Let Me Entertain You: On the Bias of Editorial and Algorithmic Recommendations in Public Service Media

Marc de Zoeten
Johannes Gutenberg University Mainz
Mainz, Germany *mdezoet@uni-mainz.de*

Marcel Hauck
Mainz University of Applied Sciences
Mainz, Germany *marcel.hauck@hs-mainz.de*

Martin Briesch
Johannes Gutenberg University Mainz
Mainz, Germany *briesch@uni-mainz.de*

Dominik Sobania
Johannes Gutenberg University Mainz
Mainz, Germany *dsobania@uni-mainz.de*

Franz Rothlauf
Johannes Gutenberg University Mainz
Mainz, Germany *rothlauf@uni-mainz.de*

Abstract

In today's online world, bias is ubiquitous. In a Video-on-Demand (VOD) context, there are two main sources for bias: (1) content providers and their methods that generate recommendations; (2) user preferences. We use real world data from one of the largest German Public Service Media (PSM) platforms to study both possible causes by examining content recommendations and content consumption (classified as Education, Information, Culture, and Entertainment according to the regulations governing PSM operations). We find that Education and Entertainment content to be recommended disproportionately. Algorithmic personalized recommendations lead to a bias towards both Education and Entertainment for the recommendations. Editorial content recommendations favor Entertainment content. Additionally, we find that the bias introduced by the recommender systems to be larger than the one introduced by the viewer. We recommend testing an increased use of algorithmic recommendations as these might lead to a lower bias towards Entertainment content.

Keywords: Public Service Media, Editorial Control, Recommendations, Bias, Curation Systems.

1. Introduction

Recommender systems are widely deployed by commercial and Public Service Media (PSM) video-on-demand (VOD) platforms alike [2, 3]. While it is widely assumed that virtually all recommender systems introduce some bias(es) when presenting items to users, the specific biases and their respective impacts are of interest [6], [11]. Commercial VOD services often ultimately care about profit and adjust their recommender systems accordingly [4, 5], [13]. Conversely, PSM platforms in Germany are tasked to provide the general population with a supply of content for five content categories, namely Education, Information, Culture, Entertainment, and Advice [8].

To fulfill this mission, PSM produce and buy content which is broadcast on PSM linear television (TV) and is made available on VOD platforms. The latter make use of two main methods to generate recommendations for videos. First, manual editorial recommendations are used, typically generated by experts and are identical for all users.

Second, algorithmic recommendations generated automatically by algorithms. In the context of this investigation, we specifically examine recommendations that are personalized, which we will refer to as algorithmic recommendations. When deployed, these different methods to generate recommendations for users may produce unintended biases and should therefore be carefully examined.

The generated recommendations are displayed to users of the PSM VOD platform, who have specific preferences for certain content [17], [19]. These preferences might lead them to favor one content category over another, which in turn, influences the distribution of consumed content and introduces bias in the consumption patterns on the platform. To gain insights into how specifically these biases affect consumption in a live setting, we analyze a large real-world dataset from a German PSM platform with regard to the biases introduced when the platforms makes recommendations to its users and biases introduced by user preferences.

Overall, our research has three goals: (1) explore the role of algorithmic and editorial recommendations as well as potential bias induced by these methods to produce recommendations (2) compare the offerings and recommendations to other available data, and finally (3) identify current consumption patterns or distributions for content.

More specifically, we first analyze the distribution of available content on the platform regarding the content categories mentioned above. Subsequently, we analyze the recommendations regarding the distribution of the content categories from (1) the editorial, and (2) the algorithmic recommendations. We then compare these distributions to the content offerings of the two largest German PSM providers on linear TV, as no comparable data is available for the field of PSM VOD-platforms. Afterwards, we analyze the consumption patterns on the platform. The global consumption patterns are analyzed and subsequently broken up into the consumption patterns based on each of the methods to generate recommendations.

Additionally, we compare the distributions for consumed content with the distribution of content users consume after searching for it manually making use of the search functionality to investigate what content might not be covered by either the editorial or algorithmic recommendations.

Our analysis of real-world data is highly relevant for the field of recommender systems in the field of PSM. While many past studies have shown a variety of effects in the lab [7], this research often faces numerous limitations with regard to its practical applicability [14]. Real-world insights can address some of these limitations.

Following this introduction, we briefly discuss related work and define which results we view as unbiased and which we view as biased. Subsequently we explain our methodology including how we processed our dataset. We then show and discuss our results.

2. Related Work

2.1. Recommender-induced biases

Biases in the field of VOD-platforms are frequently caused by the methods employed by VOD platform operators deploying bias-inducing recommender systems [1], [15]. Among the most influential biases affecting recommendations are (1) popularity bias, which refers to popular items being recommended with increased frequency, resulting in these popular items making up a disproportional share of recommendations, (2) presentation bias, which refers to the influence displaying items in a certain way or order has on consumption, and (3) selection bias, which refers to the (systematic) restrictions for recommendations of certain items [12], [18].

While personalized recommendations, which were generated algorithmically, commonly fit the preferences of users quite well, these recommendations are commonly assumed to have a number of downsides such as forming filter bubbles, or offering very little diversity in the recommended content [16], [24, 25]. With regard to our analysis of PSM data, this could indicate a high share of recommendations for items from just one content category that performs really well with regard to clicks, such as Entertainment.

Conversely, editorial recommendations, or at least some degree of editorial oversight over recommendations, is assumed to be beneficial to the user, increasing exposure to new and diverse content offerings and thus potentially negating some biases [10], [21].

Unbiased or fair recommendations from the perspective of the PSM provider refers to the equal chance for all items/content categories to be recommended and displayed to the user, which might be referred to as equality of opportunity [cf. 6], [cf. 9]. However, to the best of our knowledge, the case of bias regarding recommendations with content categories deducted from German (or any other) regulations governing PSM operations [cf. 8] has not been explored in the literature so far. Consequently, we provide novel insights into this field.

We define recommendations as unbiased in the realm of PSM platforms, if the distribution of recommendations is identical (similar) to the distribution of available content. The intensity of the bias is considered higher, if the distribution of recommended content is further away from the distribution of available content [cf. 23].

2.2. User-induced biases

Even in the case of completely unbiased recommendations, consumption of content may still be heavily skewed towards specific items or content categories, as consumers may have preferences towards some of these [6]. Additionally, some content categories such as Information, may only have limited utility that comes with additional consumption, whereas content categories such as Entertainment and Education are more likely to be stable in utility to users [12]. Furthermore, Information content might be outdated and consequently irrelevant fairly quickly. Furthermore, some items within a content category may be more relevant to certain users. In the case of relevant items being presented for one content category and irrelevant items being presented in another, consumption is likely to be skewed towards the content category with relevant items [22].

We define consumption as unbiased by user preferences, if the distribution of the consumption is identical (similar) to the distribution of recommended content, with intensity of the bias the higher, the further the distribution of consumed content is from the distribution of recommended content. The user-induced bias therefore does not refer to any changes in the algorithmic recommendation, but solely to which content a user chooses from the content that is presented to them.

3. Methodology

3.1. Approach

To enable us to investigate the data, we analyzed the specific data we received, augmented the dataset with additional information and subsequently cleaned the dataset to prepare it for our analyses. Subsequently, we started our investigation by analyzing if the methods to generate recommendations (editorial and algorithmic) cause a bias by comparing the distribution of content that is available on the platform to the distribution of content recommended by the two investigated methods to generate recommendations (editorial and algorithmic). We then analyze the observed bias. We additionally compare the methods that generate recommendations to that from the offering to the user on the VOD-platform to the offerings on PSM TV.

To prepare our analysis of the consumed content split by the methods the content was recommended, we shortly examine if the clicks are indicative of playback time. Afterwards, our investigation of biases shifts to potential user-induced biases. We analyze the distribution of consumed content caused by each of the two methods to generate recommendations. Additionally, we further examine the distribution of content that users consume when making of the platform's search functionality to investigate whether any particular content is missing from the recommendations made by the two investigated methods to generate recommendations.

3.2. Dataset

In our study, we use a large real-world dataset provided by one of the largest German PSM providers. This dataset contains (1) all assets (websites containing full videos (the focus of this study) as well as pages that do not contain a full video, e.g., a news article or a site containing a multitude of videos to click) that were available at the time, (2) a video-specific playback time in an aggregated format, allowing us to track how long a particular video has been viewed with two months of aggregated playback data, and (3) data about user consumption behavior covering 15 days, particularly the click-behavior, and (4) data covering algorithmic recommendations for 15 days. All data is from 2022.

To ensure that we are unable to track individual users, 50% of click data (3) was deleted at random before it was provided to us. This restriction was agreed upon with the responsible data protection officers to ensure that individuals cannot be tracked.

We were not provided with data on editorial recommendations. We therefore made use of the internet archive "Wayback Machine" (https://web.archive.org/) to obtain the data about the editors' recommendations and augment our existing data. Due to restrictions of the archive, we only have complete data on 11 out of 15 days. To this end, we programmed a web-scraper to extract relevant information from the archive, which we matched with our existing asset data (see (1) above).

3.3. Data processing

We have excluded all content (partially) aimed at minors (below the age of 18) from all analyses. This has been implemented for two reasons: algorithmic recommendations aimed at minors are not representative of the wider recommendations made by the system and are also generated differently. Furthermore, the metadata for this content did not allow for any insight into the actual content of these videos. Consequently, it was not possible to assign this content to any specific content category. The number of clicks and recommended items for the content category of Advice was <1%, we excluded it from all analyses of the PSM VOD platform and base all analyses for which this is appropriate on the four content categories Education, Information, Culture, and Entertainment.

Additionally, we excluded a number of videos, where the metadata could also not be assigned to a content category. The sum of the clicks and the videos excluded for this reason makes up less than .5% of clicks on videos. Videos that have been excluded for this reason include videos that are marked as focused on providing content for the visually or hearing impaired.

When processing the data obtained through the "Wayback Machine", ~3% of the collected recommendations could not be attributed to any content category. Examples for these items include links to live-television broadcasts in general (which cannot be attributed to any content category), links that are used to link to television of the previous days as well as links that do not contain videos, as they are lists of content that cannot be classified as belonging to any content category.

4. Results and discussion

In a first step, we analyze the distribution of all full-length videos. Table 1 displays (1) the share of the number of videos available, (2) the share of total video duration across the content categories. The abbreviations for the content categories are used throughout all tables (Edu: Education, Inf: Information, Cul: Culture, Ent: Entertainment). It can be noted that the content category Information has the largest share of videos available at 42%. However, the Information videos present only 22% of the video length. This can be attributed to the on average much shorter videos in this content category, as news are typically shorter when compared to, e.g., a feature-length film in the Entertainment category.

Table 1. Available Content

Share of	Edu	Inf	Cul	Ent
Number of videos	24%	42%	16%	18%
Total video duration	27%	22%	20%	31%

In a second step, we analyze the distribution of recommended content by (1) the editors and (2) via personalization based on algorithms to the distribution of available content to assess whether any of the ways to recommend content exhibit a bias and if so, the magnitude of that bias. A Chi² test (performed in SPSS 27) indicates that the distribution across the content categories is statistically significantly different for both the editors ($p = 0.000$) and personalized recommendations ($p = 0.000$). In the case of editorial recommendations, 69% of the recommendations are for the content category of Entertainment, while only 18% of the videos on the platform are in the content category of Entertainment. The share of editorial recommendations for Entertainment makes Entertainment the sole and overwhelming focus of the editorial recommendations.

For the case of personalized algorithmic recommendations, Education and Entertainment content is recommended almost in equal proportions (43% and 47% respectively). Both content categories are recommended out of proportion, with Entertainment further from the available content (47% versus 18%) compared to Education (43% versus 24%).

When comparing the distribution of content provided by both the editors and the personalized algorithms to the linear content provision of the largest German PSM providers for linear TV, it is obvious that in linear TV, Information content is quite common with 35% and 33% for Das Erste and ZDF (Germany’s two largest PSM providers), respectively. The provision of Entertainment in linear TV (57% and 54%) is right in between the Editor's provision (69%) and the provision of personalized algorithms (47%). For Education, both the editors (16%) and the personalized algorithms (43%) have substantially higher shares than linear TV (6% and 9%).

It can be assumed that the linear television offers are closely monitored and editorially curated.

Table 2. Recommendations on the PSM platform and values for the linear TV PSM channels Das Erste & ZDF from [20]. 4% of data could not be assigned to a content category. Culture was not disclosed. Remaining content was scaled to 100%.

	Edu	Inf	Cul	Ent	Advice
Editorial	16%	6%	8%	69%	<1%
Algorithmic	43%	3%	6%	47%	<1%
Das Erste 2022	6%	35%	-/-	57%	2%
ZDF, 2022	9%	33%	-/-	54%	4%

In a third step, we investigate the content consumption, as measured in clicks and playback time. It is displayed in Table 3. Overall, most clicks go towards Entertainment with 72% of clicks. Similarly, 82% of playback time falls to the content category of Entertainment. Generally, the share of clicks is quite indicative of the share of playback time. The only exception is the content category of Information, for which the share of playback time is substantially smaller than the share of clicks.

Table 3. Content Consumption.

Share of	Edu	Inf	Cul	Ent
Clicks	11%	8%	9%	72%
Playback time	10%	2%	6%	82%

To further investigate how this distribution is formed, we investigate three different paths users can take to consume content: Users can (1) click on editorial recommendations (2) algorithmic recommendations, or they can (3) actively search for an asset using the search function (user’s choice). There exist additional options for choosing a video, such as clicking on a link provided by an external source (e.g., a person shares a video, which is subsequently clicked or a direct link is provided elsewhere on the internet). These additional options are not investigated. Furthermore, clicking a video that is recommended in any way while already watching a video is also not accounted for, as there is no click stream analysis. Consequently, these three ways to click content do not account for all clicks. While clicks can be split up this way, playback time cannot be analyzed on this level due to a lack of provided/received data. However, while the share of clicks regarding a certain content category is not a perfect predictor for the share of total playback time, we can assume that the ratio does not change in any major way. Additionally, except for the content category of Information the share of clicks is a rather reliable metric for the share of total playback time.

Overall, the numbers from Table 4 strongly suggest that the higher the presented share of a content category, the higher the share of clicks, which indicates that recommendations are mainly responsible for consumption on the platform, as users will largely consume what is recommended to them. The user-induced bias exists, but is comparably small compared to the bias introduced by the methods to generate recommendations. This highlights the high level of responsibility required when designing recommender systems in the field of PSM.

Table 4. Content Consumption across different consumption channels

Clicks caused by	Edu	Inf	Cul	Ent
Editorial recommendations	11%	5%	4%	80%
Algorithmic recommendations	38%	<1%	8%	54%
Search engine	9%	6%	18%	66%

5. Conclusion

We have analyzed a large real-world dataset from a large German PSM VOD platform. Specifically, we have reviewed the content available on the platform, which content is recommended either editorially or algorithmically (personalized) and how the users react to those recommendations, and the resulting consumption patterns. Furthermore, we have compared our findings regarding the offerings to the offerings available in linear TV programs and we have compared the consumption patterns to consumption, which are not based on recommendations. We found that the distribution of the editors' recommendations as well as personalized algorithmic recommendations to be heavily skewed compared to the distribution of available content, with editorial recommendations being skewed towards Entertainment and algorithmic recommendations towards both Entertainment and Education.

We found the influence of consumer preferences to be small compared to the investigated methods that generate recommendations.

Some content (categories), specifically Culture, seem to be underserved by both algorithmic and editorial recommendations, prompting users to disproportionally find and consume this content via other means (i.e., search functionality).

The platform only provides substantial content for four out of five content categories that PSM providers should serve, with a severe lack of Advice content in available assets, recommendations via any method and consumption. This may indicate that Advice content is not produced in the same quantities, or the metadata does not classify this content correctly (see our limitations).

We recommend that PSM platforms carefully review how they provide content and which content exactly is recommended to users. This includes both algorithmic and editorial recommendations. This further includes reviewing the "mix" of editorial and algorithmic recommendations, potentially conducting real-world experiments to test the effects of additional algorithm-based recommendations.

6. Limitations

When using the metadata available to us to match an asset to the relevant content categories for PSM, we assumed that a given asset can only be assigned in full to a single content category. This can be considered a strong assumption. In practice, comedy-content (Entertainment) can inform about current events in the world (Information). Likewise, a documentary (Education) can have strong elements of culture and additionally entertain its audience. We assume that all content categories receive and lose assets in a similar way. Furthermore, regarding the recommendations, both editorial and algorithmic, the existence of a recommendation does not imply that it has been perceived by a user, as the exact position of a recommendation is at least somewhat ambiguous.

As 50% of click-data was deleted at random alongside some identifiers to ensure adequate privacy for the users of the platform, virtually any (long) click-stream analyses regarding individuals that consume content on the platform were interrupted.

Furthermore, the consumption patterns as well as recommendations made both by editors as well as algorithms may depend on other external factors and not be stable over time, highlighting a need for continuous monitoring to expand on our analysis.

7. Future Research

In the future, research can be conducted to address at least some of the limitations in our study and gain further insights into the topic. First, the design of specific experiments that can be conducted both in a laboratory and later on with a live-system can offer deeper insights into the metrics and topics explored in this paper. Additionally, the content available could be reclassified, either by the PSM provider or researchers. This would enable content to belong to more than one content category or enable researchers to add additional content categories such as Infotainment (Information & Entertainment). This ultimately allows for deeper insights into the structure of the available content and the consumed content. Additionally, while ensuring the privacy of individual users, click-stream could be further analyzed to gain insights into the consumption patterns of individual users additionally to the platform-wide consumption view. Additionally, a complete and continuous analysis of recommendations and user behavior can provide additional insights.

References

1. Adomavicius, G., Bockstedt, J., Curley, S., Zhang, J.: Reducing Recommender Systems Biases: An Investigation of Rating Display Designs. MIS Quarterly 43, 1321–1342 (2019)
2. Alfano, M., Fard, A.E., Carter, J.A., Clutton, P., Klein, C.: Technologically Scaffolded Atypical Cognition: the Case of YouTube's Recommender System. Synthese 199, 835–858 (2021)
3. Álvarez, M.V., López, J.M.T., Ruíz, M.J.U.: What Are You Offering?: An Overview of VODs and Recommender Systems in European Public Service Media. In: Information Technology and Systems Proceedings, vol. 1137, pp. 725–732. Springer International Publishing, Cham (2020)
4. Aridor, G., Goncalves, D., Sikdar, S.: Deconstructing the Filter Bubble: User Decision-Making and Recommender Systems. In: Fourteenth ACM Conference on Recommender Systems, pp. 82–91. ACM, New York, NY, USA (2020)
5. Azaria, A., Hassidim, A., Kraus, S., Eshkol, A., Weintraub, O., Netanely, I.: Movie Recommender System for Profit Maximization. In: Proceedings of the 7th ACM conference on Recommender systems, pp. 121–128. ACM, New York, NY, USA (2013)
6. Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., He, X.: Bias and Debias in Recommender System: A Survey and Future Directions. ACM Trans. Inf. Syst. 41, 1–39 (2023)

7. Chen, J., Wang, X., Feng, F., He, X.: Bias Issues and Solutions in Recommender System. In: Fifteenth ACM Conference on Recommender Systems, pp. 825–827. ACM, New York, NY, USA (2021)
8. Die Medienanstalten: Medienstaatsvertrag (in German) (2024)
9. Elahi, M., Abdollahpouri, H., Mansoury, M., Torkamaan, H.: Beyond Algorithmic Fairness in Recommender Systems. In: Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, pp. 41–46. ACM, New York, NY, USA (2021)
10. Herm-Stapelberg, N., Rothlauf, F.: The crowd against the few: Measuring the Impact of Expert Recommendations. Decision Support Systems 138, 113345 (2020)
11. Hildén, J.: The Public Service Approach to Recommender Systems: Filtering to Cultivate. Television & New Media 23, 777–796 (2022)
12. Huang, J., Oosterhuis, H., Rijke, M. de: It Is Different When Items Are Older. In: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, pp. 381–389. ACM, New York, NY, USA (2022)
13. Jannach, D., Jugovac, M.: Measuring the Business Value of Recommender Systems. ACM Transactions on Management Information Systems 10, 1–23 (2019)
14. Jannach, D., Lerche, L., Kamehkhosh, I., Jugovac, M.: What Recommenders Recommend: an Analysis of Recommendation Biases and Possible Countermeasures. User Modeling and User-Adapted Interaction 25, 427–491 (2015)
15. Lin, C., Liu, X., Xv, G., Li, H.: Mitigating Sentiment Bias for Recommender Systems. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 31–40. ACM, New York, NY, USA (2021)
16. Michiels, L., Leysen, J., Smets, A., Goethals, B.: What Are Filter Bubbles Really? A Review of the Conceptual and Empirical Work. In: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, pp. 274–279. ACM, New York, NY, USA (2022)
17. Prince, J., Greenstein, S.: Measuring Consumer Preferences for Video Content Provision via Cord-Cutting Behavior. Journal of Economics & Management Strategy 26, 293–317 (2017)
18. Qian, K., Jain, S.: Digital Content Creation: An Analysis of the Impact of Recommendation Systems. Management Science 70, 8668–8684 (2024)
19. Rodríguez, D., Rosa, R., Costa, E., Abrahão, J., Bressan Graca: Video quality assessment in video streaming services considering user preference for video content. IEEE Transactions on Consumer Electronics 60 (2014)
20. Rühle, A.: Ergebnisse der ARD-Programmanalyse 2023. Programmprofile von Das Erste, ZDF, RTL, Vox, Sat.1 und ProSieben. (in German) Media Perspektiven (2024)
21. Sørensen, J.K.: Public Service Media, Diversity and Algorithmic Recommendation: Tensions between Editorial Principles and Algorithms in European PSM Organizations. In: CEUR Workshop Proceedings, pp. 6–11 (2019)
22. Yue, Y., Patel, R., Roehrig, H.: Beyond position bias. In: Proceedings of the 19th international conference on World wide web. ACM, New York, NY, USA (2010)
23. Zafar, M.B., Valera, I., Gomez Rodriguez, M., Gummadi, K.P.: Fairness Beyond Disparate Treatment & Disparate Impact. In: Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2017)
24. Zangerle, E., Bauer, C.: Evaluating Recommender Systems: Survey and Framework. ACM Computing Surveys 55, 1–38 (2023)
25. Zuiderveen, B., Frederik J., Trilling, D., Möller, J., Bodó, B., Vreese, C.H. de, Helberger, N.: Should we Worry about Filter Bubbles? Internet Policy Review 5 (2016)