Legal but Unfair: Auditing the Impact of Data Minimization on Fairness and Accuracy Trade-off in Recommender Systems

Salvatore Bufi*
Politecnico di Bari
Bari, Italy
s.bufi@phd.poliba.it

Vincenzo Paparella*†
ISTI-CNR
Pisa, Italy
vincenzo.paparella@isti.cnr.it

Vito Walter Anelli
Department of Electrical and Information Engineering
Politecnico di Bari
Bari, Italy
vitowalter.anelli@poliba.it

Tommaso Di Noia
Department of Electrical and Information Engineering
Politecnico di Bari
Bari, Italy
tommaso.dinoia@poliba.it

Abstract

Data minimization, required by recent data privacy regulations, is crucial for user privacy, but its impact on recommender systems remains largely unclear. The core problem lies in the fact that reducing or altering the training data of these systems can drastically affect their performance. While previous research has explored how data minimization affects recommendation accuracy, a critical gap remains: How does data minimization impact consumers' and providers' fairness? This study addresses this gap by systematically examining how data minimization influences multiple objectives in recommender systems, i.e., the trade-offs between accuracy, user fairness, and provider fairness. Our investigation includes (i) an analysis of how the data minimization strategies affect RS performance across these objectives, (ii) an assessment of data minimization techniques to determine those that can balance better the trade-off among the considered objectives, and (iii) an evaluation of the robustness of different recommendation models under diverse minimization strategies to identify those that best maintain performance. The findings reveal that data minimization can sometimes undermine provider fairness, albeit enhancing group-based consumer fairness to the detriment of accuracy. Additionally, different strategies can offer diverse trade-offs for the assessed objectives. The source code supporting this study is available at https://github.com/salvatore-bufi/DataMinimizationFairness.

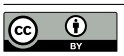
CCS Concepts

• **Information systems** → **Recommender systems**; • **Social and professional topics** → **Governmental regulations**.

Keywords

Data Minimization, Recommender Systems, Fairness, Multi-Objective Evaluation

*Corresponding authors.
†Work done while being a PhD student at Politecnico di Bari.



UMAP '25, New York City, NY, USA

ACM ISBN 979-8-4007-1313-2/25/06
https://doi.org/10.1145/3699682.3728356

ACM Reference Format:
Salvatore Bufi, Vincenzo Paparella, Vito Walter Anelli, and Tommaso Di Noia. 2025. Legal but Unfair: Auditing the Impact of Data Minimization on Fairness and Accuracy Trade-off in Recommender Systems. In *33rd ACM Conference on User Modeling, Adaptation and Personalization (UMAP '25), June 16–19, 2025, New York City, NY, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3699682.3728356

1 Introduction

Recommender Systems (RSs) have become a cornerstone for modern digital platforms, powering personalized content delivery across e-commerce, media streaming, and social networks [24, 27, 28, 35]. These systems rely on extensive user data to accurately infer preferences and produce highly tailored recommendations. However, growing concerns about user privacy, data security, and system scalability create a significant tension between the need to maintain high-quality recommendations and the increasingly important requirement to minimize data collection [14, 16]. This tension is further intensified by rigid regulatory frameworks like the GDPR [3], CCPA [1], CPRA [2], and PIPL [4], which command principles such as data minimization. The legal principle of data minimization underscores the need to collect only the strictly necessary data for a specific purpose. Given these regulations, many works have started to study the effects and risks of data minimization on machine learning-driven systems [9, 18, 39, 44]. In the realm of recommendation, Biega et al. [7] performed pioneering work to understand the performance-based effects of data minimization on system accuracy. Hence, other works have analyzed the influence of data minimization on privacy [15], algorithmic bias [9], and individual users' accuracy [37] performance of recommendations. Nonetheless, data minimization reshapes available user-item interactions by discarding or selectively retaining interactions, potentially altering the distribution of users and items. These alterations can amplify existing biases and undermine system fairness from consumer and producer perspectives. For this reason, it is pertinent to ask: *To what extent does data minimization risk to impact fair exposure of items? Can data minimization favor privileged user groups with higher-quality recommendations?* For example, a popular item may appear underrepresented in a minimized dataset if many interactions are removed, potentially pushing it into the long-tail category and reducing its visibility. Similarly, specific user groups may receive

less accurate or less diverse recommendations, further widening existing disparities. While these risks are evident, a comprehensive understanding of the broader ramifications of data minimization on accuracy and fairness remains lacking, highlighting a critical gap in current research, given the increasing emphasis on fairness [6, 12, 31, 32, 43] in RSs.

This study bridges the gap by systematically investigating the interplay between data minimization, accuracy, and fairness in recommendation. To this end, we investigate how data minimization singularly impacts the accuracy and fairness of recommendations by identifying observable trade-offs. Then, we adopt a multi-objective evaluation approach to audit to what extent different data minimization strategies can preserve RS performance on multiple dimensions. Finally, we analyze how different recommendation models are robust to data minimization, maintaining a balance between accuracy and fairness. In summary, our key contributions are the following:

- We experimentally evaluate the impact of data minimization on recommendation accuracy, consumer fairness, and provider fairness across various data minimization strategies and real-world datasets. The evaluation demonstrates that the effectiveness of data minimization cannot be evaluated by accuracy alone but must consider how it affects different stakeholders in the system.
- We adopt a multi-objective evaluation approach to study which data minimization strategies can simultaneously preserve recommendation performance when considering multiple aspects. Our findings reveal that strategies that can shape better user profiles are crucial to avoid alterations in recommendation performance.
- We investigate the effects of data reduction on various recommendation algorithms, including neighborhood-based, matrix factorization, autoencoder-based, and graph-based methods. We understand how different model architectures react to data minimization when balancing the competing objectives of accuracy and fairness. We show that graph-based and factorization-based models can preserve recommendation performance under strong minimization conditions.

2 Related Work

User privacy [10, 13, 26] and data privacy [22, 45] have become a critical concern in recommender systems research, driven by regulations such as the GDPR, CCPA, CPRA, and PIPL. These legal frameworks mandate that only data necessary for a specific purpose be collected, creating tension between the extensive datasets typically required by recommender systems and the principle of data minimization.

Several works have begun to explore how to reconcile the need for personalized recommendations with the constraints imposed by these regulations. Biega et al. [7] introduced operational definitions of data minimization aimed at maintaining system quality while satisfying legal requirements. However, they also pointed out that minimization can worsen outcomes for specific user groups if not carefully managed. A family of works focuses on designing data minimization techniques to be integrated during model training to retain acceptable accuracy performance. Shanmugam et al. [36] proposed FIDO, a strategy that adaptively restricts data collection based on performance scaling rules, illustrating how data minimization can be integrated directly into training workflows. Sonboli et al. [37] have explored the trade-off between reducing data and maintaining accuracy at the user level by using active learning approaches that select only highly informative user-item interactions. While these methods can improve efficiency, they can also worsen outcomes for groups with sparse interaction histories. Other works [38, 42] study the effects of data reduction on environmental factors, showing that smaller datasets obtained through data reduction lead to reduced energy consumption but may compromise recommendation accuracy. Regarding the interplay between data minimization and fairness, Clavell et al. [9] highlight the complexities of simultaneously addressing privacy, performance, and fairness, emphasizing the need for more comprehensive frameworks to evaluate these trade-offs.

The existing works concentrate on a single objective—such as privacy, bias, or accuracy—rather than investigating how these factors interact in a unified manner. Our work extends the current state-of-the-art by systematically examining how data minimization strategies simultaneously affect both accuracy and fairness in recommender systems.

3 Experimental Setup

This research carries out a comprehensive investigation into the effects of data minimization on recommender systems, moving beyond a simple accuracy analysis to consider its broader implications for both users and item providers. In this section, we detail the experimental setup employed to address the following research questions:

RQ1: How does data minimization impact the accuracy, provider fairness, and consumer fairness of recommendations? What are the observable trade-offs at the strategy level?
RQ2: Which data minimization strategies most effectively preserve recommender system performance compared to training on the full dataset?
RQ3: How robust are different recommendation models to data minimization, particularly concerning maintaining a balance between accuracy and fairness?

3.1 Dataset

Our experiments were conducted using two datasets to ensure a comprehensive evaluation: *MovieLens1M* (ML1M) [20] and *Ambar* [19]. Setting a minimum interaction threshold is essential when studying data minimization since it ensures each user profile is represented well enough for reliable model training and evaluation [7]. For this reason, we preprocessed both datasets to guarantee sufficient representation for each user and item, aligning with the methodology of Biega et al. [7]. In ML1M, we selected users who have at least 45 ratings in their profile. Furthermore, we subsampled the dataset to 2500 users, covering 3605 items and totaling 522152 ratings. The resulting dataset has a density of 0.0579, with a median of 137 interactions per user and an average of 208.86. In Ambar, we applied an iterative 45-core filtering, ensuring that each user rated at least 45 items and that at least 45 users rated each item. We then subsampled the dataset to 2500 users, 4996 items, and 214678 total ratings, resulting in a density of 0.0172. The median number of interactions per user is 74, with an average of 85.87.

3.2 Evaluation Protocol

Our experimental evaluation aims to emulate a real-world scenario in which users can grant permission for their data storage. Hence, solely data related to users who allowed their data storage explicitly are included in the analysis, resulting in a reduced dataset called *Minimization Data* (D_M).

To create this scenario, the users of the entire dataset are randomly divided into two groups. The first group, representing 70% of all users, simulates those who do not consent to data storage and are excluded from the experiments. The second group consists of the remaining 30% of users who have consented to their data being stored. The interaction data related to these users constitute D_M. Within D_M, each user's ratings are further partitioned into three sets: a candidate set (70%), a validation set (10%), and a test set (20%). The candidate set serves as a pool of user data from which different minimization strategies can select a subset of ratings to train the system. Data is selected from this candidate set using a specific data minimization strategy and a minimization parameter n, which indicates the number of items to choose for each user. The study uses the following values for $n = \{1, 3, 7, 15, 100\}$. This approach draws inspiration from Biega et al. [7] but adapts it to a practical scenario, where individual user choices determine the availability of their data to the system.

3.3 Algorithms and Metrics

To provide a comprehensive view of how data minimization affects different types of recommender systems, we consider five representative algorithms spanning different methodological families:

- **LightGCN** [21] is a graph-based model that simplifies the aggregation process of Graph Convolutional Networks by eliminating the feature transformation step and non-linear activation functions. It aggregates messages from neighboring nodes linearly, assigning equal weights to all edges.
- **User-kNN** [34] is a neighborhood-based model that computes recommendations by identifying the nearest neighbors of a target user and aggregating their preferences. Similarity between users is typically defined based on the overlap of items they have co-rated, and final predictions are derived from a weighted combination of neighbor ratings.
- **BPRMF** [33] is a pairwise learning-to-rank method based on matrix factorization.
- **EASE^R** [41] is a linear model conceptually resembling an autoencoder by computing a closed-form item-item similarity matrix.
- **MultiVAE** [25] harnesses a deep variational autoencoder to learn non-linear latent representations from entire user interactions.

We evaluate the quality of recommendations of the selected algorithms through nDCG [23]. Furthermore, we select two fairness-related metrics to assess the impact of data minimization on recommendation performance on consumer and provider fairness:

- **Mean Absolute Deviation (MAD)** [11]: it measures the disparity in recommendation quality across different user groups. Formally, let Q_u denote the quality of the recommendation for the user u and let $\mathcal{U}_x$ and $\mathcal{U}_y$ be two distinct groups of users. The mean absolute deviation (MAD) between these groups is defined as:

$$MAD\left(\mathbf{Q}^{(\mathcal{U}_x)}, \mathbf{Q}^{(\mathcal{U}_y)}\right) = \left| \frac{\sum_{x \in \mathcal{U}_x} Q_x}{|\mathcal{U}_x|} - \frac{\sum_{y \in \mathcal{U}_y} Q_y}{|\mathcal{U}_y|} \right|. \quad (1)$$

  When the users are partioned in more than two groups, MAD values are calculated for all possible pairwise group combinations. The final metric is derived as the mean of these computed MAD values. A smaller MAD value implies more equitable treatment across the groups, whereas a larger MAD reveals greater discrepancies and, thus, a higher degree of group-level unfairness. In our experiments, we use nDCG as the quality metric. In the ML1M dataset, users are partitioned by gender, while users are grouped by continent for the Ambar dataset.
- **Ranking-based Statistical Parity (RSP)** [46]: it measures how evenly different item categories are exposed to users in the top-k positions. Let us denote c_i as the item categories, and $P\big(R@k \mid c_i\big)$ the probability of a user encountering items in category c_i among the top-k recommendations ($R@k$). The RSP@k is defined as:

$$RSP@k = \frac{std\Big(P\big(R@k \mid c_1\big), \ldots, P\big(R@k \mid c_n\big)\Big)}{mean\Big(P\big(R@k \mid c_1\big), \ldots, P\big(R@k \mid c_n\big)\Big)} \quad (2)$$

  where $std(\cdot)$ and $mean(\cdot)$ denote the standard deviation and mean, respectively. A lower RSP indicates that the ranking probabilities are more evenly distributed, signifying less bias in how items are exposed to users. For ML1M, we categorize items into four release-year ranges. In the Ambar dataset, we group items by their continent of origin.

We also employ the **Hypervolume (HV)** [17] as a multi-objective metric to assess the model performance on accuracy, consumer, and provider fairness simultaneously. We choose the reference point $r = (1, 1.5, 1)$ for the three objectives: nDCG, RSP, and MAD.

3.4 Reproducibility Details

We adopt the all-unrated-items protocol [40], where the recommendable items are those not previously interacted with for each user. In our experiments, the embedding size is fixed at 64, and the batch size and the maximum number of epochs are 1024 and 300, respectively. Furthermore, we tune the model's hyper-parameters through grid search. For BPRMF and LightGCN, the learning rate and the regularization term are tuned in $\{0.01, 0.005, 0.001\}$. The number of layers in LightGCN is set to 2. EASE^R is tuned by varying its regularization factor in $\{500, 1000, 5000\}$. Moreover, for UserKNN, the neighborhood size is searched in $\{20, 30, 50\}$, while we use Cosine and Jaccard as similarity measures. Finally, for MultiVAE, the learning rate is tuned in $\{0.01, 0.001\}$ while the dimension of both the latent representation and the hidden layer is searched in $\{100, 200\}$. Combining all hyperparameter explorations for every model and strategy resulted in 2345 total trained models. We select the best models by adopting nDCG@10 as the validation metric, employing a patience value of 5 epochs for early stopping. Then, model performance results are computed on the test set. All the metrics are measured with a cutoff at 10, i.e., considering the top 10 suggested items to each user. To foster reproducibility, the models are trained using the reproducibility framework Elliot [5].

3.5 Data Minimization Strategies

In our study, we evaluate several approaches to minimize user data. These strategies are designed to select subsets of user-item interactions that serve as input to the system. In this paper, we study the data minimization strategies explored by Biega et al. [7], aiming at broadening their assessment to fairness issues. We briefly describe such strategies in the following.

Full data. This baseline utilizes the complete set of user interactions without minimization. This strategy serves as a reference point against which the effectiveness of all other strategies is measured. Formally, the system observes all interactions in the user history.

Random selection. This approach randomly selects n interactions from each user's history. Although randomness may introduce variability, the average profile characteristics of users are often preserved due to the sampling process.

Most recent. This strategy selects the n most recent user interactions by assuming that recent behaviors are more relevant to the recommendation task. The effectiveness of this method depends on the temporal data distribution. Long observation periods may lead to performance variation when compared to random selection. We do not apply this strategy to the Ambar dataset due to lack of timestamp information.

Most and least favorite. These techniques select the n items with the highest or lowest scores for users, respectively. While the former emphasizes user preferences, the latter explores how negative interactions influence recommendations.

Most rated. This method selects n items rated most frequently in the dataset for each user. This global popularity measure identifies items that are statistically significant within the system.

Most characteristic. It selects items based on their proximity to the average item profile. Each item is represented as a binary vector indicating which users have interacted with it. The average item profile is computed by averaging all these binary vectors in the system. We calculate the Euclidean distance for each item between its binary representation and this average vector. Items with lower distances are deemed more representative of the overall trends and are therefore considered "most characteristic." For each user, we select the n items they have interacted with closest to the system-wide profile, ensuring that the retained items best capture standard user behavior while avoiding niche or rarely rated items.

Highest variance. This strategy prioritizes items with the highest variability in their ratings across all users. Items with high variance are considered more informative as they capture diverse or polarizing opinions. The n items with the highest rating variance are selected for each user.

4 Results and Discussion

This section presents the empirical findings that address our three research questions.

4.1 Impact of Data Minimization (RQ1)

We answer RQ1 by discussing the impact of data minimization strategies on the accuracy, provider fairness, and consumer fairness of recommendation algorithms. Tables 1 and 2 report the model's performance measured with nDCG, RSP, MAD, and HV (see Section 3.3) for various minimization strategies and parameter n values for the ML1M and Ambar datasets, respectively. For each metric, we report the percentage differences in the performance between the considered strategy and the "full" strategy in parenthesis. Following Biega et al. [7], we say that minimization has a low impact on accuracy if the difference in nDCG is statistically insignificant or minimal. From the fairness perspective, minimization positively (negatively) impacts fairness if it strongly reduces (increases) the values of RSP and MAD, for which the lower, the better.

Impact on accuracy. Data minimization seriously alters the accuracy of recommendation algorithms. Reducing the parameter n affects recommendation accuracy, measured by nDCG, across datasets, models, and minimization strategies. This analysis corroborates the insights from Biega et al. [7], revealing that smaller n leads to a consistent drop in accuracy. Specifically, extreme minimization ($n = 1$) results in accuracy reductions that generally exceed 90% for many configurations, rendering recommendations impractical, as users cannot receive relevant suggestions. The accuracy performance loss is smaller but still noteworthy for moderate minimization ($n = 3$ and $n = 7$). At $n = 15$, dataset-specific differences emerge. In ML1M, the drop in accuracy remains pronounced, while the nDCG values become more acceptable for specific strategies in Ambar. Interestingly, the "Most Favorite" strategy even improves accuracy compared to the "full" strategy in some cases. At $n = 100$, accuracy differences become minimal or statistically insignificant in Ambar, demonstrating the feasibility of data minimization in this dataset under moderate conditions. In ML1M, performance differences remain measurable but less severe than smaller n values.

Impact on provider fairness. Data minimization introduces variability in provider fairness, as measured by RSP. Lower n values lead to inconsistent changes in RSP, with both improvements (negative changes compared to the "full" strategy) and deteriorations (positive changes) observed. This variability indicates a loss of control over provider fairness, particularly at smaller n, where fairness outcomes become unpredictable. Regarding the Ambar dataset, the relationship between accuracy and provider fairness is evident: strategies achieving higher nDCG often induce worse provider fairness. For example, the "Most Rated" strategy, which excels in accuracy, consistently increases RSP compared to the "full" strategy, indicating fairness degradation. In contrast, the "Most Characteristic" strategy sacrifices accuracy but consistently improves provider fairness, highlighting a trade-off between these two objectives.

Impact on consumer fairness. Data minimization generally enhances consumer fairness, as lower MAD values indicate. This improvement mainly stems from a reduction in overall nDCG, which levels the quality of recommendations across user groups to the detriment of overall recommendation relevance.

Impact of strategies. The interplay between data minimization and the objectives of accuracy, provider fairness, and consumer fairness is intricate, with varying impacts on recommendation performance. To unravel these complexities, we analyze the minimization strategies to identify patterns in their performance when $n \leq 15$. The "Most Rated" strategy, which emphasizes the presence of popular items in the training dataset, generally achieves the best or

Table 1: Performance results of all models on the ML1M dataset across different minimization strategies and values of n. The percentage difference between each value and the corresponding value obtained with the full strategy is indicated below each result. For each scenario, bold and underline stand for best and second-to-best values, respectively. All results are statistically significant based on paired t-tests ($p < 0.05$), except for the values denoted with †.

Strategy	$n=1$				$n=3$				$n=7$				$n=15$				$n=100$			
	nDCG↑	RSP↓	MAD↓	HV↑	nDCG↑	RSP↓	MAD↓	HV↑	nDCG↑	RSP↓	MAD↓	HV↑	nDCG↑	RSP↓	MAD↓	HV↑	nDCG↑	RSP↓	MAD↓	HV↑
UserKNN																				
Random	0.0405	0.7432	0.0028	0.7852	0.0497	0.2577	0.009	1.2922	0.0711	0.379	0.0155	1.182	0.0977	0.5374	0.0306	1.0243	0.2351	0.4241	0.023	1.2982
	(-87.52)	(73.69)	(-93.62)	(-21.48)	(-84.7)	(-39.77)	(-79.76)	(29.22)	(-78.09)	(-11.43)	(-65.0)	(18.2)	(-69.89)	(25.59)	(-31.09)	(2.43)	(-27.55)	(-0.88)	(-48.17)	(29.82)
Most Recent	0.027	0.9834	0.0052	0.5277	0.0443	0.573	0.0061	0.9622	0.0736	0.4608	0.0046	1.1106	0.1165	0.4273	**0.0002**	1.1975	0.2556	0.4617	0.0216	1.2756
	(-91.68)	(129.83)	(-88.26)	(-47.23)	(-86.34)	(33.91)	(-86.32)	(-3.78)	(-77.3)	(7.68)	(-89.57)	(11.06)	(-64.11)	(-0.15)	(-99.64)	(19.75)	(-21.21)	(7.89)	(-51.36)	(27.56)
Most Favorite	0.0509	**0.6137**	0.0098	**0.9223**	0.0713	0.4552	0.0105	1.1075	0.0984	0.4927	0.0165	1.0882	0.1395	0.5473	0.0192	1.0647	0.2838	0.4362	0.0508	1.2964
	(-84.31)	(43.43)	(-78.0)	(-7.77)	(-78.02)	(6.38)	(-76.33)	(10.75)	(-69.67)	(15.14)	(-62.94)	(8.82)	(-57.01)	(27.9)	(-56.69)	(6.47)	(-12.53)	(1.93)	(14.34)	(29.64)
Least Favorite	0.0442	1.0051	0.0054	0.514	0.0336	0.6857	0.0041	0.8382	0.053	0.3819	**0.0022**	1.1748	0.0782	0.6994	0.0047	0.8591	0.2248	0.4354	0.0248	1.2716
	(-86.38)	(134.9)	(-87.91)	(-48.6)	(-89.65)	(60.25)	(-90.73)	(-16.18)	(-83.66)	(-10.75)	(-95.05)	(17.48)	(-75.9)	(63.45)	(-89.33)	(-14.09)	(-30.73)	(1.74)	(-44.09)	(27.16)
Most Rated	**0.0921**	1.0067	0.0159	0.5302	**0.1447**	1.0039	0.0157	0.559	**0.2069**	0.9652	0.0297	0.6262	**0.2518**	0.7458	0.0491	0.8978	**0.3105**	0.3669	0.0403	**1.425**
	(-71.61)	(135.26)	(-64.2)	(-46.98)	(-55.4)	(134.61)	(-64.54)	(-44.1)	(-36.23)	(125.57)	(-33.02)	(-37.38)	(-22.38)	(74.3)	(10.58)	(-10.22)	(-4.31)	(-14.25)	(-9.15)	(42.5)
Most Characteristic	0.0025	0.9607	0.0008	0.5402	0.0049	0.4391	**0.0005**	1.0656	0.011	**0.1799**	0.0024	**1.3313**	0.033	**0.286**	0.0048	**1.2481**	0.2122	**0.3446**	**0.0178**	1.3757
	(-99.22)	(124.51)	(-98.12)	(-45.98)	(-98.48)	(2.61)	(-98.87)	(6.56)	(-96.62)	(-57.95)	(-94.54)	(33.13)	(-89.82)	(-33.16)	(-89.23)	(24.81)	(-34.59)	(-19.47)	(-59.9)	(37.57)
Highest Variance	0.0152	0.8777	**0.0004**	0.6315	0.0381	**0.1019**	0.004	**1.4455**	0.0687	0.3048	0.0102	1.2643	0.1198	0.4355	0.01	1.1801	0.2756	0.4153	0.0251	1.3489
	(-95.32)	(105.12)	(-99.03)	(-36.85)	(-88.27)	(-76.19)	(-90.91)	(44.55)	(-78.82)	(-28.77)	(-77.08)	(26.43)	(-63.07)	(1.78)	(-77.5)	(18.01)	(-15.06)	(-2.94)	(-43.52)	(34.89)
BPRMF																				
Random	0.0358	0.3937	0.0043	1.1409	0.0453	0.4242	0.009	1.1144	0.0538	0.4735	0.0108	1.0701	0.0923	0.6841	0.013	0.8796	0.2329	0.4274	0.0161	1.3011
	(-88.82)	(-30.35)	(-82.55)	(14.09)	(-85.85)	(-24.95)	(-63.94)	(11.44)	(-83.18)	(-16.23)	(-56.69)	(7.01)	(-71.16)	(21.02)	(-47.48)	(-12.04)	(-27.25)	(-24.38)	(-35.35)	(30.11)
Most Recent	0.0275	0.456	0.0003	1.0723	0.0388	0.5041	0.0122	1.0219	0.0471	0.448	0.0071	1.0937	0.0924	0.5417	0.0115	1.0348	0.2396	0.4838	0.0012	1.2582
	(-91.4)	(-19.32)	(-98.72)	(7.23)	(-87.87)	(-10.82)	(-50.97)	(2.19)	(-85.3)	(-20.75)	(-71.31)	(9.37)	(-71.15)	(-4.17)	(-53.72)	(3.48)	(-25.14)	(-14.41)	(-95.35)	(25.82)
Most Favorite	0.0348	0.2236	0.0023	1.3178	0.0588	0.3714	0.011	1.1818	0.0889	0.5047	0.0288	1.0526	0.1215	0.6164	0.0226	0.9685	0.2823	0.4733	0.0307	1.2762
	(-89.13)	(-60.45)	(-90.81)	(31.78)	(-81.63)	(-34.29)	(-55.82)	(18.18)	(-72.24)	(-10.71)	(15.71)	(5.26)	(-62.04)	(9.05)	(-8.95)	(-3.15)	(-11.79)	(-16.27)	(23.54)	(27.62)
Least Favorite	0.025	0.5616	0.0041	0.958	0.021	0.4908	0.0004	1.03	0.0392	0.7105	0.0047	0.8165	0.0587	0.9444	0.0144	0.5798	0.225	0.6161	0.0053	1.077
	(-92.18)	(-0.64)	(-83.66)	(-4.2)	(-93.43)	(-13.18)	(-98.19)	(3.0)	(-87.76)	(25.7)	(-80.91)	(-18.35)	(-81.66)	(67.07)	(-42.24)	(-42.02)	(-29.69)	(9.0)	(-78.6)	(7.7)
Most Rated	**0.116**	0.7281	0.0247	0.8402	**0.1504**	1.0669	0.0374	0.4796	**0.1923**	1.0935	0.0443	0.4632	**0.2443**	0.865	0.0496	0.751	**0.299**	0.5114	0.0421	1.2302
	(-63.76)	(28.8)	(-0.47)	(-15.98)	(-53.01)	(88.75)	(50.58)	(-52.04)	(-39.92)	(93.45)	(78.26)	(-53.68)	(-23.66)	(53.02)	(99.48)	(-24.9)	(-6.58)	(-9.53)	(69.27)	(23.02)
Most Characteristic	0.0021	**0.1067**	**0.0001**	**1.396**	0.0033	**0.1248**	**0.0001**	1.3796	0.0067	**0.0427**	**0.0019**	**1.4642**	0.0171	**0.0989**	**0.0038**	**1.4197**	0.1957	**0.3798**	**0.0001**	**1.3393**
	(-99.36)	(-81.12)	(-99.67)	(39.6)	(-98.98)	(-77.93)	(-99.54)	(37.96)	(-97.9)	(-92.44)	(-92.26)	(46.42)	(-94.65)	(-82.51)	(-84.58)	(41.97)	(-38.84)	(-32.81)	(-99.53)	(33.93)
Highest Variance	0.011	0.1258	0.0023	1.3861	0.025	0.1452	0.0037	**1.3835**	0.0628	0.7276	0.0097	0.8129	0.1004	0.6421	0.0165	0.9284	0.2648	0.5155	0.0066	1.237
	(-96.56)	(-77.74)	(-90.68)	(38.61)	(-92.19)	(-74.31)	(-85.28)	(38.35)	(-80.38)	(28.72)	(-61.0)	(-18.71)	(-68.65)	(13.59)	(-33.52)	(-7.16)	(-17.27)	(-8.8)	(-73.51)	(23.7)
EASE^R																				
Random	0.0377	0.77	0.0011	0.7567	0.0495	0.2089	0.01	1.3415	0.0742	0.3458	0.0129	1.2239	0.115	0.7393	0.014	0.8363	0.2539	0.5541	0.0324	1.1477
	(-88.79)	(31.47)	(-97.93)	(-24.33)	(-85.28)	(-64.33)	(-80.69)	(34.15)	(-77.96)	(-40.96)	(-75.15)	(22.39)	(-65.83)	(26.22)	(-72.95)	(-16.37)	(-24.55)	(-5.41)	(-37.36)	(14.77)
Most Recent	0.0268	1.0037	0.0049	0.5072	0.0453	0.4955	0.005	1.0447	0.0718	0.4212	0.008	1.147	0.1246	0.4998	**0.001**	1.1237	0.2719	0.5382	0.0276	1.1894
	(-92.03)	(71.36)	(-90.61)	(-49.28)	(-86.55)	(-15.4)	(-90.27)	(4.47)	(-78.67)	(-28.09)	(-84.54)	(14.7)	(-62.98)	(-14.67)	(-98.07)	(12.37)	(-19.19)	(-8.11)	(-46.57)	(18.94)
Most Favorite	0.0491	**0.6675**	0.0091	**0.8654**	0.0676	0.425	0.0045	1.1424	0.115	0.6754	0.0178	0.903	0.1482	0.7227	0.0303	0.8655	0.2969	0.5889	0.0491	1.1236
	(-85.41)	(13.96)	(-82.37)	(-13.46)	(-79.91)	(-27.43)	(-91.23)	(14.24)	(-65.82)	(15.32)	(-65.53)	(-9.7)	(-55.96)	(23.39)	(-41.51)	(-13.45)	(-11.75)	(0.55)	(-5.06)	(12.36)
Least Favorite	0.0454	1.0169	0.0044	0.5028	0.034	0.6346	**0.0005**	0.8944	0.0498	0.2519	0.0069	**1.3013**	0.0913	0.7488	0.0093	0.8122	0.2418	0.5392	0.0244	1.164
	(-86.51)	(73.61)	(-91.47)	(-49.72)	(-89.88)	(8.35)	(-99.01)	(-10.56)	(-85.19)	(-57.0)	(-86.7)	(30.13)	(-72.86)	(27.84)	(-82.06)	(-18.78)	(-28.13)	(-7.94)	(-52.87)	(16.4)
Most Rated	**0.0898**	1.0172	0.0162	0.5177	**0.16**	0.9405	0.0264	0.6318	**0.209**	1.1198	0.0405	0.441	**0.257**	0.9422	0.0494	0.6666	**0.325**	0.6228	0.0654	1.0862
	(-73.3)	(73.67)	(-68.76)	(-48.23)	(-52.43)	(60.58)	(-48.89)	(-36.82)	(-37.88)	(91.19)	(-21.68)	(-55.9)	(-23.62)	(60.86)	(-4.55)	(-33.34)	(-3.4)	(6.33)	(26.49)	(8.62)
Most Characteristic	0.0026	1.0692	0.0009	0.4315	0.0044	**0.1381**	0.0007	1.3669	0.0115	**0.2157**	**0.0017**	1.2968	0.0315	**0.321**	0.4998	**1.2138**	0.2253	**0.3865**	**0.0204**	**1.3365**
	(-99.22)	(82.55)	(-98.19)	(-56.85)	(-98.71)	(-76.42)	(-98.7)	(36.69)	(-96.59)	(-63.17)	(-96.72)	(29.68)	(-90.65)	(-45.2)	(-96.36)	(21.38)	(-33.04)	(-34.01)	(-60.51)	(33.65)
Highest Variance	0.0138	1.0179	**0.0003**	0.4886	0.0414	0.1476	0.001	**1.4069**	0.0792	0.5182	0.0083	1.0508	0.1174	0.7187	0.0078	0.8663	0.2932	0.5941	0.0321	1.1339
	(-95.9)	(73.79)	(-99.45)	(-51.14)	(-87.7)	(-74.8)	(-98.0)	(40.69)	(-76.46)	(-11.53)	(-83.91)	(5.08)	(-65.11)	(22.7)	(-85.0)	(-13.37)	(-12.84)	(1.43)	(-37.88)	(13.39)
MultiVAE																				
Random	0.0559	0.4359	0.0183	1.1029	0.0943	0.8057	0.02	0.7446	0.0987	0.9279	0.0193	0.6164	0.1028	0.933	0.0093	0.6195	0.1513	1.0855	0.0286	0.4636
	(-80.5)	(32.32)	(-42.06)	(10.29)	(-67.12)	(144.55)	(-36.94)	(-25.54)	(-65.57)	(181.64)	(-38.98)	(-38.36)	(-64.13)	(183.18)	(-70.48)	(-38.05)	(-47.24)	(229.47)	(-9.83)	(-53.64)
Most Recent	0.0408	0.5061	0.01	1.0242	0.0608	0.8357	0.0032	0.7025	0.0771	0.9861	**0.0023**	0.5522	0.086	0.9924	0.0228	0.5387	0.207	**0.2829**	0.0015	**1.4667**
	(-85.75)	(53.6)	(-68.54)	(2.42)	(-78.78)	(153.66)	(-90.05)	(-29.75)	(-73.11)	(199.32)	(-92.69)	(-44.78)	(-70.01)	(201.21)	(-28.15)	(-46.13)	(-27.82)	(-14.14)	(-95.14)	(46.67)
Most Favorite	0.0797	0.9686	0.0324	0.5552	0.0948	0.8103	0.026	0.7355	0.108	1.0646	0.0336	0.4662	0.1312	0.3815	0.0241	1.2347	0.2225	0.4687	0.046	1.2027
	(-72.19)	(193.99)	(2.42)	(-44.48)	(-66.95)	(145.94)	(-17.95)	(-26.45)	(-62.32)	(223.14)	(5.95)	(-53.38)	(-54.24)	(15.8)	(-23.8)	(23.47)	(-22.4)	(42.26)	(45.27)	(20.27)
Least Favorite	0.0366	1.0074	**0.0005**	0.5104	0.0549	1.4399	0.0101	0.0627	0.0521	1.0557	0.0093	0.4631	0.0809	1.0609	0.0143	0.4678	0.2036	0.3799	0.0015	1.3462
	(-87.24)	(205.77)	(-98.36)	(-48.96)	(-80.86)	(337.06)	(-68.0)	(-93.73)	(-81.83)	(220.43)	(-70.56)	(-53.69)	(-71.77)	(222.02)	(-54.83)	(-53.22)	(-28.97)	(15.3)	(-95.19)	(34.62)
Most Rated	**0.1211**	0.8012	0.029	0.7607	**0.152**	1.1149	0.0342	0.4285	**0.1987**	1.1363	0.0496	0.4143	**0.2553**	0.7976	0.0576	0.8309	0.2021	1.094	0.0513	0.463
	(-57.76)	(143.18)	(-8.36)	(-23.93)	(-46.99)	(238.39)	(7.86)	(-57.15)	(-30.7)	(244.91)	(56.5)	(-58.57)	(-10.97)	(142.08)	(81.89)	(-16.91)	(-29.5)	(232.07)	(62.13)	(-53.7)
Most Characteristic	0.0041	**0.0942**	0.0048	**1.4048**	0.0026	**0.0305**	**0.0005**	**1.4726**	0.0058	**0.0781**	0.0025	**1.4265**	0.0147	**0.1887**	**0.0029**	**1.3268**	0.1884	0.3015	**0.0002**	1.424
	(-98.57)	(-71.41)	(-84.93)	(40.48)	(-99.1)	(-90.75)	(-98.52)	(47.26)	(-97.99)	(-76.3)	(-92.06)	(42.65)	(-94.86)	(-42.73)	(-90.9)	(32.68)	(-34.3)	(-8.49)	(-99.38)	(42.4)
Highest Variance	0.0268	0.4372	0.0014	1.0898	0.0462	0.5804	0.0021	0.96	0.0615	1.0367	0.0039	0.4899	0.1054	0.3263	0.0141	1.279	**0.2468**	0.3662	0.0127	1.3957
	(-90.66)	(32.7)	(-95.68)	(8.98)	(-83.87)	(76.17)	(-93.23)	(-4.0)	(-78.54)	(214.67)	(-87.73)	(-51.01)	(-63.25)	(-0.95)	(-55.36)	(27.9)	(-13.9)	(11.14)	(-59.85)	(39.57)
LightGCN																				
Random	0.0437	0.306	0.0034	1.242	0.0436	0.2697	0.0038	1.2791	0.0687	0.4756	0.0127	1.0809	0.1183	0.8726	0.0161	0.6903	0.2439	0.3928	0.0226	1.3461
	(-87.0)	(-24.73)	(-90.65)	(24.2)	(-87.01)	(-33.65)	(-89.37)	(27.91)	(-79.56)	(16.98)	(-64.51)	(8.09)	(-64.79)	(114.66)	(-54.96)	(-30.97)	(-27.38)	(-3.36)	(-36.99)	(34.61)
Most Recent	0.032	0.5049	0.0077	1.019	0.0427	0.4607	0.0093	1.0736	0.0641	0.4211	0.0031	1.1444	0.1183	0.4871	0.0035	1.1288	0.2406	0.4163	**0.0003**	1.344
	(-90.48)	(24.2)	(-78.41)	(1.9)	(-87.28)	(13.34)	(-74.02)	(7.36)	(-80.93)	(3.59)	(-91.22)	(14.44)	(-64.79)	(19.83)	(-90.37)	(12.88)	(-28.39)	(2.41)	(-99.13)	(34.4)
Most Favorite	0.0536	0.411	0.009	1.1371	0.0617	0.2915	0.0108	1.2692	0.112	0.5898	0.0137	0.9982	0.1625	0.5028	0.0255	1.1297	0.2905	0.4596	0.0365	1.2935
	(-84.05)	(1.1)	(-74.95)	(13.71)	(-81.63)	(-28.29)	(-69.97)	(26.92)	(-66.67)	(45.1)	(-61.73)	(-0.18)	(-51.63)	(23.68)	(-28.82)	(12.97)	(-13.53)	(13.07)	(1.9)	(29.35)
Least Favorite	0.0318	0.6495	0.0015	0.8762	0.0292	0.6542	0.0028	0.868	0.0406	0.6807	0.0033	0.8498	0.084	0.8337	0.0149	0.7115	0.2403	0.4117	0.0071	1.3402
	(-90.55)	(59.78)	(-95.76)	(-12.38)	(-91.31)	(60.94)	(-92.12)	(-13.2)	(-87.92)	(67.45)	(-90.81)	(-15.02)	(-75.0)	(105.08)	(-58.38)	(-28.85)	(-28.46)	(1.29)	(-80.06)	(34.02)
Most Rated	**0.1094**	0.5996	0.0146	0.9844	**0.1524**	1.0152	0.0296	0.5422	**0.184**	1.1128	0.0467	0.4371	**0.2627**	0.8014	0.0479	0.84	**0.3179**	0.5199	0.0373	1.2435
	(-67.42)	(47.49)	(-59.38)	(-1.56)	(-54.62)	(149.73)	(-17.56)	(-45.78)	(-45.23)	(173.74)	(30.21)	(-56.29)	(-21.79)	(97.13)	(33.53)	(-16.0)	(-5.36)	(27.89)	(4.1)	(24.35)
Most Characteristic	0.0037	**0.0842**	**0.0011**	**1.4195**	0.0048	**0.0837**	**0.0019**	**1.4203**	0.0115	**0.0927**	**0.0018**	**1.421**	0.0304	**0.1471**	**0.002**	**1.3911**	0.2145	**0.3243**	0.0048	**1.421**
	(-98.9)	(-79.3)	(-96.93)	(41.95)	(-98.58)	(-79.41)	(-94.64)	(42.03)	(-96.57)	(-77.2)	(-95.11)	(42.1)	(-90.95)	(-63.8)	(-94.29)	(39.11)	(-36.15)	(-20.23)	(-86.55)	(42.1)
Highest Variance	0.0147	0.1839	0.0041	1.33	0.0374	0.3612	0.0049	1.1757	0.0729	0.3729	0.0069	1.2009	0.1185	0.4996	0.0148	1.1024	0.2864	0.4566	**0.0003**	1.3417
	(-95.62)	(-54.75)	(-88.58)	(33.0)	(-88.86)	(-11.14)	(-86.46)	(17.57)	(-78.3)	(-8.28)	(-80.68)	(20.09)	(-64.73)	(22.9)	(-58.66)	(10.24)	(-14.75)	(12.33)	(-99.09)	(34.17)

Table 2: Performance results of all models on the Ambar dataset across different minimization strategies and values of n. The percentage difference between each value and the corresponding value obtained with the full strategy is indicated below each result. For each scenario, bold and underline stand for best and second-to-best values, respectively. All results are statistically significant based on paired t-tests ($p < 0.05$), except for the values denoted with †.

Strategy	$n=1$				$n=3$				$n=7$				$n=15$				$n=100$			
	nDCG↑	RSP↓	MAD↓	HV↑	nDCG↑	RSP↓	MAD↓	HV↑	nDCG↑	RSP↓	MAD↓	HV↑	nDCG↑	RSP↓	MAD↓	HV↑	nDCG↑	RSP↓	MAD↓	HV↑
UserKNN																				
Random	0.0036 (-97.26)	1.4422 (75.43)	0.0049 (-91.79)	0.0577 (-94.23)	0.0243 (-81.53)	0.9005 (9.54)	0.0756 (25.48)	0.5676 (-43.24)	0.0572 (-56.55)	0.7075 (-13.95)	0.0324 (-46.17)	0.8107 (-18.93)	0.1147 (-12.88)	0.8074 (-1.8)	**0.0463** (-23.21)	0.7363 (-26.37)	0.1373† (4.29)	0.8199 (-0.27)	0.062 (2.93)	0.7255 (-27.45)
Most Favorite	0.0101 (-92.35)	1.4621 (77.84)	0.0044 (-92.72)	0.0381 (-96.19)	0.049 (-62.78)	0.8691 (5.71)	0.0241 (-60.03)	0.6459 (-35.41)	**0.1092** (-17.03)	0.7307 (-11.12)	0.0361 (-40.13)	0.8225 (-17.75)	**0.1498** (13.77)	0.8174 (-0.57)	0.0483 (-19.8)	0.7469 (-25.31)	0.1404† (6.64)	0.8755 (6.5)	0.06 (-0.42)	0.6694 (-33.06)
Least Favorite	0.0055 (-95.84)	1.4863 (80.79)	0.0354 (-41.26)	0.0133 (-98.67)	0.0115 (-91.3)	0.9036 (9.91)	0.0053 (-91.19)	0.6 (-40.0)	0.0345 (-73.81)	0.5747 (-30.1)	0.0153 (-74.6)	0.9426 (-5.74)	0.0806 (-38.75)	0.9176 (11.61)	0.1341 (122.41)	0.545 (-45.5)	0.1362† (3.48)	**0.8072** (-1.82)	0.0683 (13.31)	**0.7335** (-26.65)
Most Rated	**0.0164** (-87.55)	1.447 (76.0)	0.0101 (-83.21)	0.0534 (-94.66)	**0.0769** (-41.56)	1.1329 (37.8)	0.0644 (6.79)	0.3699 (-63.01)	0.0938 (-28.74)	1.0878 (32.31)	0.0525 (-12.91)	0.4272 (-57.28)	0.1103 (-16.2)	0.9898 (20.39)	0.0524 (-13.05)	0.5369 (-46.31)	0.1311† (-0.4)	0.8177 (-0.53)	**0.0594** (-1.54)	0.7259 (-27.41)
Most Characteristic	0.0016 (-98.75)	1.5222 (85.16)	**0.0005** (-99.18)	0.0 (-100.0)	0.01 (-92.44)	1.1757 (43.01)	**0.0041** (-93.14)	0.3262 (-67.38)	0.0285 (-78.37)	**0.491** (-40.28)	**0.0077** (-87.17)	**1.0297** (2.97)	0.0407 (-69.12)	**0.6132** (-25.41)	0.0504 (-16.35)	**0.8763** (-12.37)	**0.1437**† (9.13)	0.8548 (3.97)	0.0707 (17.29)	0.6858 (-31.42)
Highest Variance	0.0015 (-98.83)	**0.9225** (12.21)	0.0007 (-98.84)	**0.578** (-42.2)	0.0237 (-82.02)	**0.6099** (-25.81)	0.0121 (-79.92)	**0.9001** (-9.99)	0.062 (-52.9)	0.6658 (-19.01)	0.0396 (-34.26)	0.8508 (-14.92)	0.0976 (-25.85)	0.7875 (-4.21)	0.0553 (-8.25)	0.7388 (-26.12)	0.1318† (0.09)	0.8138 (-1.01)	0.0598 (-0.72)	0.7301 (-26.99)
BPRMF																				
Random	0.0064 (-94.64)	0.7976 (-1.26)	0.0091 (-78.83)	0.7005 (-29.95)	0.004 (-96.61)	0.6976 (-13.63)	0.0022 (-94.82)	0.8038 (-19.62)	0.0329 (-72.3)	0.7796 (-3.49)	0.0168 (-61.05)	0.7316 (-26.84)	0.0884 (-25.52)	0.7978 (-1.22)	0.0484 (12.26)	0.7272 (-27.28)	0.1169† (-1.52)	0.8874 (9.87)	0.0476 (10.4)	0.6516 (-34.84)
Most Favorite	0.0089 (-92.53)	0.5577 (-30.96)	0.0099 (-77.09)	0.9413 (-5.87)	0.0324 (-72.69)	0.6987 (-13.49)	0.0179 (-58.62)	0.8125 (-18.75)	0.0798 (-32.8)	0.6438 (-20.29)	0.0266 (-38.33)	0.8999 (-10.01)	**0.1227** (3.34)	0.7991 (-1.07)	0.0451 (4.57)	0.7514 (-24.86)	**0.1235**† (3.99)	0.9042 (11.94)	0.0442 (2.56)	0.6398 (-36.02)
Least Favorite	0.0062 (-94.78)	0.5746 (-28.86)	0.0034 (-92.04)	0.9279 (-7.21)	0.0098 (-91.73)	0.7216 (-10.66)	0.0041 (-90.49)	0.7828 (-21.72)	0.0157 (-86.74)	0.7214 (-10.69)	**0.0051** (-88.22)	0.7869 (-21.31)	0.0657 (-44.66)	0.7916 (-1.99)	0.125 (189.85)	0.6605 (-33.95)	0.118† (-0.62)	0.8898 (10.17)	0.0485 (12.51)	0.6491 (-35.09)
Most Rated	**0.0245** (-79.38)	1.2525 (55.06)	0.01 (-76.84)	0.2511 (-74.89)	**0.0744** (-37.33)	1.1952 (47.97)	0.0634 (46.92)	0.3067 (-69.33)	**0.0917** (-22.79)	1.1311 (40.04)	0.0459 (6.35)	0.3842 (-61.58)	0.0957 (-19.41)	1.0444 (29.31)	0.0374 (-13.31)	0.4805 (-51.95)	0.1064 (-10.36)	0.8685 (7.53)	**0.0398** (-7.73)	0.6709 (-32.91)
Most Characteristic	0.0008 (-99.35)	**0.3023** (-62.58)	**0.0003** (-99.29)	**1.1983** (19.83)	0.0039 (-96.74)	**0.5627** (-30.34)	**0.0017** (-96.03)	**0.9393** (-6.07)	0.0087 (-92.71)	**0.413** (-48.87)	0.008 (-81.56)	**1.0877** (8.77)	0.034 (-71.34)	**0.5905** (-26.9)	**0.012** (-72.24)	**0.9292** (-7.08)	0.1192 (0.35)	0.8914 (10.37)	0.0439 (1.7)	0.6512 (-34.88)
Highest Variance	0.0025 (-97.89)	0.5701 (-29.42)	0.0012 (-97.13)	0.9311 (-6.89)	0.0046 (-96.11)	0.6246 (-22.67)	**0.0017** (-96.09)	0.878 (-12.2)	0.0437 (-63.22)	0.6134 (-24.06)	0.0377 (-12.58)	0.8904 (-10.96)	0.0809 (-31.85)	0.8204 (1.58)	0.0451 (4.44)	0.7015 (-29.85)	0.1177† (-0.91)	**0.8217** (1.73)	0.0441 (2.34)	**0.7247** (-27.53)
EASE^R																				
Random	0.0033 (-97.75)	1.4402 (64.81)	0.0049 (-92.29)	0.0597 (-94.03)	0.0255 (-82.42)	0.8246 (-5.63)	0.077 (20.7)	0.6393 (-36.07)	0.0583 (-59.83)	0.7536 (-13.76)	0.0469 (-26.4)	0.7529 (-24.71)	0.1119 (-22.89)	0.8115 (-7.13)	**0.0449** (-29.64)	0.7312 (-26.88)	0.149 (2.71)	0.8715 (-0.26)	0.0646 (1.3)	0.6755 (-32.45)
Most Favorite	0.0101 (-93.06)	1.4621 (67.32)	0.0044 (-93.12)	0.0381 (-96.19)	0.0543 (-62.61)	0.8143 (-6.82)	0.0349 (-45.32)	0.6977 (-30.23)	**0.1058** (-27.06)	0.7562 (-13.46)	0.0338 (-47.0)	0.7947 (-20.53)	**0.1508** (3.95)	0.8305 (-4.96)	0.0477 (-25.27)	0.7338 (-26.62)	0.1507† (3.85)	0.8769 (0.35)	0.0646 (1.33)	0.6706 (-32.94)
Least Favorite	0.0049 (-96.6)	1.5398 (76.21)	0.0018 (-97.12)	0.0 (-100.0)	0.0128 (-91.17)	0.7903 (-9.56)	0.0426 (-33.24)	0.6882 (-31.18)	0.0363 (-74.95)	0.6153 (-29.59)	0.0787 (23.34)	0.8447 (-15.53)	0.0709 (-51.15)	0.9006 (3.06)	0.1432 (124.6)	0.55 (-45.0)	0.15† (3.36)	0.8504 (-2.68)	0.0717 (12.37)	0.6935 (-30.65)
Most Rated	**0.0164** (-88.7)	1.447 (65.59)	0.0101 (-84.13)	0.0534 (-94.66)	**0.0741** (-48.93)	1.1241 (28.64)	0.055 (-13.73)	0.3816 (-61.84)	0.0937 (-35.41)	1.0882 (24.53)	0.0513 (-19.52)	0.4273 (-57.27)	0.1088 (-25.03)	1.0503 (20.2)	0.0561 (-12.05)	0.4706 (-52.94)	0.1436 (-1.06)	0.8698 (-0.47)	**0.063** (-1.15)	0.6753 (-32.47)
Most Characteristic	0.0016 (-98.87)	1.5222 (74.2)	**0.0005** (-99.22)	0.0 (-100.0)	0.01 (-93.14)	0.6672 (-23.64)	**0.004** (-93.71)	0.8377 (-16.23)	0.0265 (-81.77)	**0.4624** (-47.08)	**0.0072** (-88.75)	**1.0574** (5.74)	0.0387 (-73.35)	**0.5933** (-32.11)	0.0597 (-6.39)	**0.8856** (-11.44)	**0.152**† (4.77)	**0.8489** (-2.86)	0.0733 (14.93)	**0.6952** (-30.48)
Highest Variance	0.0008 (-99.44)	**0.9525** (9.01)	**0.0005** (-99.22)	**0.5476** (-45.24)	0.0226 (-84.45)	**0.5795** (-33.69)	0.0111 (-82.56)	**0.9309** (-6.91)	0.0602 (-58.52)	0.7098 (-18.77)	0.0386 (-39.5)	0.8054 (-19.46)	0.0989 (-31.81)	0.7882 (-9.8)	0.0567 (-11.05)	0.7378 (-26.22)	0.1445 (-0.38)	0.8624 (-1.31)	0.0633 (-0.76)	0.6836 (-31.64)
MultiVAE																				
Random	0.0116 (-88.45)	0.8987 (3.88)	0.0094 (-78.09)	0.6026 (-39.74)	0.0192 (-80.87)	0.8336 (-3.64)	0.0079 (-81.63)	0.6739 (-32.61)	0.0284 (-71.79)	0.9992 (15.5)	0.0175 (-59.08)	0.506 (-49.4)	0.0676 (-32.78)	0.8202 (-5.19)	0.0471 (10.01)	0.6915 (-30.85)	0.101† (0.48)	0.8094 (-6.44)	0.044 (2.88)	0.7269 (-27.31)
Most Favorite	0.0198 (-80.31)	0.8718 (0.77)	0.0262 (-38.81)	0.6239 (-37.61)	0.0419 (-58.31)	0.6107 (-29.41)	0.0192 (-55.24)	**0.9088** (-9.12)	0.073 (-27.36)	0.7678 (-11.25)	0.0244 (-43.08)	0.7666 (-23.34)	**0.1015** (0.92)	0.7589 (-12.28)	0.032 (-25.2)	0.7902 (-20.98)	**0.1056**† (5.07)	0.7931 (-8.32)	0.0455 (6.4)	0.746 (-25.4)
Least Favorite	0.0085 (-91.52)	0.5952 (-31.21)	0.0033 (-92.28)	0.9095 (-9.05)	0.0096 (-90.42)	0.9784 (13.09)	0.0523 (22.25)	0.4991 (-50.09)	0.0167 (-83.43)	0.9492 (9.72)	0.0263 (-38.67)	0.5453 (-45.47)	0.042 (-58.21)	0.8537 (-1.32)	0.0614 (43.4)	0.6321 (-36.79)	0.1031† (2.54)	0.789 (-8.8)	0.0517 (20.68)	0.7438 (-25.62)
Most Rated	**0.0212** (-78.94)	1.2257 (41.68)	0.0158 (-63.14)	0.2757 (-72.43)	**0.0723** (-28.09)	1.1437 (32.2)	0.0713 (66.69)	0.3548 (-64.52)	**0.087** (-13.43)	1.2147 (40.4)	0.0462 (8.03)	0.2958 (-70.42)	0.1009 (0.38)	0.9789 (13.15)	0.0403 (-5.9)	0.5506 (-44.94)	0.1016† (1.1)	0.8352 (-3.46)	0.0436 (1.78)	0.7005 (-29.95)
Most Characteristic	0.0018 (-98.24)	**0.4404** (-49.1)	**0.0006** (-98.54)	**1.0608** (6.08)	0.0006 (-99.36)	**0.5933** (-31.42)	**0.0006** (-98.49)	0.9067 (-9.33)	0.014 (-86.11)	**0.389** (-55.03)	**0.013** (-69.63)	**1.1119** (11.19)	0.0242 (-75.89)	**0.5258** (-39.22)	**0.0128** (-70.08)	**0.9851** (-1.49)	0.1023† (1.78)	0.813 (-6.03)	**0.0331** (-22.59)	0.7323 (-26.77)
Highest Variance	0.0029 (-97.16)	0.5896 (-31.85)	0.0016 (-96.31)	0.9115 (-8.85)	0.0241 (-76.0)	0.6378 (-26.27)	0.0114 (-73.47)	0.873 (-12.7)	0.0492 (-51.02)	0.6604 (-23.67)	0.0271 (-36.64)	0.8571 (-14.29)	0.0787 (-21.69)	0.637 (-26.37)	0.0388 (-9.39)	0.8948 (-10.52)	0.101† (0.43)	**0.7737** (-10.57)	0.0478 (11.57)	**0.7614** (-23.86)
LightGCN																				
Random	0.0083 (-93.95)	0.7742 (-9.77)	0.0302 (-46.03)	0.7097 (-29.03)	0.0233 (-82.92)	0.796 (-7.24)	0.1292 (130.96)	0.6273 (-37.27)	0.053 (-61.18)	0.7235 (-15.69)	0.0525 (-6.11)	0.7747 (-22.53)	0.1171 (-14.27)	0.9015 (5.06)	**0.0281** (-49.79)	0.6498 (-35.02)	0.1291 (-5.46)	**0.7887** (-8.09)	0.0532 (-4.85)	**0.7604** (-23.96)
Most Favorite	0.0108 (-92.1)	0.5157 (-39.9)	0.0058 (-89.55)	0.9891 (-1.09)	0.0497 (-63.59)	0.7441 (-13.28)	0.0266 (-52.5)	0.7724 (-22.76)	**0.1078** (-21.1)	0.6993 (-18.5)	**0.0231** (-58.74)	0.8665 (-13.35)	**0.1488** (8.97)	0.8195 (-4.5)	0.0443 (-20.8)	0.7471 (-25.29)	0.1287 (-5.73)	0.8545 (-0.42)	**0.0496** (-11.31)	0.6924 (-30.76)
Least Favorite	0.008 (-94.13)	0.6632 (-22.72)	0.0132 (-76.33)	0.8324 (-16.76)	0.0133 (-90.28)	0.671 (-21.8)	0.0779 (39.24)	0.7746 (-22.54)	0.0353 (-74.15)	0.6342 (-26.1)	0.1252 (123.74)	0.7842 (-21.58)	0.08 (-41.44)	0.8752 (2.0)	0.1024 (83.11)	0.6056 (-39.44)	**0.134** (-1.9)	0.8338 (-2.84)	0.062 (10.85)	0.7086 (-29.14)
Most Rated	**0.0233** (-82.98)	1.1165 (30.11)	0.0082 (-85.38)	0.3892 (-61.08)	**0.0786** (-42.46)	1.1528 (34.34)	0.0712 (27.2)	0.3478 (-65.22)	0.0853 (-37.51)	1.1706 (36.42)	0.044 (-21.29)	0.3417 (-65.83)	0.1039 (-23.91)	1.022 (19.1)	0.0389 (-30.46)	0.5071 (-49.29)	0.1278† (-6.46)	0.8384 (-2.3)	0.0545 (-2.66)	0.7055 (-29.45)
Most Characteristic	0.001 (-99.26)	**0.3557** (-58.55)	**0.0004** (-99.31)	**1.145** (14.5)	0.009 (-93.39)	**0.3682** (-57.09)	**0.0042** (-92.41)	**1.1372** (13.72)	0.0256 (-81.23)	**0.4283** (-50.09)	0.028 (-50.0)	**1.0684** (6.84)	0.0442 (-67.66)	**0.5738** (-33.13)	0.029 (-48.23)	**0.9391** (-6.09)	0.1302 (-4.66)	0.8584 (0.04)	0.0584 (4.33)	0.6828 (-31.72)
Highest Variance	0.0047 (-96.59)	0.5715 (-33.4)	0.0026 (-95.35)	0.9304 (-6.96)	0.0272 (-80.12)	0.5781 (-32.63)	0.0128 (-77.15)	0.9348 (-6.52)	0.0587 (-57.02)	0.6896 (-19.64)	0.0317 (-43.4)	0.8308 (-16.92)	0.0902 (-33.96)	0.8125 (-5.32)	0.051 (-8.82)	0.7113 (-28.87)	0.1244 (-8.88)	0.8497 (-0.98)	0.0544 (-2.84)	0.6915 (-30.85)

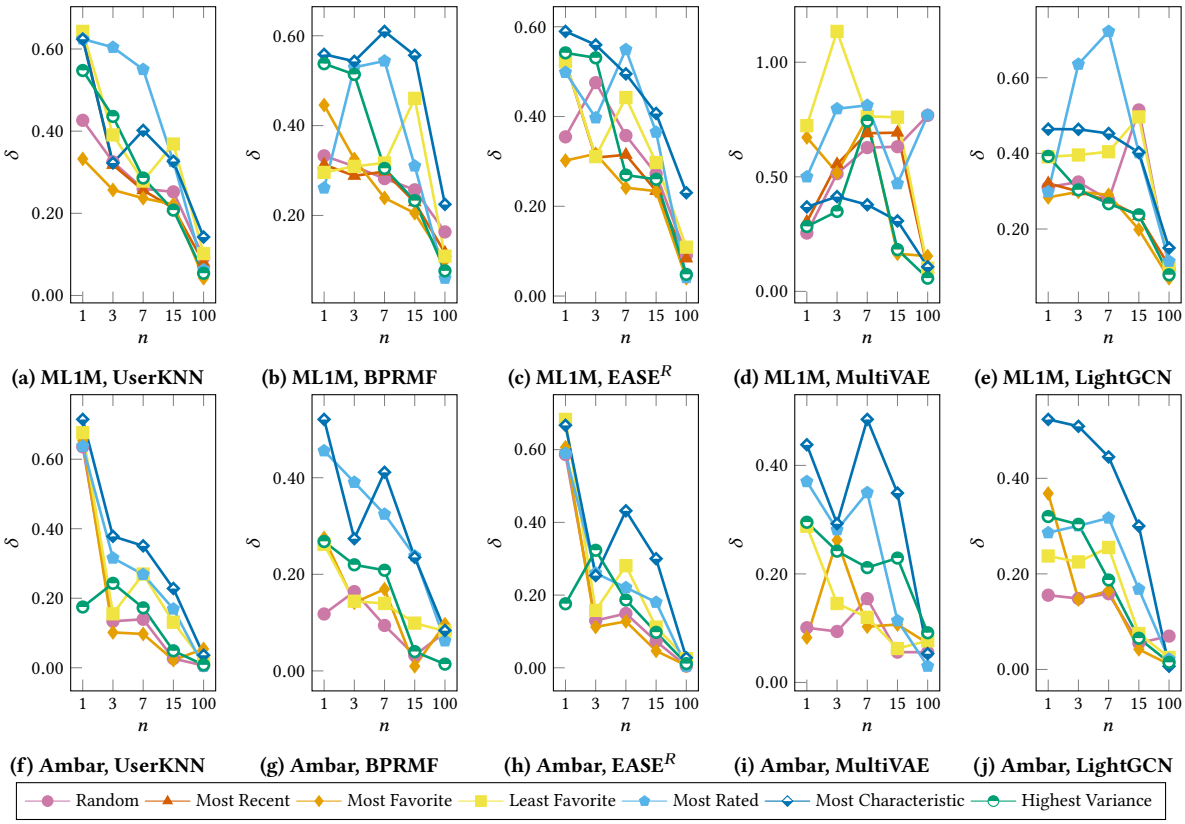


Figure 1: Comparison of the Euclidean distance (δ) between the point representing the model performance under a minimization strategy with value n and the point representing the model performance under the "full" data strategy. The colored lines represent the different minimization strategy. Smaller δ values indicate better preservation of the model's overall performance.

second-best nDCG values. However, this focus on popularity comes at a cost: provider fairness, as measured by RSP, deteriorates significantly, especially at smaller n. As n increases, the items are exposed more fairly due to the inclusion of a broader range of items in the training set. Nevertheless, the low values of HV indicate that this strategy predominantly optimizes for accuracy, with limited alignment to beyond-accuracy objectives. The "Most Favorite" strategy prioritizes the items most liked by individual users, resulting in competitive nDCG values (often the second-best) while mitigating the fairness degradation seen in "Most Rated." This strategy strikes a better balance between accuracy and fairness since the items most liked by users are not necessarily the most popular. Indeed, the acceptable HV values suggest a more harmonious fulfillment of accuracy and fairness objectives. In contrast, the "Most Characteristic" strategy consistently achieves the worst performance in terms of accuracy but demonstrates improvements in provider fairness. This strategy favors the representation of items most interacted with by the most active users, effectively tailoring user profiles for this subgroup. However, on most platforms, most users are less active, resulting in a decline in overall recommendation quality when adopting this strategy. The best HV values associated with this strategy indicate that it excels in achieving beyond-accuracy objectives, albeit with poor accuracy performance. The remaining strategies introduce higher variability in user profiles and tend to achieve a better balance between accuracy and fairness. This balance is reflected in their HV values, which suggest moderate fulfillment of all objectives.

To answer RQ1, we conclude that data minimization can lead to a decrease in the accuracy performance of recommendation models. While this outcome directly offers fairer recommendations across different user groups, data minimization can be responsible for unfair exposure of items. Hence, these objectives potentially show a critical trade-off, underscoring the need for nuanced strategies to balance these competing objectives in recommendation systems.

4.2 Data Minimization Strategies Performance (RQ2)

To answer RQ2, we evaluate how effectively data minimization strategies preserve the performance of recommender systems compared to the "full" strategy. We adopt a multi-objective evaluation approach [29, 30] that simultaneously considers multiple dimensions of analysis. Fixing a value of n and a recommendation model, we measure the model's performance regarding nDCG, RSP, and MAD for each minimization strategy applied. These metrics define a three-dimensional objective function space, where each model trained under a specific minimization strategy is represented as

Table 3: Mean (μ) and standard deviation (σ) of the distances among the points represented by each model performance under different strategies and the point corresponding to the model performance under the "full" data strategy, across different n. For each scenario, bold and underline stand for best and second-to-best values, respectively.

Model	$n=1$		$n=3$		$n=7$		$n=15$		$n=100$		Global	
	$\mu\downarrow$	$\sigma\downarrow$	$\mu\downarrow$	$\sigma\downarrow$	$\mu\downarrow$	$\sigma\downarrow$	$\mu\downarrow$	$\sigma\downarrow$	$\mu\downarrow$	$\sigma\downarrow$	$\mu\downarrow$	$\sigma\downarrow$
Movielens1M												
UserKNN	0.547	0.1216	**0.379**	0.1146	**0.3241**	**0.1134**	**0.2736**	**0.0656**	**0.0821**	0.0339	0.3212	0.1777
BPRMF	0.3921	0.1211	0.4027	0.1188	0.3705	0.1443	0.3217	0.1346	0.1216	0.056	0.3217	**0.1537**
EASER	0.4758	0.1058	0.4141	**0.1082**	0.3814	0.1169	0.2957	0.0666	0.0918	0.0667	0.3317	0.1622
MultiVAE	0.4435	0.1916	0.6115	0.2701	0.6814	0.1461	0.4583	0.2455	0.2935	0.3259	0.4977	0.2676
LightGCN	**0.3515**	**0.0657**	0.3887	0.1253	0.384	0.1662	0.3551	0.1309	0.1014	**0.0261**	**0.3161**	0.1538
Ambar												
UserKNN	0.5821	0.2017	0.2212	0.1098	**0.2163**	**0.0956**	**0.1043**	**0.085**	0.0212	0.02	0.229	0.2236
BPRMF	0.3166	0.1467	0.2223	0.0969	0.2245	0.1203	0.109	0.1031	0.0698	0.0294	0.1884	0.1339
EASER	0.5516	0.1884	**0.2063**	0.0847	0.2328	0.1116	0.135	0.0926	**0.0133**	**0.0105**	0.2278	0.21
MultiVAE	**0.2625**	0.1433	0.2196	**0.0809**	0.2372	0.1506	0.1529	0.1145	0.0632	0.0216	**0.1871**	**0.1275**
LightGCN	0.3151	**0.1254**	0.2722	0.1351	0.2546	0.1113	0.1177	0.1	0.0249	0.0231	0.1969	0.1483

a point identified by the corresponding metric values. The model trained using the "full" strategy serves as a reference point in this space. To quantify the similarity of performance between the "full" strategy and each minimization strategy, we compute the Euclidean distance δ of every point (i.e., the model trained under a minimization strategy) from the reference point. A smaller Euclidean distance indicates that the minimization strategy better preserves the model's overall performance compared to the "full" strategy. Figure 1 presents the Euclidean distance results for each model and dataset as n varies.

Consistent with the trends observed in Tables 1 and 2, setting $n = 100$ enables most strategies to effectively preserve model performance, as evidenced by the lowest δ values. No single strategy emerges as a definitive winner at this level, reflecting similar performance across strategies. Conversely, as n decreases, a more evident hierarchy among the strategies emerges.

The "Most Rated" and "Most Characteristic" strategies generally exhibit higher δ values, indicating poorer performance preservation—except for MultiVAE trained on ML1M (Figure 1d). This behavior aligns with the observations in Section 4.1: the "Most Rated" strategy degrades model performance primarily regarding fairness, while the "Most Characteristic" strategy compromises accuracy.

In contrast, the "Most Favorite" strategy consistently preserves model performance, frequently achieving the lowest distance values. As discussed in Section 4.1, this strategy effectively delineates user profiles without adversely affecting fairness. Similar performance trends can be observed for the "Random," "Most Recent," and "Highest Variance" strategies. These strategies introduce variability into user profiles, yet this variability enables the retention of average user profile characteristics. Consequently, these strategies maintain acceptable accuracy levels while avoiding a disproportionate emphasis on specific items in the training dataset. The resulting datasets closely resemble those generated by the "full" strategy, which explains their superior performance compared to the "Most Rated" and "Most Characteristic" strategies. Among these, the "Random" strategy is the most effective. As noted in Section 3.5, it best maintains the average profile of each user. Finally, the "Least Favorite" strategy generally underperforms compared to the "Random" strategy. This behavior is likely because models find learning effectively from negative feedback more challenging, resulting in more significant performance degradation.

To answer RQ2, we say that data minimization strategies that can shape better user profiles and introduce variability strike a better balance among the considered objectives, being able to retain the model's performance compared to the "full" strategy.

4.3 Robustness of Recommenders to Data Minimization (RQ3)

To address RQ3, we evaluate the robustness of recommendation models under data minimization while considering both accuracy and fairness. Here, robustness is defined as a model's ability to preserve its performance under data minimization compared to the full-data setting. We adopt the multi-objective approach outlined in Section 4.2 to quantify this. We compute the performance achieved under each data minimization strategy for each model in the three-dimensional objective function space—defined by the nDCG, RSP, and MAD metrics. Given a model and fixed n, we measure the Euclidean distances of these performance points from the reference point, i.e., the model's performance under the "full" strategy. The mean μ and standard deviation σ of these distances capture the model's robustness: a lower μ indicates better overall robustness, while a lower σ suggests more consistent performance across strategies. Table 3 reports μ and σ for each model across different values of n and their global averages across all settings (i.e., without fixing n).

In Ambar and ML1M, UserKNN demonstrates strong robustness for $n \geq 3$, consistently achieving the lowest (or second-lowest) values of μ and σ. This finding aligns with previous works [7, 8], which

showed that neighborhood-based models can maintain performance even after removing random interactions. Similarly, EASE^R is robust only when data minimization is moderate. Its reliance on item co-occurrence statistics makes it highly sensitive to severe sparsity. If the goal is to ensure robustness even with minimal data ($n = 1$), LightGCN emerges as the most reliable model across both datasets. This finding aligns with its graph-based structure, which propagates information from neighboring users and items, mitigating the impact of data sparsity. Moreover, LightGCN maintains low μ and σ across all scenarios, making it the most robust model overall. BPRMF follows closely behind LightGCN, demonstrating strong generalization under data minimization. Its factorization-based approach enables it to retain performance even with limited interactions, though it lacks the graph-based smoothing effect of LightGCN, leading to slightly higher variability. Finally, MultiVAE exhibits the most dataset-dependent behavior, showing weaker robustness in ML1M but achieving the best performance in Ambar at the global level. This observation suggests that its variational latent space representations are highly sensitive to dataset characteristics, which may impact its stability under data minimization.

To answer RQ3, we conclude that graph-based modeling is crucial for robustness under extreme data sparsity when considering accuracy and fairness objectives, enhancing the classic factorization-based representation of users and items. Additionally, traditional neighborhood-based and non-linear methods show limitations in such scenarios, suffering from high data reduction.

5 Conclusion and Future Work

In this study, we systematically investigated the impact of data minimization on recommender systems, with particular attention to the trade-offs among accuracy, consumer fairness, and provider fairness. Our contributions extend the current understanding of data minimization in three key ways. First, we empirically evaluate how different data minimization strategies across real-world datasets impact the accuracy, provider, and consumer fairness of recommendations. Second, we employ a multi-objective evaluation to investigate which data minimization strategies can simultaneously preserve recommendation performance across multiple objectives. Third, we explore how different model classes respond to varying degrees of data reduction, shedding light on the relationship between architectural design and data scarcity.

Our findings show that although data minimization can diminish accuracy performance, it levels the recommendation quality of different user groups, enhancing consumer fairness. However, reducing data can inadvertently lead to the unfair exposure of specific items, highlighting a critical trade-off between data minimization, accuracy, and fairness. Further analysis highlights specific behaviors of data minimization strategies. Particularly, approaches that introduce variability when shaping the user profiles can strike a better balance among data minimization, accuracy, and fairness of recommendations. Finally, our insights indicate that graph-based recommender systems remain more resilient under extreme data minimization when fairness and accuracy are jointly considered, in contrast to traditional methods that exhibit lower robustness.

Future works may seek to devise more refined data minimization strategies that promote equitable recommendations while minimizing adverse effects on accuracy performance. In addition, our insight can guide the development of recommender systems specifically devised to be robust under severe data minimization conditions while accomplishing fairness objectives.

Acknowledgments

The authors acknowledge partial support of the following projects: OVS: Fashion Retail Reloaded, ePansa, Lutech Digitale 4.0, Secure Safe Apulia, Patti Territoriali WP1, BIO-D, Meditech. We acknowledge the CINECA award under the ISCRA initiative for the availability of high-performance computing resources and support.

References

[1] [n. d.]. California Consumer Privacy Act (CCPA). https://cloud.google.com/security/compliance/ccpa. accessed 2024-02-13.
[2] [n. d.]. The California Privacy Rights Act of 2020. https://thecpra.org/. accessed 2024-02-13.
[3] [n. d.]. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679. accessed 2024-03-15.
[4] [n. d.]. Translation: Personal Information Protection Law of the People's Republic of China – Effective Nov. 1, 2021. https://digichina.stanford.edu/work/translation-personal-information-protection-law-of-the-peoples-republic-of-china-effective-nov-1-2021/. accessed 2024-02-13.
[5] Vito Walter Anelli, Alejandro Bellogín, Antonio Ferrara, Daniele Malitesta, Felice Antonio Merra, Claudio Pomo, Francesco Maria Donini, and Tommaso Di Noia. 2021. Elliot: A Comprehensive and Rigorous Framework for Reproducible Recommender Systems Evaluation. In *SIGIR*. ACM, 2405–2414.
[6] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Daniele Malitesta, Vincenzo Paparella, and Claudio Pomo. 2023. Auditing Consumer- and Producer-Fairness in Graph Collaborative Filtering. In *Advances in Information Retrieval - 45th European Conference on Information Retrieval, ECIR 2023, Dublin, Ireland, April 2-6, 2023, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 13980)*, Jaap Kamps, Lorraine Goeuriot, Fabio Crestani, Maria Maistro, Hideo Joho, Brian Davis, Cathal Gurrin, Udo Kruschwitz, and Annalina Caputo (Eds.). Springer, 33–48. doi:10.1007/978-3-031-28244-7_3
[7] Asia J. Biega, Peter Potash, Hal Daumé III, Fernando Diaz, and Michèle Finck. 2020. Operationalizing the Legal Principle of Data Minimization for Personalization. In *SIGIR*. ACM, 399–408.
[8] Richard Chow, Hongxia Jin, Bart P. Knijnenburg, and Gökay Saldamli. 2013. Differential data analysis for recommender systems. In *Seventh ACM Conference on Recommender Systems, RecSys '13, Hong Kong, China, October 12-16, 2013*, Qiang Yang, Irwin King, Qing Li, Pearl Pu, and George Karypis (Eds.). ACM, 323–326. doi:10.1145/2507157.2507190
[9] Gemma Galdon Clavell, Mariano Martín Zamorano, Carlos Castillo, Oliver Smith, and Aleksandar Matic. 2020. Auditing Algorithms: On Lessons Learned and the Risks of Data Minimization. In *AIES*. ACM, 265–271.
[10] Enyan Dai, Tianxiang Zhao, Huaisheng Zhu, Junjie Xu, Zhimeng Guo, Hui Liu, Jiliang Tang, and Suhang Wang. 2024. A Comprehensive Survey on Trustworthy Graph Neural Networks: Privacy, Robustness, Fairness, and Explainability. *Mach. Intell. Res.* 21, 6 (2024), 1011–1061.
[11] Yashar Deldjoo, Vito Walter Anelli, Hamed Zamani, Alejandro Bellogín, and Tommaso Di Noia. 2021. A flexible framework for evaluating user and item fairness in recommender systems. *User Model. User Adapt. Interact.* 31, 3 (2021), 457–511. doi:10.1007/S11257-020-09285-1
[12] Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogín, Alessandro Difonzo, and Dario Zanzonelli. 2024. Fairness in recommender systems: research landscape and future directions. *User Model. User Adapt. Interact.* 34, 1 (2024), 59–108.
[13] Yashar Deldjoo, Tommaso Di Noia, and Felice Antonio Merra. 2022. A Survey on Adversarial Recommender Systems: From Attack/Defense Strategies to Generative Adversarial Networks. *ACM Comput. Surv.* 54, 2 (2022), 35:1–35:38.
[14] Tobias Eichinger and Axel Küpper. 2023. Distributed Data Minimization for Decentralized Collaborative Filtering Systems. In *ICDCN*. ACM, 140–149.
[15] Tobias Eichinger and Axel Küpper. 2024. On data minimization and anonymity in pervasive mobile-to-mobile recommender systems. *Pervasive Mob. Comput.* 103 (2024), 101951.
[16] Michèle Finck and Asia Biega. 2021. Reviving Purpose Limitation and Data Minimisation in Personalisation, Profiling and Decision-Making Systems. *CoRR*

abs/2101.06203 (2021).
[17] M. Fleischer. 2003. The Measure of Pareto Optima. In *EMO (Lecture Notes in Computer Science, Vol. 2632)*. Springer, 519–533.
[18] Prakhar Ganesh, Cuong Tran, Reza Shokri, and Ferdinando Fioretto. 2024. The Data Minimization Principle in Machine Learning. *CoRR* abs/2405.19471 (2024).
[19] Elizabeth Gómez, David Contreras, Ludovico Boratto, and Maria Salamó. 2024. AMBAR: A dataset for Assessing Multiple Beyond-Accuracy Recommenders. In *RecSys*. ACM, 137–147.
[20] F. Maxwell Harper and Joseph A. Konstan. 2016. The MovieLens Datasets: History and Context. *ACM Trans. Interact. Intell. Syst.* 5, 4 (2016), 19:1–19:19.
[21] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yong-Dong Zhang, and Meng Wang. 2020. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. In *SIGIR*. ACM, 639–648.
[22] Yassine Himeur, Shahab Saquib Sohail, Faycal Bensaali, Abbes Amira, and Mamoun Alazab. 2022. Latest trends of security and privacy in recommender systems: A comprehensive review and future perspectives. *Comput. Secur.* 118 (2022), 102746.
[23] Walid Krichene and Steffen Rendle. 2020. On Sampled Metrics for Item Recommendation. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash (Eds.). ACM, 1748–1757. doi:10.1145/3394486.3403226
[24] Nian Li, Chen Gao, Depeng Jin, and Qingmin Liao. 2023. Disentangled Modeling of Social Homophily and Influence for Social Recommendation. *IEEE Trans. Knowl. Data Eng.* 35, 6 (2023), 5738–5751.
[25] Dawen Liang, Rahul G. Krishnan, Matthew D. Hoffman, and Tony Jebara. 2018. Variational Autoencoders for Collaborative Filtering. In *WWW*. ACM, 689–698.
[26] Jianghao Lin, Xinyi Dai, Yunjia Xi, Weiwen Liu, Bo Chen, Hao Zhang, Yong Liu, Chuhan Wu, Xiangyang Li, Chenxu Zhu, et al. 2023. How can recommender systems benefit from large language models: A survey. *ACM Transactions on Information Systems* (2023).
[27] Qin Liu, Xuan Feng, Tianlong Gu, and Xiaoli Liu. 2024. FairCRS: Towards User-oriented Fairness in Conversational Recommendation Systems. In *RecSys*. ACM, 126–136.
[28] Alberto Carlo Maria Mancino, Antonio Ferrara, Salvatore Bufi, Daniele Malitesta, Tommaso Di Noia, and Eugenio Di Sciascio. 2023. KGTORe: Tailored Recommendations through Knowledge-aware GNN Models. In *RecSys*. ACM, 576–587.
[29] Vincenzo Paparella, Vito Walter Anelli, Franco Maria Nardini, Raffaele Perego, and Tommaso Di Noia. 2023. Post-hoc Selection of Pareto-Optimal Solutions in Search and Recommendation. In *CIKM*. ACM, 2013–2023.
[30] Vincenzo Paparella, Dario Di Palma, Vito Walter Anelli, and Tommaso Di Noia. 2023. Broadening the Scope: Evaluating the Potential of Recommender Systems beyond prioritizing Accuracy. In *Proceedings of the 17th ACM Conference on Recommender Systems, RecSys 2023, Singapore, Singapore, September 18-22, 2023*, Jie Zhang, Li Chen, Shlomo Berkovsky, Min Zhang, Tommaso Di Noia, Justin Basilico, Luiz Pizzato, and Yang Song (Eds.). ACM, 1139–1145. doi:10.1145/3604915.3610649
[31] Hossein A. Rahmani, Mohammadmehdi Naghiaei, and Yashar Deldjoo. 2024. A Personalized Framework for Consumer and Producer Group Fairness Optimization in Recommender Systems. *Trans. Recomm. Syst.* 2, 3 (2024), 19:1–19:24.
[32] Theresia Veronika Rampisela, Maria Maistro, Tuukka Ruotsalo, and Christina Lioma. 2023. Evaluation Measures of Individual Item Fairness for Recommender Systems: A Critical Study. *CoRR* abs/2311.01013 (2023).
[33] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2012. BPR: Bayesian Personalized Ranking from Implicit Feedback. *CoRR* abs/1205.2618 (2012).
[34] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. 1994. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In *CSCW*. ACM, 175–186.
[35] Abdulaziz Samra, Evgeny Frolov, Alexey Vasilev, Alexander Grigorevskiy, and Anton Vakhrushev. 2024. Cross-Domain Latent Factors Sharing via Implicit Matrix Factorization. In *RecSys*. ACM, 309–317.
[36] Divya Shanmugam, Fernando Diaz, Samira Shabanian, Michèle Finck, and Asia Biega. 2022. Learning to Limit Data Collection via Scaling Laws: A Computational Interpretation for the Legal Principle of Data Minimization. In *FAccT*. ACM, 839–849.
[37] Nasim Sonboli, Sipei Li, Mehdi Elahi, and Asia Biega. 2024. The trade-off between data minimization and fairness in collaborative filtering. *CoRR* abs/2410.07182 (2024).
[38] Giuseppe Spillo, Allegra De Filippo, Cataldo Musto, Michela Milano, and Giovanni Semeraro. 2024. Towards Green Recommender Systems: Investigating the Impact of Data Reduction on Carbon Footprint and Algorithm Performances. In *RecSys*. ACM, 866–871.
[39] Robin Staab, Nikola Jovanovic, Mislav Balunovic, and Martin T. Vechev. 2024. From Principle to Practice: Vertical Data Minimization for Machine Learning. In *SP*. IEEE, 4733–4752.
[40] Harald Steck. 2013. Evaluation of recommendations: rating-prediction and ranking. In *RecSys*. ACM, 213–220.
[41] Harald Steck. 2019. Embarrassingly Shallow Autoencoders for Sparse Data. In *WWW*. ACM, 3251–3257.
[42] Tobias Vente, Lukas Wegmeth, Alan Said, and Joeran Beel. 2024. From Clicks to Carbon: The Environmental Toll of Recommender Systems. In *Proceedings of the 18th ACM Conference on Recommender Systems, RecSys 2024, Bari, Italy, October 14-18, 2024*, Tommaso Di Noia, Pasquale Lops, Thorsten Joachims, Katrien Verbert, Pablo Castells, Zhenhua Dong, and Ben London (Eds.). ACM, 580–590. doi:10.1145/3640457.3688074
[43] Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. 2023. A Survey on the Fairness of Recommender Systems. *ACM Trans. Inf. Syst.* 41, 3 (2023), 52:1–52:43.
[44] Hongyi Wen, Longqi Yang, Michael Sobolev, and Deborah Estrin. 2018. Exploring recommendations under user-controlled data filtering. In *RecSys*. ACM, 72–76.
[45] Keping Yu, Zhiwei Guo, Yu Shen, Wei Wang, Jerry Chun-Wei Lin, and Takuro Sato. 2022. Secure Artificial Intelligence of Things for Implicit Group Recommendations. *IEEE Internet Things J.* 9, 4 (2022), 2698–2707.
[46] Ziwei Zhu, Jianling Wang, and James Caverlee. 2020. Measuring and Mitigating Item Under-Recommendation Bias in Personalized Ranking Systems. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*, Jimmy X. Huang, Yi Chang, Xueqi Cheng, Jaap Kamps, Vanessa Murdock, Ji-Rong Wen, and Yiqun Liu (Eds.). ACM, 449–458. doi:10.1145/3397271.3401177