0.00	0.41	0.00	0.41	0.00	0.00	0.63	0.71	0.60	0.05	0.69

Algorithm

Figure 4: Exposure ratio across different themes and algorithms

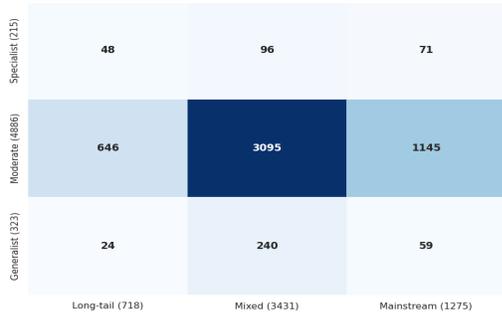|	Long-tail (718)	Mixed (3431)	Mainstream (1275)
Specialist (215)	48	96	71
Moderate (4886)	646	3095	1145
Generalist (323)	24	240	59

Figure 5: Intersectional subgroups

5 Conclusion and Future Work

This work examined thematic bias in book recommender systems, focusing on how such bias arises from both data imbalance and algorithmic reinforcement. We enriched the Book-Crossing dataset with thematic metadata and proposed a framework to evaluate recommendation fairness through a theme-centric lens. Our analysis clearly demonstrated presence of thematic bias, driven by a two-fold phenomenon: an initial imbalance in theme representation within the dataset's unique book content, which is further reinforced by user engagement patterns. This sets a clear precedent for downstream recommendation behavior, as algorithms are more likely to learn from dominant themes.

Instead of evaluating bias along a single dimension, we jointly considered user's preferences for popular content and thematic breadth. In particular, users with niche and long-tail reading habits often received poorly aligned recommendations, as algorithms nudged them towards more mainstream content. In contrast, users whose preferences aligned with dominant patterns in the dataset received consistent recommendations. These findings make it clear that evaluating recommender systems through a universal lens is not enough. To truly support a wide range of users, we need to understand how different groups are impacted in different ways.

While our approach offers valuable insights, it comes with certain limitations. The topic modeling technique assigned only one theme to each book, even though many books naturally cover multiple themes. While this simplification made our analysis more manageable, it might have affected the granularity of bias detection. Future work could improve thematic labeling by using multi-label annotation or LLM enhanced topic modeling techniques such as the one by Kapoor et al. [16]. Additionally, we focused solely on collaborative filtering algorithms without extensive fine-tuning and evaluated only the top-10 recommendations. Future studies could explore content-based or hybrid methods and assess performance across broader recommendation ranges. Finally, due to lack of directly comparable literature addressing thematic bias in book recommender systems, our findings largely stood alone without direct benchmarks for comparison. Looking ahead, applying the framework to other content domains such as movies, music, or podcasts could reveal how thematic bias differs across industries and whether common patterns emerge. Exploring the long-term effects of repeated exposure to biased recommendations on user behavior and engagement would highlight the broader consequences of bias and reinforce the urgency of developing effective bias-mitigation strategies.

References

[1] Himan Abdollahpouri, Masoud Mansoury, Robin Burke, and Bamshad Mobasher. 2019. The unfairness of popularity bias in recommendation. *arXiv* (2019). arXiv:cs.IR [cs.IR] http://arxiv.org/abs/1907.13286

[2] Dimo Angelov. 2020. Top2Vec: Distributed representations of topics. *arXiv* (2020). arXiv:2008.09470 [cs.CL] http://arxiv.org/abs/2008.09470

[3] Gitte Balling and Marianne Martens. 2024. BookTok Helped Us Sell It: How TikTok Disrupts Publishing and Fuels the Romantasy Boom. *Convergence: The International Journal of Research into New Media Technologies* (2024). doi:10.1177/13548565241301271

[4] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3 (2003), 993–1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

[5] Pierre Bourdieu. 1987. *Distinction: A social critique of the judgement of taste.* Harvard University Press.

[6] Lisa Bowleg. 2012. The Problem With the Phrase Women and Minorities: Intersectionality—An Important Theoretical Framework for Public Health. *American Journal of Public Health* 102, 7 (2012), 1267–1273. doi:10.2105/AJPH.2012.300750

[7] Jiawei Chen, Huimin Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. 2023. Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems* 41, 3 (2023), 1–39. doi:10.1145/3564284

[8] Sumi Cho, Kimberlé Williams Crenshaw, and Leslie McCall. 2013. Toward a Field of Intersectionality Studies: Theory, Applications, and Praxis. *Signs* 38, 4 (2013), 785–810. http://www.jstor.org/stable/10.1086/669608

[9] Savvina Daniil, Mirjam Cuper, Cynthia C. S. Liem, Jacco van Ossenbruggen, and Laura Hollink. 2022. Hidden author bias in book recommendation. *arXiv* (2022). arXiv:cs.IR doi:10.5281/zenodo.3865344

[10] Michael D. Ekstrand and Daniel Kluver. 2021. Exploring Author Gender in Book Rating and Recommendation. *User Modeling and User-Adapted Interaction* (2021). doi:10.1007/s11257-020-09284-2

[11] Daniel M. Fleder, Kartik Hosanagar, and Andreas Buja. 2010. Recommender systems and their effects on consumers: The fragmentation debate. In *Proceedings of the 11th ACM Conference on Electronic Commerce.* ACM, New York, NY, USA, 229–230. doi:10.1145/1807342.1807378

[12] Melanie Green and Timothy Brock. 2000. The Role of Transportation in the Persuasiveness of Public Narrative. *Journal of personality and social psychology* 79 (11 2000), 701–21. doi:10.1037/0022-3514.79.5.701

[13] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* (2022). arXiv:2203.05794 [cs.CL] http://arxiv.org/abs/2203.05794

[14] Sara Hajian, Francesco Bonchi, and Carlos Castillo. 2016. Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM.

[15] Patricia Hill Collins. 1990. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment.* Unwin Hyman.

[16] Satya Kapoor, Alex Gil, Sreyoshi Bhaduri, Anshul Mittal, and Rutu Mulkar. 2024. Qualitative Insights Tool (QualIT): LLM Enhanced Topic Modeling. *arXiv preprint arXiv:2409.15626* (2024). arXiv:2409.15626 [cs.IR] https://arxiv.org/abs/2409.15626

[17] Emily E. Lawrence. 2020. On the problem of oppressive tastes in the public library. *The Journal of Documentation; Devoted to the Recording, Organization and Dissemination of Specialized Knowledge* 76, 5 (2020), 1091–1107. doi:10.1108/jd-01-2020-0002

[18] Daniel D Lee and H Sebastian Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 6755 (1999), 788–791.

[19] Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. 2020. Feedback loop and bias amplification in recommender systems. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management.* ACM.

[20] Mahdi Naghiaei, Hossein Amir Rahmani, and Mostafa Dehghan. 2022. The unfairness of popularity bias in book recommendation. In *Communications in Computer and Information Science.* Springer International Publishing, 69–81.

[21] Ehsan Noordeh, Roi Levin, Ruocong Jiang, and Hamid Shadmany. 2020. Echo chambers in collaborative filtering based recommendation systems. *arXiv* (2020). arXiv:2011.03890 [cs.IR] http://arxiv.org/abs/2011.03890

[22] Martina Ridzuan and Norsyuhada bt Ahmadrashidi. 2023. The Phenomenon of Booktok and Its Impact on The Transformation of Literary Culture and The Publishing Industry. *Forum Komunikasi* 18, 2 (2023), 74–83. https://ir.uitm.edu.my/id/eprint/112680/1/112680.pdf

[23] Shlomith Rimmon-Kenan. 1995. What is theme and how do we get at it? In *Thematics: New Approaches*, Claude Bremond, Joshua Landy, and Thomas G. Pavel (Eds.). SUNY Press, 9–20.

[24] Will Self. 2014. The novel is dead (this time it's for real). https://www.theguardian.com/books/2014/may/02/will-self-novel-dead-literary-fiction. *The Guardian* (2 May 2014). https://www.theguardian.com/books/2014/may/02/will-self-novel-dead-literary-fiction Retrieved on June 18, 2025.

[25] Harald Steck. 2018. Calibrated recommendations. In *Proceedings of the 12th ACM Conference on Recommender Systems* (Vancouver, British Columbia, Canada) *(RecSys '18).* Association for Computing Machinery, New York, NY, USA, 154–162. doi:10.1145/3240323.3240372

[26] Ilya Vayansky and Satish Ananda Prabhu Kumar. 2020. A review of topic modeling methods. *Information Systems* 94 (2020). doi:10.1016/j.is.2020.101582

[27] Chong Wang and David M. Blei. 2011. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.* 448–456.

[28] Wikipedia contributors. 2024. *Lists of The New York Times number-one books.* https://en.wikipedia.org/wiki/Lists_of_The_New_York_Times_number-one_books Accessed June 30, 2025.

[29] Wikipedia contributors. 2024. *Publishers Weekly list of bestselling novels in the United States in the 2000s.* https://en.wikipedia.org/wiki/Publishers_Weekly_list_of_bestselling_novels_in_the_United_States_in_the_2000s Accessed June 30, 2025.

[30] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving Recommendation Lists Through Topic Diversification. In *Proceedings of the 14th International Conference on World Wide Web.* ACM, 22–32. doi:10.1145/1060745.1060754