Article

Modeling and Validating a News Recommender Algorithm in a Mainstream Medium-Sized News Organization: An Experimental Approach

Paschalia (Lia) Spyridou *, Constantinos Djouvas and Dimitra Milioni

Department of Communication and Internet Studies, Cyprus University of Technology, Saripolou 33, 3036 Limassol, Cyprus
* Correspondence: l.spyridou@cut.ac.cy

Abstract: News recommending systems (NRSs) are algorithmic tools that filter incoming streams of information according to the users' preferences or point them to additional items of interest. In today's high-choice media environment, attention shifts easily between platforms and news sites and is greatly affected by algorithmic technologies; news personalization is increasingly used by news media to woo and retain users' attention and loyalty. The present study examines the implementation of a news recommender algorithm in a leading news media organization on the basis of observation of the recommender system's outputs. Drawing on an experimental design employing the 'algorithmic audit' method, and more specifically the 'collaborative audit' which entails utilizing users as testers of algorithmic systems, we analyze the composition of the personalized MyNews area in terms of accuracy and user engagement. Premised on the idea of algorithms being black boxes, the study has a two-fold aim: first, to identify the implicated design parameters enlightening the underlying functionality of the algorithm, and second, to evaluate in practice the NRS through the deployed experimentation. Results suggest that although the recommender algorithm manages to discriminate between different users on the basis of their past behavior, overall, it underperforms. We find that this is related to flawed design decisions rather than technical deficiencies. The study offers insights to guide the improvement of NRSs' design that both considers the production capabilities of the news organization and supports business goals, user demands and journalism's civic values.

Keywords: news personalization; news recommender systems; algorithmic design; algorithmic journalism; algorithmic agenda; beyond accuracy

1. Introduction

More than eight in ten Americans consume news from digital devices, with 60% claiming to do so often [1]. Similar trends are documented in Europe, yet recent research [2] found that the youngest cohort represents a more casual, less loyal news user. Social natives tend to heavily rely on social media for news, while both digital and social natives share a weak connection with brands, making it harder for media organizations to attract and engage them. At the same time, younger audiences are also particularly suspicious and less trusting of all information provided by news outlets (p. 45). These findings reflect the ongoing turmoil of the news media industry. In a high-choice media environment where multiple players offer news and users can access news content from a variety of pathways, and in many different modalities [3], media organizations find it increasingly hard to woo and retain users' attention. Myllylahti (2020) [4] defines attention as "a scarce and fluid commodity which carries monetary value" (p. 568); it is based on individual user interaction which can be measured and analyzed through detailed web metrics and analytics and exchanged for revenue [5,6]. As a result, news media resort to multiple strategies and techniques aiming to control what news people pay attention to and the conditions under which they do so in order to generate revenue [7]. A prominent technique increasingly used

by news organizations is News Recommending Systems (NRSs). NRSs are algorithmic tools that filter incoming streams of information according to the users' preferences or point them to additional items of interest. Harnessing the deluge of big data and machine learning [8], these technical systems aggregate, filter, select and prioritize information [9]. The *Daily Me* project foreseen by Nicholas Negroponte [10] in 1995 is becoming common practice as news publishers are increasingly experimenting with recommender systems to extent the provision of personalized news [11] hoping to increase their sites' 'stickiness', capture user data and reduce their dependence on external suppliers of such information [12].

However, their job is not easy; in today's high-choice media environment, attention shifts easily between platforms and (news) sites [4] and is greatly affected by algorithmic technologies [9,13]. More importantly, the task of recommending appropriate and relevant news stories to readers proves particularly challenging; in addition to the technical and design challenges associated with the news recommendations of most media offerings, personalized systems of news delivery present a special case given the impact of news for an informed citizenry [14,15].

Taking into consideration the business challenges of media organizations and the particularities of news content, the present study probes the algorithmic design of a news recommender introduced in the newsroom of a leading news portal in Cyprus. The study follows a design-oriented approach [16] aiming to identify the implicated parameters enlightening the underlying functionality of the algorithm and evaluate the NRS in hand to offer insights that can guide the improvement of NRS that support and align with business goals, user demands and journalism's civic values.

2. Problem Definition and Motivation: Modeling a News Recommending System

The collapse of the traditional advertising model [17], the web's free news culture [18] and the growing role of the platforms in news distribution [19,20] have made it increasingly difficult for news organizations to cope with editorial and commercial standards [21]. Initially, publishers saw the platforms as an ally to help them boost content visibility and brand awareness; soon, they realized that this evolving publisher–platform partnership is unequal; platforms wield more power over user data and earn significantly more advertising revenue than publishers [22,23]. Amid increasing demands to woo and retain users' attention, the use of algorithmic and data-driven techniques is gaining more and more ground in the media industry; they are used to automate workflows by modeling human-centered practices, thereby assigning new roles to both machines and media professionals [24,25].

News recommendations bear substantial benefits for all parties involved: first, they comprise an effective content monetization tool in terms of building traffic, engagement and loyalty [26]. Second, they help readers discover the depth of the outlets' reporting; The *New York Times*, for example, publishes approximately 250 stories per day. Algorithmic curation is used to propose content facilitating users to encounter news stories that might prove helpful and interesting to them and might otherwise not find while motivating the medium to keep producing a wide range of content [27], as the tailored delivery of news allows republishing content on a much broader scale. Finally, NRS have shown to be a useful tool for helping users deal with information overload [28]. Users are bombarded with news and information from different news outlets, social network posts, notifications, emails, etc., a situation which affects their attention span while making it increasingly difficult for them to find content of interest, at the right moment, in the right form [29].

Despite the acknowledged merits of news personalization, the technical challenge for offering effective recommendations is high [9,30] and particularly expensive [31]. Karimi and his colleagues [32] provide a comprehensive review of the many challenges associated with algorithmic accuracy and user profiling in news recommender systems. However, apart from accuracy and user profiling which comprise typical algorithmic challenges, news recommendation systems present a special case considering their civic function in the direction of an informed citizenry [33]. From a normative perspective, news provides the necessary information for citizens to think, discuss and decide wisely, to participate

in political life and thus democracy to function [34]. For that reason, much work on news recommendations focuses on exposure diversity as a design principle for news recommender systems (see [35–37]). The idea is that for a functioning democracy, users should encounter a variety of themes, opinions and ideas. Helberger and her colleagues [38] argue that "recommendation systems can be instrumental in realizing or obstructing public values and freedom of expression in a digital society; much depends on the design of these systems" (p. 3).

Recent diversity preoccupations echo older concerns arguing that the retreat from human editorial decision making in favor of machine choices might prove problematic. Pariser's [39] widely known hypothesis over filter bubbles refers to algorithmic filtering that tends to promote tailored content based on users' pre-existing attitudes, interests and prejudices, leading citizens into content 'bubbles'. In other words, personalized news offerings may result in people losing the common ground of news [40] by producing "different individual realities" [41] that hinder debate and amplify audience fragmentation and polarization [42]. Relevant work so far provides mixed results. While some studies provide evidence supporting the idea of filter bubble effects (e.g., amplification of selective exposure, negative effects on knowledge gain) [43,44], others argue that these fears are severely exaggerated [45,46]. Considering the social impact of news delivery, and Broussard's [30] argument that the mathematical logic of computers may do well in calculating but often falls short in complex tasks with social or ethical consequences, the possibility that news recommenders can lead to information inequalities or amplify existing biases and thus undermine the democratic functions of the media cannot be excluded [33]. Going a step further, Helberger and her colleagues [35] argue that filtering and recommendation systems can, at least in principle, be designed to offer a personalized news diet which both serves (assumed) individual needs and interests while catering for the provision of a democratic news diet. Recent work by Vrijenhoek and his colleagues [47] formulates the problem explicitly: the question is whether news recommenders are merely designed to generate clicks and short-term engagement or if they are programmed to balance relevance along with helping users discover diverse news and not miss out on important information (p. 173).

Figure 1 depicts the proposed model as a generic approach to describe NRS data flows and processes. The news organization produces a number of news items to be delivered. These items are classified into thematic categories or assigned multiple tags. At the same time, preferences set by the users and/or their previous browsing experience are correlated and matched with the features of the associated stories. Then, the algorithm ranks the matching stories, aiming to deliver accurate recommendations. When developing an effective NRS, one must also consider the beyond-accuracy aspects to evaluate the quality of news recommendations [14]. Therefore, apart from accuracy, the model incorporates the elements of diversity, serendipity and novelty (further defined in the following section) as input features used to limit the possibility of echo-chamber effects while enhancing both users' news experience and their engagement. Overall, the proposed model incorporates baseline data flows and processes that common NRS employ, elaborating on the civic role of journalism (reddish routes) through novel metrics to make the underlying mechanisms more robust and useful.
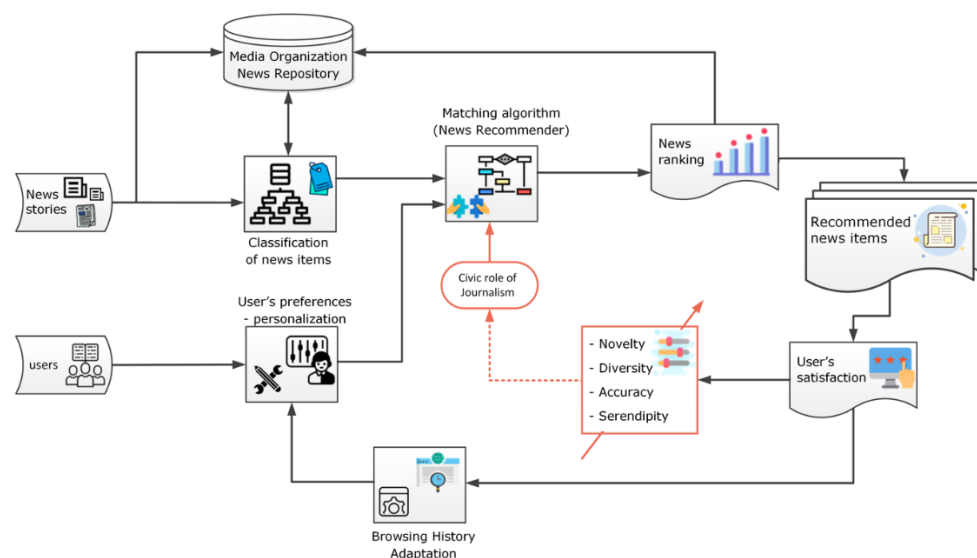
Figure 1. The proposed model incorporates baseline NRS data flows and processes, elaborating on the civic role of journalism (reddish routes) through novel metrics to make the underlying mechanisms more robust and useful.

3. Design Challenges for News Recommenders

An algorithm can be defined as a series of steps undertaken to solve a particular problem or accomplish a defined outcome [48]. In the context of news recommenders, the task of algorithms is to structure and order the pool of available news according to predetermined principles [49]. Algorithms though are embedded with values, biases or ideologies [50] that can influence an effective and unbiased provision of information [40]. Diakopoulos [48] speaks of algorithmic power premised in algorithms' atomic decisions, including prioritization, classification, association and filtering. Some algorithmic power may be exerted intentionally, while other aspects might be unintended side effects rooted in design decisions, objective descriptions, constraints and business rules embedded in the system, major changes that have happened over time, as well as implementation details that might be relevant (p. 404). News recommender systems are often classified into four main categories:

(1) *Collaborative filtering*: Items are recommended to a user based upon values assigned by other people with similar taste. Users are grouped into clusters on the basis of their preferences, habits or content ranking [51]; in practice, collaborative filtering automates the process of 'word-of-mouth' recommendations [52] and is found to be the most common approach in the recommender system literature [32].

(2) *Popularity filtering*: In this case, items are rated for their general popularity among all users; it is the simplest approach, as all users receive the same recommendations, potentially leading to 'popularity biases' and 'bandwagon effects' in which consumers gravitate toward already popular items [33].

(3) *Content-based filtering:* The main idea here is to create clusters of content and associate these clusters with user profiles [51]. Content-based filtering tries to recommend items similar to those a given user has liked in the past based on similarity scores of a user toward all the items [53].

(4) *Hybrid approaches:* Often, news recommender systems use a hybrid approach combining content-based filtering and collaborative filtering (Karimi et al., 2018), also including other methods such as weighing items by recency or pushing content that has specific features (e.g., paid content) [49].

The aforementioned types describe data-driven algorithms, but most news recommender systems employ additional rules which basically shape the overall design of the system. In most cases, these rules are jointly decided by the engineers and the editorial

team of the news organization [16]. So, while in rule-based systems, the automated process is carefully designed to reflect the choices and the criteria that are set by the creators; in data-driven systems, algorithmic bias is less direct, as it is caused by the attributes of the available data that are used to build the decision-making model through the algorithmic process.

When designing NRSs, several issues need to be taken under consideration. Most recommender algorithms are based on a topic-centered approach aiming to meet the different interests of users [54]. However, this approach can prove ineffective considering the specific nature of news items compared to other media offerings [55]. To begin with, news classification (tagging) is a difficult task, as news items may belong to more than one news category. Most often, news organizations follow their own typology of tagging and classifying news stories, which is a practice with a substantial degree of subjectivity. Additionally, most news items have a short lifespan, and thus, it is necessary to process them as fast as possible and start recommending them because their information value degrades [51]. In addition to the element of recency, the popularity of news items may differ dramatically, thereby rendering the traditional recommendation methods unsuccessful [32]. Figure 2 summarizes the main design challenges associated with the particularities of news as a media offering. Furthermore, news recommender systems must deal with a large and heterogenous corpus of news items generating scalability issues. A common strategy for solving scalability is clustering; effective clustering though requires a very thorough classification of news stories and detailed user profiling based on the reading behavior of users and their short-term and long-term profiles [54], which often proves a difficult task. The cold start problem stemming from insufficient user information to estimate user preferences is a common challenge in NRS [32,56]. User registration, among others, comprises a standard method to overcome it. Lavie and her colleagues [57], however, found significant differences between declared and actual interests in news topics, especially in broad news categories containing many subtopics (for instance, politics). They concluded that users cannot accurately assess their interest in news topics and argue that news recommender systems should apply different filtering mechanisms for different news categories. In other words, the depth of personalization should be adjusted to cater for both declared interests and assumed interests of important events (see Figure 3, Challenges associated with profiling).
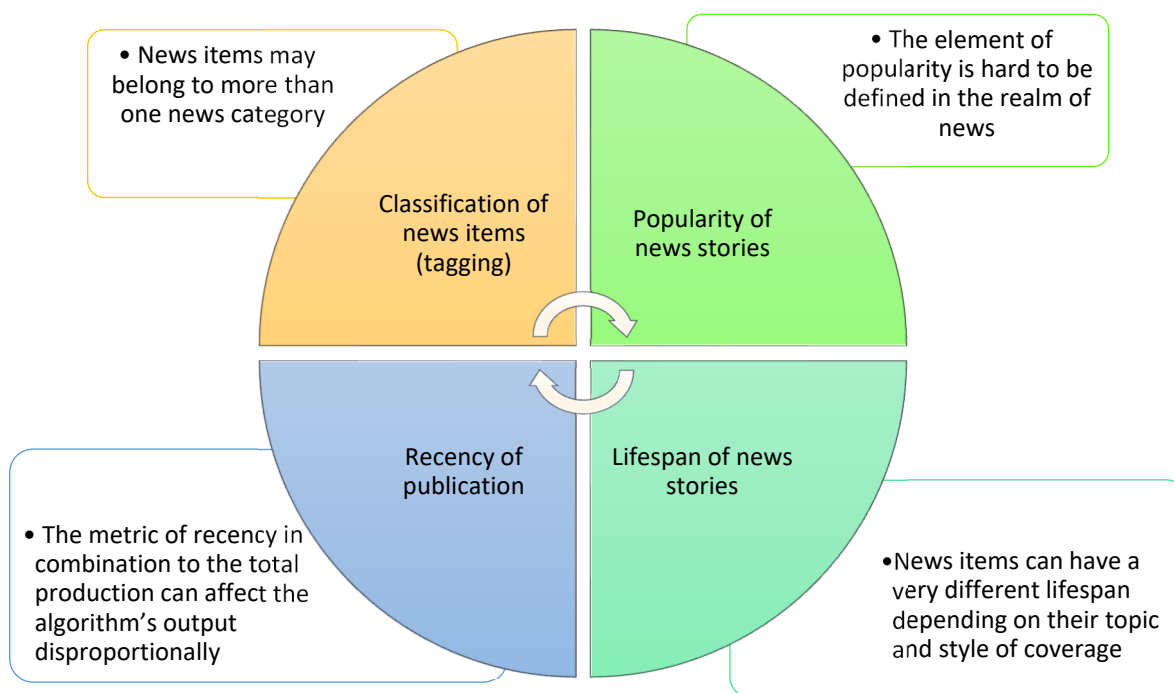

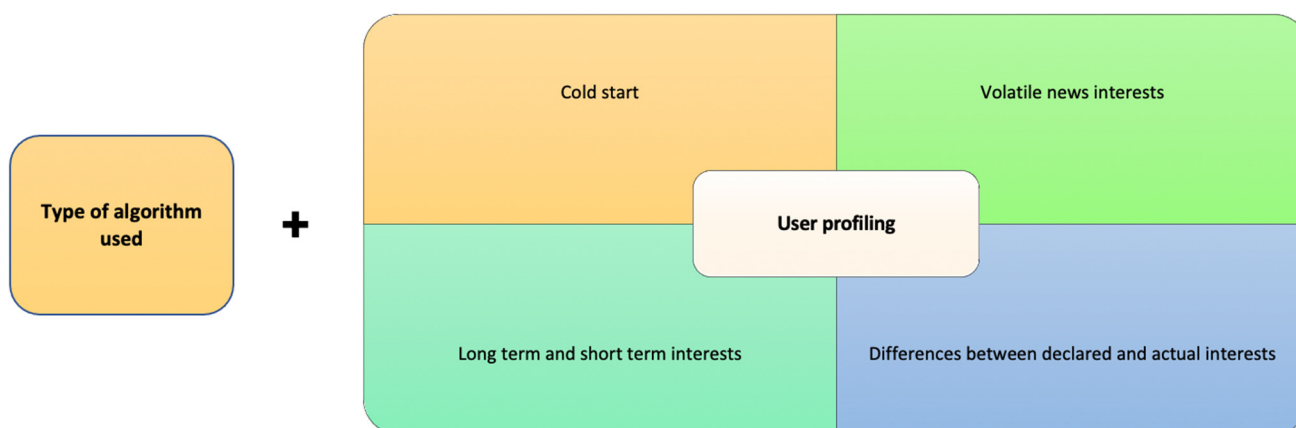**Figure 2.** Design challenges for NRS related to the nature of news.

Figure 3. Challenges relating to user profiling in news recommender systems.

An important factor influencing how specific challenges will be treated pertains to the purpose of the recommender system as shaped by competing logics in the news organization. Smets et al. [58] argue that there is a crucial stage in the design of a recommender system in which the organization decides on the business and strategic goals they want to reach with the personalization service. A purpose-driven evaluation of recommender systems brings along the question of which stakeholder(s) are defining the recommendation purpose and how the conflicts and trade-offs between stakeholders are resolved and embedded in the system. More sophisticated algorithms not only combine hybrid logics in their filtering techniques but also include an element of surprise: serendipity [33]. The serendipity principle posits that the algorithm recommends items that are not only novel but also positively surprising for the user and propose a generic metric based on the concepts of unexpectedness and usefulness [32]. A key challenge of serendipitous recommendations is setting a balance between novelty and users' expectations. Serendipity is considered a quality factor for improving algorithmic output; it helps users keep an open window to new discoveries, it broadens the recommendation spectrum to avoid cases of users losing interest because the choice set is too uniform and narrow and helps integrate new items in order to acquire information on their appeal [49]. In addition to serendipity, the principles of novelty and diversity are deemed quality factors that can broaden the news menu and improve user perceptions and engagement [59]. Again, the 'right' level of novelty and diversity works as a trade-off for accuracy and can depend on the user's current situation and context [32]. Figure 4 depicts the main parameters for evaluating the quality of news recommendations. Recent initiatives aiming to run 'diversity-enhancing' algorithms focus on nudge-like personalization features, for instance, visuals to increase item salience, or item re-ranking in an attempt to curb rigid algorithmic recommendations [60]. Although nudging involves the steering of people in news paths that can enhance their news diet and knowledge [61], initiatives to nudge people toward diversity in their information exposure raise questions of autonomy and freedom of choice [62]. Non-transparent algorithmic nudging may undermine users' freedom of choice even if the principal objective (stimulating diversity) is a noble one [35,60].
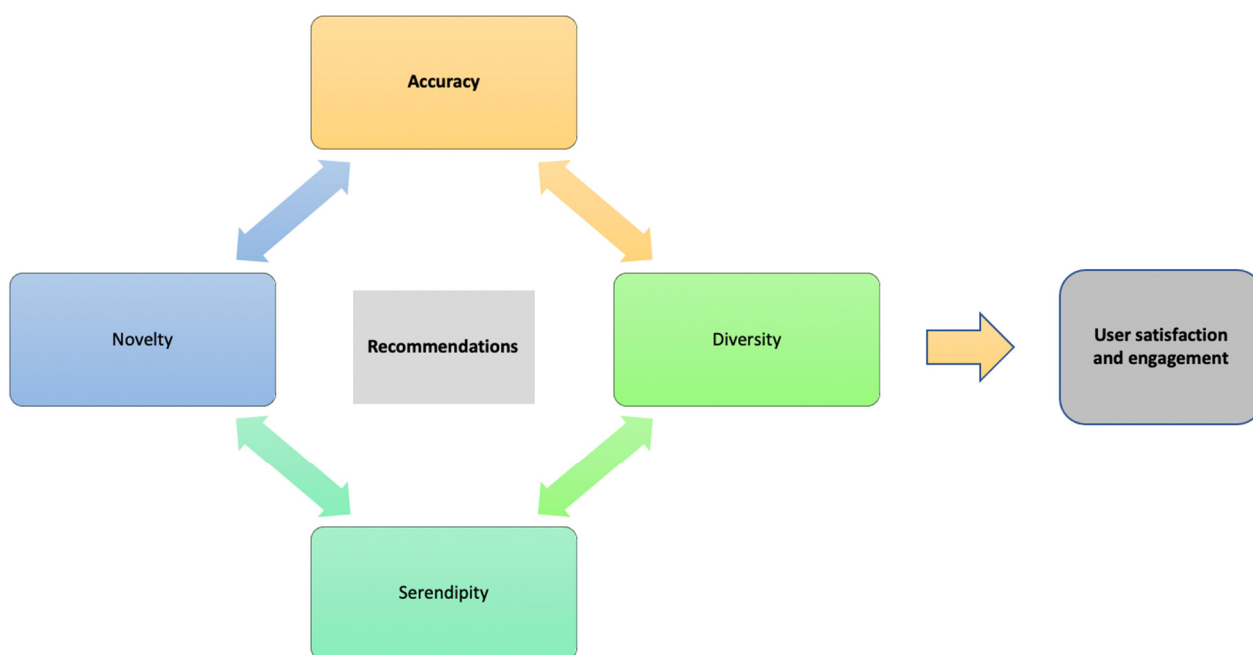
Figure 4. Parameters to evaluate the quality of news recommendations.

4. Experimental Setup for Validating an NRS Algorithm and Its Outcomes

The actual process of developing and rolling out news recommender systems within media companies remains largely under-researched [26]. In practice, explaining the workings of algorithmic systems is notoriously difficult [63] because of the opacity in their automated decision-making capabilities. Drawing on theoretical elements of algorithmic design [48,64], this study explores the implementation of a news recommender system within a leading news media organization in Cyprus, which started offering personalized news in January 2020.

The website is part of one of the largest media houses in Cyprus owning a television station, two radio stations, one newspaper and four established magazine titles. It is the leading online news player with 1.5 million users per month and 30 million page views monthly (source: Google Analytics). It covers political, economic and social affairs through a mainstream perspective. The website includes five sub-domains focusing on economy, sports, features, lifestyle and cooking, respectively. In addition to the website, the content is available on mobile apps. The website operates in a converged newsroom along with journalists from television, print and radio. Web metrics and analytics comprise an integral part of the outlet's content strategy. It manages the largest Facebook page for news in Cyprus with more than 140,000 followers and is also active on Twitter and Instagram.

The news recommender was developed as part of a research project. The outlet advertised its new service prompting users to register. Registration entailed basic user information, such as gender, age and interests. After visiting the website, registered users had the opportunity to log into MyNews, which was a webpage offering personalized stories. Each time, a user entered MyNews, a webpage containing 28–32 news items. These items appeared in a box-type layout, each one containing a photograph and the title of the news item. The boxes appeared in rows, each one displaying five articles.

The study draws on an experimental design employing the 'algorithmic audit' method [65] and more specifically the 'collaborative audit' which entails utilizing users as testers of algorithmic systems. The experiment was set up in a way to monitor the achieved accuracy of the algorithm and to probe the input parameters. To do so, users were divided into two groups: (a) *collaborating users* instructed to only view specific kinds of items (e.g., political ones) and (b) *ordinary users* who could elect freely what news item to engage with. The aim of dividing users into these two distinct groups was to monitor po-

tential accuracy differences stemming from users browsing behavior. Both groups involved registered users of the MyNews NRS service (see Figure 5).
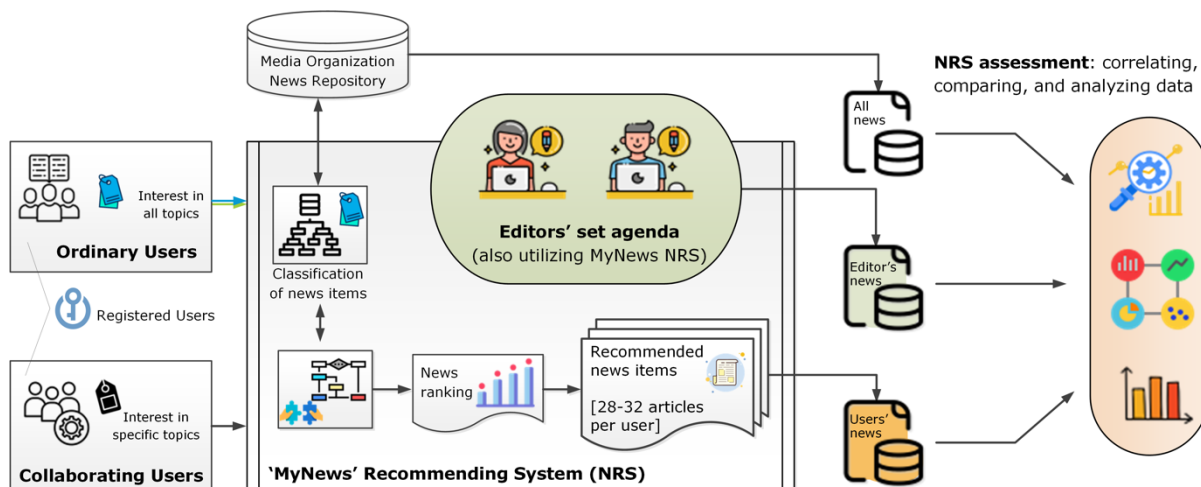

Figure 5. Visualization of the experimental approach: Design and datasets.

We posited the following research questions:

RQ1: When it comes to the news diet outputs produced by the editor-defined agenda and the algorithmic agenda, how different are they and in what ways?

RQ2: How effective is the recommender system in terms of accuracy? In other words, how likely is that the recommender system offers stories a user would be interested in, which is judged on the basis of the two distinct types of users?

RQ3: Are algorithmic recommendations more likely to lead to more clicks, i.e., does the recommender system increase users' engagement?

4.1. Participants

The study draws on the behavior of a total of 18 individuals who registered to the personalized news service of the outlet under study. These users were split into two groups. The first group comprised six individuals who were in essence collaborators of the researchers (*collaborating users*), each of whom was instructed to only click on news items that were of a specific type; e.g., one person was asked to only click on news relating to economy, while another was asked to consume only lifestyle-type news. More specifically, for these *collaborating users*, each was asked to read news pertaining to lifestyle, international news, local (Cyprus) news, politics, the economy and sports. The second group (*ordinary users*) included the remaining 12 individuals that were not given specific instructions and were asked to consume stories that were of interest to them. All users were additionally given the instruction to start perusing from the personalized MyNews area of the website.

4.2. Process and Data Collection

During a period of ten days (21 April 2020–1 May 2020), the following datapoints (see Table 1) were collected for each user and per session (each time they used the designated browser to visit the website under study).

Participants were fully informed that their activity was monitored in this manner; after providing their written consent to participate in the study, they were provided with an ad hoc created Chrome plugin pre-installed into a special browser to enable the collection of the aforementioned data. Users were asked to use this special browser any time they wanted to take part in this experiment (session). The plugin was transparently collecting and storing into a remote MongoDB database all the information necessary for analyzing the behavior of the news recommender algorithm implemented and applied by the medium. Data were only communicated and stored if users elected to press a specific button in their browser.

Table 1. Type of data collected.

	Type of Data	Abbreviation
1	the news items users clicked on	'clicked news items'
2	the news items contained in MyNews personalized area created for them by the recommender algorithm	'MyNews items'
3	the news items contained in the frontpage of the website at the time	'agenda items'
4	all news items published by the outlet per day	'pool items'

In addition to the aforementioned data, we collected information pertaining to the 'pool items' (the stories posted on the website) and also user information. Table 2 describes in detail the type of data collected. Having extracted the posts from the website, each entry was augmented with the user's information in order to associate users with content. User activity was captured by monitoring his/her click activity.

Table 2. Additional data collected pertaining to 'pool items' and user information.

	Pool Items	User Information
1	the section of the webpage where the post had appeared	the user's ID (usually an email address)
2	the news category assigned to the news item (as assigned by the medium)	the user's session ID (an auto-increment number)
3	the post's title	the session's timestamp
4	the post's URL	
5	the post's publication date	

Special mention needs to be made to the 'news category' variable mentioned in Table 2; this was an attribute assigned to each news item by the website, presumably the journalist responsible for writing the relevant article (i.e., these were not manually coded by the researchers). A total of 161 such categories were observed to be present on the news outlet under study; however, this included instances where mistakes had been made in tagging the article (e.g., the tags "Greece and "Greese") and duplications (e.g., using the tag "Greece" and "Ellada", a phonetic spelling out of "Greece" in the Greek language). When cleaned and grouped appropriately, a total of 33 different news categories emerged. Since however, using such a large number of categories would make the results difficult if not impossible to assess, it was decided to group these 33 categories into broader categories (e.g., "lifestyle news" were grouped together with "gossip"), ending with the nine categories reported in the results below (see Table 3). It should be mentioned, however, that this only affects the visual aspects of the results, as all calculations were made using the 33 original news categories assigned by the medium. Finally, it ought to be noted that the plugin maintained the order of posts as they appeared on the website, thereby ensuring that the order of the entries appearing into the database reflected the order the articles were read by the users. Every time a user accessed the website, the plugin collected all content and user-related information.

4.3. Datasets

In order to gauge the ability of the recommender algorithm to differentiate between users and respond to the research questions presented above, four types of datasets are necessary: a 'pool dataset', a 'MyNews dataset', an 'editor's agenda dataset' and a 'clicks dataset'.

Table 3. Relative frequencies of news items categories in the three datasets (percentage within dataset).

News Category	Pool Dataset	MyNews Dataset	Editor's Agenda
Politics	4.1%	5.9%	8.7%
Economy	15.3%	9.9%	6.2%
International	13.2%	18.4%	9.6%
Public Health	5.8%	2.6%	7.7%
Science and Technology	0.4%	0.1%	5.2%
Local (Cyprus)	12.5%	18.3%	10.6%
Greece	4.3%	6%	6.9%
Sports	32.3%	32.1%	12.2%
Lifestyle/Gossip	5.5%	3.8%	19.3%
Other Hard News (e.g., energy)	3.8%	2.6%	1.1%
Other Soft News (e.g., viral, culture)	2.6%	0.4%	7.5%
Reader Content	0.1%	0	5.1%

4.4. Pool Dataset

This dataset contained all news items published by the medium for each of the ten days of data collection, regardless of whether users interacted with them or not. The purpose of this dataset is to provide a baseline against which the output of the recommender system can be judged; if, e.g., a user was only interested in a category of news that appeared very seldomly on the website, it is natural for any algorithm to not be able to provide correct matches to this user's behavior.

The website published an average of 207.17 news items per day (median of 214). Table 3 presents the relative frequencies of the various news categories that the different news articles belonged to, from which it can be gauged that a plurality of the articles published belonged in the 'sports' category, which is followed closely by local (Cyprus-related) news, economy-related news and international news. All remaining news categories contained half or less of the aforementioned.

4.5. MyNews Dataset

This dataset contains all articles included in the personalized 'MyNews' area generated for each user and per session by the recommender algorithm. While merely by observing the relative frequencies of news categories in 'MyNews', one can reach early conclusions, e.g., that sports-related news is by far the most frequently observed category (32.3%) followed by some distance by local and international news, this might be deceptive; after all, such behavior could be indicative of the recommender algorithm successfully delivering sport-related content to users who are primarily interested in sports.

More importantly, a number of observations concerning the algorithm's behavior could be made on the basis of data other than the news category indicating designer choices. First, the algorithm consistently recommended between 28 and 32 articles to each user in each session. Secondly, there were no duplications of news items within any session, as could be expected. More interestingly, observing the times that the various items contained in 'MyNews' and the time a user session started, it became obvious that a number of rules have been set for the recommender system concerning recency: roughly 50% of the recommended items had been published within 3.5 h of the session's start, while 75% had been published within the last ten hours. The maximal time allowable lapsing between the start of a session and the publishing of a news item was 1439 min, which is one minute from 24 h. It then becomes obvious that the system had been designed to only deliver news appearing during the last day. However, this is the only hard rule that can be safely deduced; there appears to be an additional bias toward more recent items, but it is impossible to pinpoint how this operates exactly.

4.6. Editor's Agenda Dataset

This dataset refers to news items that were contained in the frontpage of the website, which was consistently structured in the same manner to contain a number of political,

economy-related, sports-related items, etc. that always appeared in the same space on the website's frontpage. This dataset indicates the agenda of the editors making these choices.

4.7. Clicks Dataset

This dataset contains the news items clicked on by the aforementioned users participating in this experiment (and the relevant attributes). Users differed in both the amount of clicks they performed in the duration of the experiment and in the amount of sessions they engaged in. On average, users clicked a total 252.1 news items, although this is significantly skewed due to the instruction to *collaborating users* to perform a minimum of 25 clicks per session, the difference between *collaborating* and *ordinary users* being statistically significant (Wlicoxon W = 73, *p* = 0.044).

5. Results

5.1. Comparing the Editor's and MyNews Agendas

The first research question concerned the structure of the news collections (in terms of news categories) offered by the different areas of the website: the Editor's Agenda (the frontpage) determined by the editorial team, the personalized MyNews Agenda produced through the algorithm, and the total pool of articles produced by the media organization, from which the aforementioned two draw their content.

Even a quick perusal of Figure 6 that compares the relative frequencies (in percentage) of the categories that the various news items are assigned to suggests substantial differences between the two (omitted from the figure for easier visualization are the near-empty categories: "reader content", "other hard" and "other soft news").
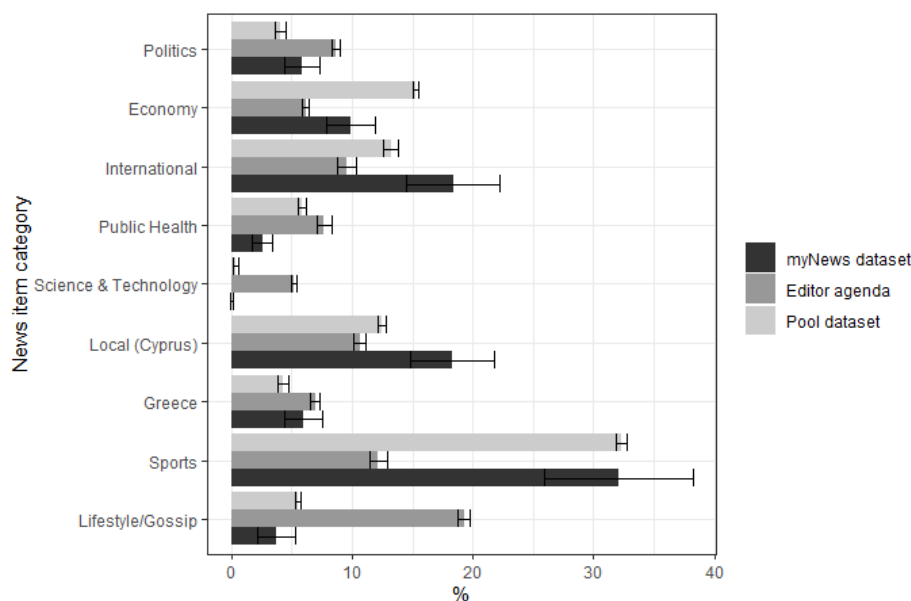

Figure 6. Relative frequencies of news items categories per dataset type.

The news environment produced by the two different processes, namely the editorial and algorithmic agendas, differ from each other. The editors have elected to follow a balanced approach with a mix of different news categories for their frontpage, with most categories being represented equally, covering between 5.2% and 12.2% of the available space. Interestingly, the exception to this rule concerns lifestyle and gossip-related items, which are overrepresented in the website's homepage (19.4%) particularly when contrasted with the relative amount produced in total (6.6% of all articles produced by the medium belong to this category). Other major categories of news (politics, international, local and sports) cover roughly the same large amount of space as would be expected from a mainstream news website aiming to cater for the diverse needs of a general audience.

To facilitate such comparisons, we take advantage of the fact that the overall pool of news items produced was collected; we divide the relative frequency (percentage) of each news category that appeared in the Editorial agenda (the frontpage) and in the MyNews area by the relative frequency of the categories of news items in the total pool of news items available for that session. If the result of this division (quotient) is 1, then the number of times a news item of the corresponding category in either the frontpage or the personalized MyNews section, is exactly as would be expected; were the two environments produced at random. Deviations from 1, on the other hand, suggest that either the algorithm or the editors who choose the frontpage items are favoring the specific news category (if the result of the quotient division is over 1) or that they are biased against the category (if the quotient is under 1).The boxplot (see Figure 7) examines exactly these ratios separately for the editorial and the MyNews agendas; the editorial agenda exhibits a soft news bias as shown from the overrepresentation of 'lifestyle and gossip' items. Similarly, news items tagged as relevant to 'Greece' and politics are over-represented, though much less so: about twice as many such items are contained on the website's frontpage (Editor agenda), as would have been expected by chance. Finally, concerning the editorial agenda, noteworthy is the larger than expected presence of 'public health'-related items, reflecting the ongoing COVID-19 pandemic. On the other hand, items dealing with the economy and international news are under-represented in the editorial agenda, as is the category of sports, although this is also the result of the large number of items in this category produced by the medium as a whole.
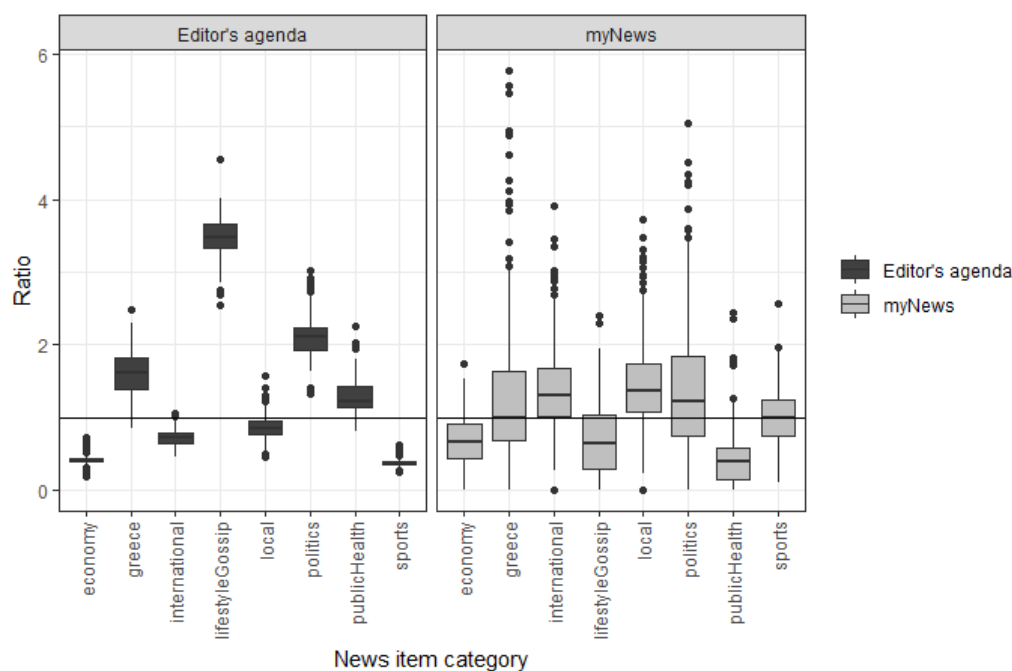

Figure 7. Boxplots of ratios of the Editor's Agenda and MyNews item categories to the available Pool items.

Concerning the MyNews area, we observe that the news stories offered are much closer to the distribution of categories in the total available pool of articles than the editorial agenda, with only economy- and public-health-related items being under-represented in the various MyNews agendas (personalized sessions) that different users in the sample viewed. However, it is not possible to reach any solid conclusions regarding an algorithmic bias from Figure 7, as it is produced on the basis of cumulative data from all users, who presumably have different interests—and the collaborating users among them, purposively so.

5.2. Evaluating Algorithmic Accuracy for Collaborating Users

While the aforementioned observations make it clear that there is a distinction between the news collection produced by the editors and the algorithm, they do not answer the second research question, i.e., whether the algorithm employed produces a distinct collection on the basis of deduced user interests. In order to examine this question, we need to take into account the type of user into our calculations.

We first focus only on data from *collaborating users*, who were instructed to only click on specific types of news (e.g., economy). While these users did not necessarily know the news category that the medium had assigned each item, it was hypothesized that they would be sufficiently accurate in clicking on only the 'right' (in accordance with the instruction they received) articles. An examination of the articles these users elected to click on suggests that this is indeed the case, since the relative frequency of clicking on the 'correct' type of article (i.e., where there was correspondence between instruction and the medium's assigned category) ranged between 78.8% and 92.7% of the articles viewed by these users (average 86%). These *collaborating users* then would be the easiest group for the algorithm to recommend, since they had highly discriminant behavior.

A dedicated examination of the MyNews/pool ratios (see Figure 7) for these users suggests that the algorithm indeed produced recommendations more likely to be clicked on by these users, with 'favored categories' being over-represented by a median of 2.16 times in the users' MyNews collections compared to the available pool, while 'non-favored categories' were under-represented by roughly 25% on average. However, this should not be taken to imply that the recommender system produces accurate recommendations; given these users' pre-designated behavior, their 'favored' categories should be over-represented to a much larger extent than observed, particularly during later sessions, when enough data on their behavior had been collected.

Indeed, on the basis of these data, we can construct a square matrix with rows representing the news categories collaborating users were instructed to click on and columns representing the categories of news items delivered to them by the algorithm averaged across sessions (Figure 8).
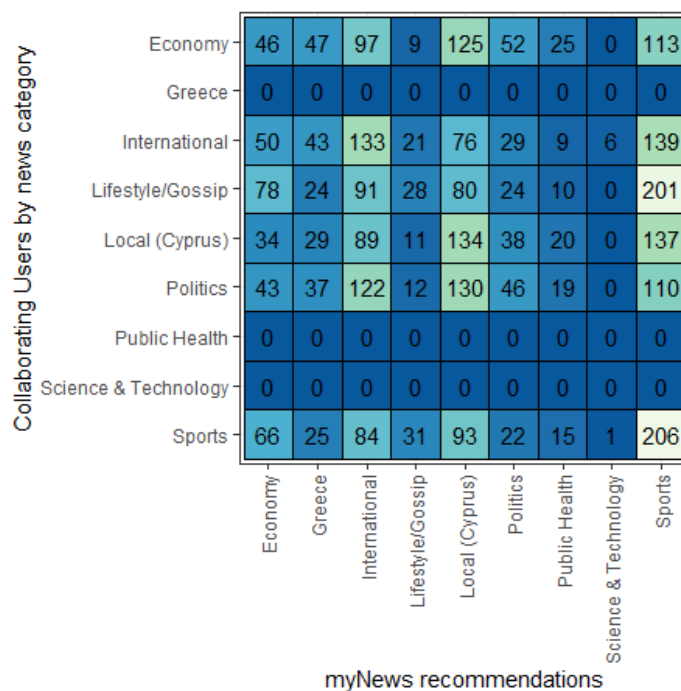

Figure 8. Matrix of frequencies of collaborating users.

This matrix can be re-composed to produce a per-news category confusion matrix for the dataset (Table 4), which can be used to produce various algorithm accuracy measures. The (unweighted) macro-averaged metrics for the algorithm was a precision of 0.211, recall of 0.19 and relevant F1-score of 0.2, confirming the aforementioned concerning the relative failure of the algorithm even for the collaborating users. The relevant overall accuracy metric was 0.747. In the calculation of these metrics, we omitted categories for which no collaborating user was designated ("Public Health", "Science and Technology" and "Greece"), since for these, no "True Positive" values could possibly exist.

Table 4. Re-composition of Figure 8 into a per-news category confusion matrix.

		True Positives	False Positives	False Negatives	True Negatives
	Economy	46	271	468	2325
	Greece	0	205	0	2905
	International	133	483	373	2121
News Category	Lifestyle	28	84	508	2490
	Local (Cyprus)	134	504	358	2114
	Politics	46	165	473	2426
	Public Health	0	98	0	3012
	Science and Technology	0	7	0	3103
	Sports	206	700	337	1867

The inability of the medium's algorithm to correctly include items that should have been recommended, at least for the relatively brief period of data collection, is also indicated by the lack of improvement in the algorithm model's accuracy score over time (Figure 9), as the flat linear trend indicates. So, while the algorithm did indeed differentiate between users, it did not create a news environment fully adapted to the assumed interests of these artificial, single-focused individuals. While this may indicate some faulty programming in the algorithm on the technical end, an alternative explanation can be offered. By design, the algorithm was required to select roughly 30 news items per session, which were additionally produced within the last 24 h (and more likely within 4 h of each session). However, the medium produced only, e.g., 11.3 and 16.4 politics- and lifestyle-related items as a whole per day. In other words, the algorithm failed to produce more politics-related items for the politics account, for the simple reason that it did not have enough such items available to pick from. The failure of the algorithm when it comes to the economy account is more perplexing, since roughly 41.8 such items are produced daily; the explanation for this might lie in the fact that these are mostly produced via the affiliated *EconomyToday* URL rather than the root website. It may be the case that the algorithm was designed to avoid cross-posting articles from such affiliated sites, although we have no means of ascertaining whether this is true on the basis of the data collected.

5.3. Evaluating Algorithmic Accuracy for Ordinary Users

Evaluating the algorithm's performance for the *ordinary users*, who received the instruction to choose whatever article they wished to view is less straightforward, since similar metrics cannot be calculated, as their interests are not known (as in the case of *collaborating users*). However, it is possible to construct a measure of distance between the items these users actually clicked on (indicating an overall 'profile') and the news collection presented to them by the algorithm in their MyNews area by calculating the distance between the relative frequency of articles in each category. Normalized, this index of distance takes values between 0 and 100, with larger numbers indicating greater distance between actual clicking behavior and the "MyNews" area. We can expect the algorithm to become better

at guessing users' behavior across time, since it has more data, and this should be reflected in smaller distances between the users' preferred environment (indicated by their clicking behavior in each session) and the algorithm's recommendations.
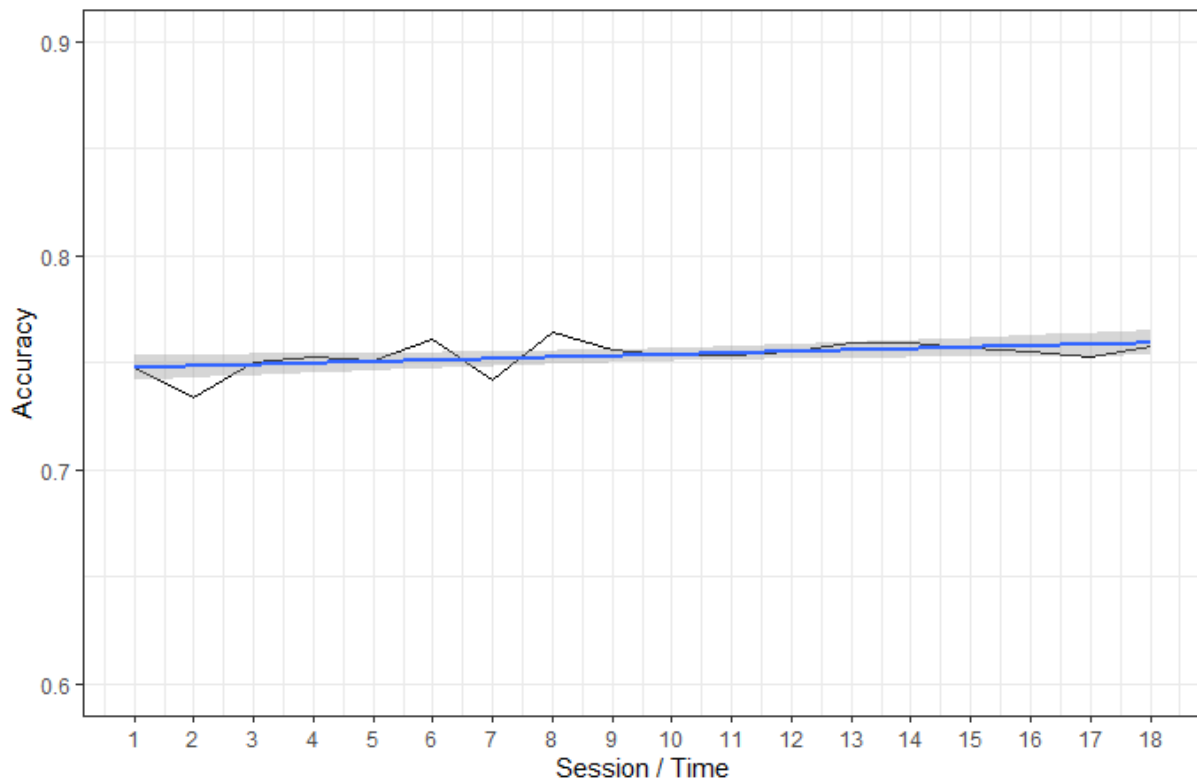

Figure 9. Accuracy score for *collaborating users* over time averaged across news type categories.

Figure 10 presents the linear trendline of exactly these distances across sessions for *ordinary users*, on average. While there is a decrease in the amount of distance between the pattern of clicking behavior and the news collection of the MyNews area, the decline is somewhat shallow with a decrease in distance of roughly 10% between the first and last sessions. However, the reader is reminded of the aforementioned design flaw in the algorithm (insufficient number of relevant articles in the pool the algorithm chooses from).

This relative absence of improvement in the algorithm's performance is also observable when considering the relative frequency with which *ordinary users* clicked on news items within their MyNews area. Results suggest that users did not choose news items from within the latter environment significantly more than they did at the beginning of the data collection period (Figure 10), suggesting a tentative negative answer to the third research question on whether the algorithm led to greater engagement.

5.4. Engagement over Time

After several sessions that would train the algorithm, one would expect that the algorithm would produce content that appealed to the users' interests and information needs, and therefore, the distance between media offerings in the MyNews area and user clicking behavior would diminish. Using the Euclidean distance metric, Figure 11 presents the polynomial trendlines of the distances across sessions for *collaborating users*, *ordinary users* and all users, on average starting from session 6 for each user. The failure of the algorithm to improve is apparent in the fact that these lines are relatively straight, whereas they would be expected to have a steep decline, although there is a small decline in distance following session 15, and the reader is reminded of the aforementioned design flaw in the algorithm (insufficient number of relevant articles in the pool the algorithm chooses from).
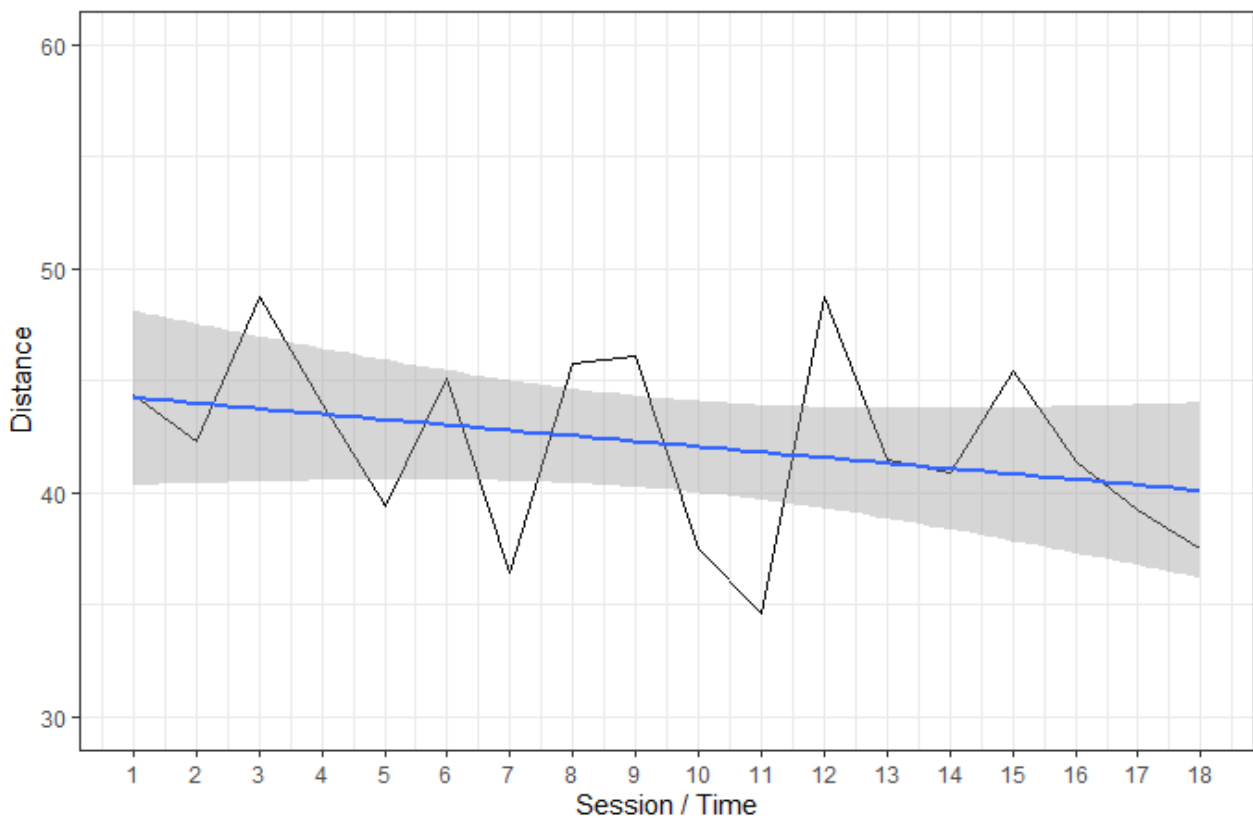
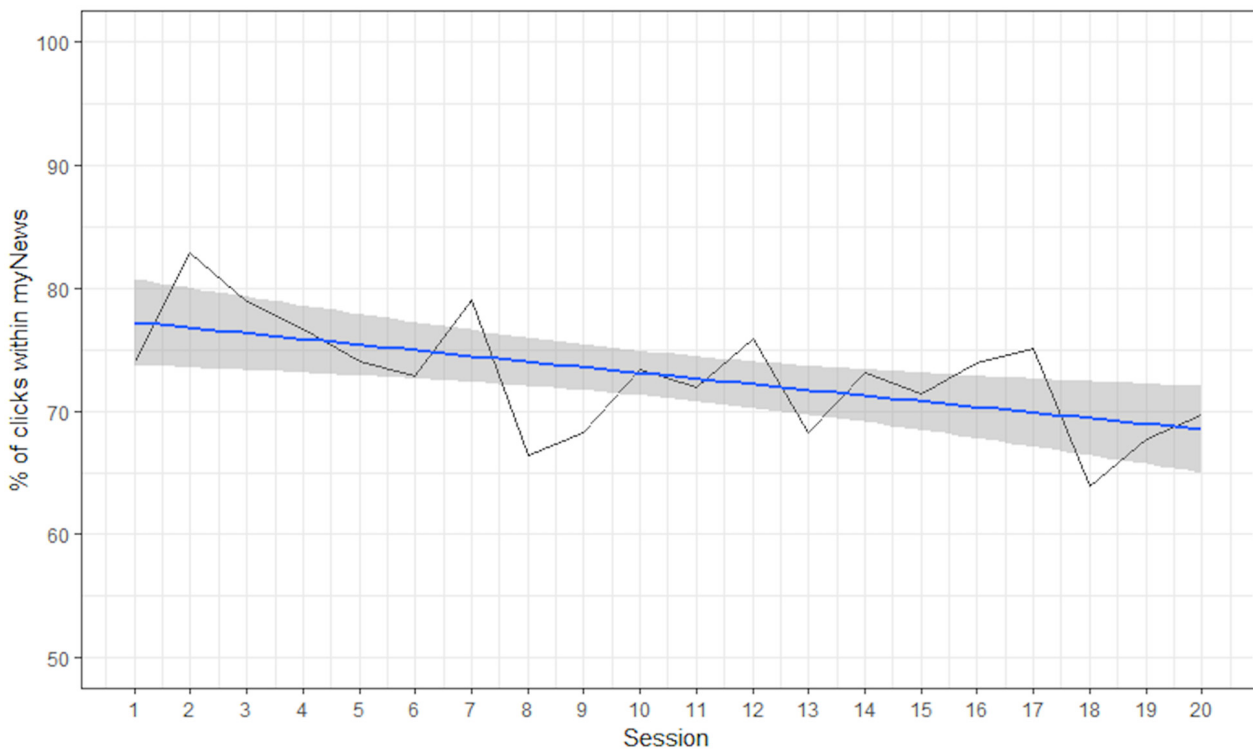**Figure 10.** Distance between profile based ordinary user clicking behavior and MyNews outputs over time.

Figure 11. Percentage of clicks ordinary users performed within the MyNews area over time.

6. Conclusions

This work attempted first to synthesize relevant work and propose a generic news recommendation system which goes beyond accuracy and considers the civic role of journalism for an informed citizenry. Second, it attempted to examine the application of a recommender system to a national-level mainstream news organization in a small European country, which was a decision driven by the participation of the medium in a funded research project. Regarding the first research question exploring the main differences between the Editor's Agenda and the algorithmic agenda, the study shows that the Editor's Agenda was based on a balanced mix of hard and soft stories aiming to cater for the diverse needs of a general audience. On the other hand, a large number of recommendations offered in the MyNews sessions were significantly influenced by the available pool of stories produced on a daily basis. In other words, the news diet offered in the MyNews area is shaped by an algorithmic logic as opposed to an editorial logic identified in the Editor's Agenda [66]. Thorough examination of the data provided by users specifically instructed to only view certain types of articles (*collaborating users*) and *ordinary users* instructed to consume news stories on their own will and interests suggests that the application of the recommender system was problematic. While the algorithm does provide recommendations significantly different from what would be expected by chance, it ultimately fails to produce a personalized environment populated primarily by news items of the type that a user would be expected to view on the basis of their past behavior; this finding holds true for both *collaborating* and *ordinary users* examined here.

A careful examination of the instances in which the algorithm fails the most suggests that this is the result of design flaws rooted in problematic rules. Our findings provide empirical evidence showing [48] that unintended side effects of design decisions undermine the accuracy capacity of algorithms. These design flaws can be summarized as follows:

1. Voracity: We introduce the term voracity to refer to the large number of news stories expected to be offered in the MyNews area per user per session. Given the relatively small output of the news organization altogether, the identified under-performance of the algorithm was partly due to its being required to populate the personalized area with too many news items (28–32 stories).
2. Recency: The recency metric needs to be used with caution. In our case, it was mostly driven by the nature of the news content produced by the medium: timely, short-form stories aiming to build traffic. However, the requirement that all personalized recommendations must have been produced within a day of viewing—or preferably within the last four hours—significantly hampered the algorithm's capacity to provide accurate and relevant recommendations.
3. Unsystematic tagging: The classification and ranking of news items depends greatly on the tags assigned to news stories. Evidence of unsystematic tagging of articles had a negative impact on the algorithm's capacity to make accurate offerings.
4. Underuse of available content: Although the news portal under study is affiliated with four other websites, this content was excluded from the algorithm's repository. Considering the ambitious expectations of the MyNews area, designers and editors should provision a greater pool of content or limit the quantity of stories provided in the MyNews area.

Finally, the decreasing engagement of users with the MyNews area can be associated not only with the problematic levels of accuracy but also with the random filling of the 30-items sessions—shaped primarily by the availability of content—as opposed to a more sophisticated algorithmic provision of recommendations including the principles of diversity, serendipity and novelty [33].

Overall, the findings provide evidence that the effective design of news recommender systems depends not only on the particularities of the news domain as a media offering but also on the special traits and ideology of the news medium implementing the recommender system. By special traits and ideology, we refer to the characteristics of the news content produced, including (a) the quantity of news items produced, (b) the style of news reporting

(e.g., short stories satisfying the value of immediacy and clickability or explanatory stories having a larger life-span), and (c) the scope of the recommender system [58] (e.g., aiming to generate clicks and short-term engagement or provide a more balanced and diverse news diet [35,47]). To sum up, when implementing NRS, two major conclusions are drawn: First, design decisions need to be carefully associated with both the scope and the production capabilities of the news organization. Some metrics, for instance recency, are commonly used in NRS, but news organizations need to carefully define the metric according to their own needs and capacity. Second, all input metrics need to be validated as a whole and not separately.

This study does not come without limitations stemming primarily from the limited timespan of the experiment and the low number of participants. On the other hand, the heterogeneity of the media landscape [67] has produced a diverse set of online news media calling for the need to decode heterogeneous and emerging needs and thus types of NRS. Under this assumption, the utility of the present study lies in revealing significant insights about other like-minded models: namely, medium-sized mainstream online media focusing on short-form, current affairs-type of journalism.

7. Future Research

News recommender systems select, filter, and personalize news content, thereby shaping the news diet and knowledge level of citizens [68]; their algorithms make highly consequential decisions and thus exercise significant power [69]. At the same time, news media are tasked with relaying information to citizens, setting an agenda of common concern, acting as watchdogs to the powerful and providing an arena for public deliberation [70]. The overarching question therefore is how to ensure that news recommender systems enact the civic values of journalism in the direction of an informed citizenry while catering for the commercial needs of news organizations. Addressing this question in a principled fashion requires technical knowledge of recommender design and operation, and it also critically depends on insights from diverse fields [71], including journalism studies, psychology, policy and law. Following our discussion on diversity, serendipity and novelty, one critical question is how to operationalize these values and turn them into metrics specifically for the case for news offerings and also decide on the resulting trade-offs between competing values, stakeholders and overall scope. In this direction, further design approaches are needed to enable news recommender systems to conform to specific values while considering the capabilities and needs of different types of media.

Author Contributions: Conceptualization, P.S. and D.M.; Data curation, C.D.; Formal analysis, C.D.; Methodology, P.S. and D.M.; Writing–original draft, P.S.; Writing–review & editing, P.S. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shearer, E. More than Eight-in-Ten Americans Get News from Digital Devices, Pew Research Center. Available online: https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/ (accessed on 21 January 2021).
2. Newman, N.; Fletcher, R.; Robertson, C.; Eddy, K.; Nielsen, R.K. Reuters Institute Digital News Report 2022. 2022. Available online: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-06/Digital_News-Report_2022.pdf (accessed on 25 September 2022).
3. Katsaounidou, A.; Dimoulas, C.; Veglis, A. *Cross-Media Authentication and Verification: Emerging Research and Opportunities*; IGI Global: Hershey, PA, USA, 2019. [CrossRef]
4. Myllylahti, M. Paying Attention to Attention: A Conceptual Framework for Studying News Reader Revenue Models Related to Platforms. *Digit. J.* **2020**, *8*, 567–575. [CrossRef]
5. Carlson, M. Confronting Measurable Journalism. *Digit. J.* **2018**, *6*, 406–417. [CrossRef]

6.   Petre, C. *All the News That's Fit to Click: How Metrics Are Transforming the Work of Journalists*; Princeton University Press: Princeton, NJ, USA, 2021.

7.   Nixon, B. The Business of News in the Attention Economy: Audience Labor and MediaNews Group's Efforts to Capitalize on News Consumption. *Journalism* **2020**, *21*, 73–94. [CrossRef]

8.   Dimoulas, C. Machine Learning. In *The SAGE Encyclopedia of Surveillance, Security, and Privacy*; Arrigo, B., Ed.; Sage Publications Inc.: Thousand Oaks, CA, USA, 2018; pp. 591–592. [CrossRef]

9.   Diakopoulos, N. *Automating the News: How Algorithms are Rewriting the Media*; Harvard University Press: Harvard, UK, 2019.

10.   Negroponte, N. *Being Digital*; Alfred, A., Ed.; Knopf, Inc.: New York, NY, USA, 1995.

11.   Thurman, N.; Lewis, S.; Kunert, J. Algorithms, Automation, and News. *Digit. J.* **2019**, *7*, 980–992. [CrossRef]

12.   Thurman, N.; Schifferes, S. The Future of Personalization at News Websites. *J. Stud.* **2012**, *13*, 775–790. [CrossRef]

13.   Van Drunen, M.-Z.; Helberger, N.; Bastian, M. Know your Algorithm: What Media Organizations Need to Explain to their Users about News Personalization. *Int. Data Priv. Law* **2019**, *9*, 220–235. [CrossRef]

14.   Raza, S.; Ding, C. News Recommender System: A Review of Recent Progress, Challenges, and Opportunities. *Artif. Intell. Rev.* **2022**, *55*, 749–800. [CrossRef]

15.   Bastian, M.; Makhortykh, M.; Harambam, J.; Van Drunen, M. Explanations of News Personalization across Countries and Media Types. *Internet Policy Rev.* **2020**, *9*, 220–235. [CrossRef]

16.   Diakopoulos, N. Computational News Discovery: Towards Design Considerations for Editorial Orientation Algorithms in Journalism. *Digit. J.* **2020**, *8*, 945–967. [CrossRef]

17.   Picard, R. Twilight or New Dawn of Journalism? *J. Pract.* **2014**, *8*, 488–498. [CrossRef]

18.   Bakker, P. Aggregation, Content Farms and Huffinization. *J. Pract.* **2012**, *6*, 627–637. [CrossRef]

19.   Smyrnaios, N. *Internet Oligopoly: The Corporate Takeover of Our Digital World*; Emerald Publishing: Bingley, UK, 2018.

20.   Rashidian, N.; Brown, P.; Hansen, E.; Bell, E.; Albright, J. Friend & Foe: The Platform Press at the Heart of Journalism, Tow Center for Digital Journalism, A Tow/Knight Report, 2018. Available online: https://www.cjr.org/tow_center_reports/the-platform-press-at-the-heart-of-journalism.php (accessed on 25 September 2022).

21.   Spyridou, L.-P.; Veglis, A. Sustainable Online News Projects: Redefining Production Norms and Practices. In *Redefining Multiplatform Digital Scenario and Added Value Networks in Media Business and Policy*; Vukanovic, Z., Powers, A., Tsourvakas, G., Faustino, P., Eds.; Media XXI Publishing: Lisbon, Portugal, 2015; pp. 63–83.

22.   Smyrnaios, N.; Rebillard, F. How infomediation platforms took over the news: A longitudinal perspective. *Political Econ. Commun.* **2019**, *7*, 30–50.

23.   Rashidian, N.; Tsiveriotis, G.; Brown, P.; Bell, E.; Hartsone, A. Platforms and Publishers—The End of an Era, Tow Center for Digital Journalism, A Tow/Knight Report. 2020. Available online: https://academiccommons.columbia.edu/doi/10.7916/d8-sc1s-2j58 (accessed on 25 September 2022).

24.   Vrysis, L.; Vryzas, N.; Kotsakis, R.; Saridou, T.; Matsiola, M.; Veglis, A.; Arcila-Calderón, C.; Dimoulas, C. A Web Interface for Analyzing Hate Speech. *Future Internet* **2021**, *13*, 80. [CrossRef]

25.   Vryzas, N.; Vrysis, L.; Dimoulas, C. Audiovisual speaker indexing for Web-TV Automations. *Expert Syst. Appl.* **2021**, *186*, 115833. [CrossRef]

26.   Bodó, B. Selling News to Audiences—A Qualitative Inquiry into the Emerging Logics of Algorithmic News Personalization in European Quality News Media. *Digit. J.* **2019**, *7*, 1054–1075. [CrossRef]

27.   Coenen, A. How The New York Times is Experimenting with Recommendation Algorithms, New York Times Open, 19 October 2019. Available online: https://open.nytimes.com/how-the-new-york-times-is-experimenting-with-recommendation-algorithms-562f78624d26 (accessed on 25 September 2022).

28.   Jannach, D.; Zanker, M.; Felfernig, A.; Friedrich, G. *Recommender Systems: An Introduction*; Cambridge University Press: Cambridge, MA, USA, 2010.

29.   Plattner, T. Why Personalization Will be the Next Revolution in the News Industry. Medium, December 15 2017. Available online: https://medium.com/jsk-class-of-2018/personalization-3a4cf928a875 (accessed on 25 September 2022).

30.   Broussard, M. *Artificial Unintelligence: How Computers Misunderstand the World*; MIT Press: Cambridge, MA, USA, 2018.

31.   Palomo, B.; Heravi, B.; Masip, P. Horizon 2030 in Journalism: APredictable Future Starring AI? In *Total Journalism*; Vázquez-Herrero, J., Silva-Rodríguez, A., Negreira-Rey, M.C., Toural-Bran, C., López-García, X., Eds.; Springer: Berlin/Heidelberg, Germany, 2022.

32.   Karimi, M.; Jannach, D.; Jugovac, M. News recommender systems—Survey and roads ahead. *Inf. Process. Manag.* **2018**, *54*, 1023–1227. [CrossRef]

33.   Møller, L.A. Recommended for You: How newspapers normalise algorithmic news recommendation to fit their gatekeeping role. *Journal. Stud.* **2022**, *23*, 800–817. [CrossRef]

34.   Fenton, N. Left out? Digital Media, Radical Politics and Social Change. *Inf. Commun. Soc.* **2016**, *19*, 346–361. [CrossRef]

35.   Helberger, N.; Karppinen, K.; D'Acunto, L. Exposure diversity as a design principle for recommender systems. *Information. Commun. Soc.* **2018**, *21*, 191–207. [CrossRef]

36.   Bernstein, A.; De Vreese, C.; Helberger, N.; Schulz, W.; Zweig, K.; Helberger, N.; Zueger, T. Diversity in news recommendation—Manifesto from Dagstuhl Perspectives Workshop 19482. *DagMan* **2021**, *9*, 43–61. [CrossRef]

37. Hendrickx, J.; Smets, A.; Ballon, P. News Recommender Systems and News Diversity, Two of a Kind? A Case Study from a Small Media Market. *Journal. Media* **2021**, *2*, 515–528. [CrossRef]
38. Helberger, N.; Bernstein, A.; Schulz, W.; De Vreese, C. Challenging Rabbit Holes: Towards more Diversity in News Recommendation Systems. Media@LSE. Available online: https://blogs.lse.ac.uk/medialse/2020/07/02/challengingrabbit-holes-towards-more-diversity-in-news-recommendation-systems/ (accessed on 25 September 2022).
39. Pariser, E. *The Filter Bubble: What the Internet is Hiding from You*; Viking: New York, NY, USA, 2011.
40. Lafrance, A. The Power of Personalization, Nieman Reports, Fall 2017. Available online: https://niemanreports.org/articles/the-power-of-personalization/ (accessed on 25 September 2022).
41. Gillespie, T. The Relevance of Algorithms. In *Media Technologies: Essays on Communication, Materiality, and Society*; Gillespie, T.P.J., Boczkowski, K.A.F., Eds.; The MIT Press: Cambridge, MA, USA, 2014; pp. 167–193.
42. Just, N.; Latzer, M. Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet. *Media Cult. Soc.* **2017**, *39*, 238–258. [CrossRef]
43. Dylko, I.; Dolgov, I.; Hoffman, W.; Eckhart, N.; Molina, M.; Aaziz, O. The Dark Side of Technology: An Experimental Investigation of the Influence of Customizability Technology on Online Political Selective Exposure. *Comput. Hum. Behav.* **2017**, *73*, 181–190. [CrossRef]
44. Beam, M.A. Automating the News: How Personalized News Recommender System Design Choices Impact News Reception. *Commun. Res.* **2014**, *41*, 1019–1041. [CrossRef]
45. Dutton, W.H.; Reisdorf, B.C.; Dubois, E.; Blank, G. Search and Politics: The Uses and Impacts of Search in Britain, France, Germany, Italy, Poland, Spain, and the United States. Available online: https://ora.ox.ac.uk/objects/uuid:2cec8e9b-cce1-4339-9916-84715a62066c (accessed on 25 September 2022).
46. Bruns, A. *Are Filter Bubbles Real?* Polity Press: Cambridge, MA, USA, 2019.
47. Vrijenhoek, S.; Kaya, M.; Metoui, N.; Möller, J.; Odijk, D.; Helberger, N. Recommenders with a Mission: Assessing Diversity in News Recommendations. In Proceedings of the 2021 Conference on Human Information Interaction and Retrieval, Canberra, Australia, 14–19 March 2021; The Association for Computing Machinery: New York, NY, USA; pp. 173–183. [CrossRef]
48. Diakopoulos, N. Algorithmic Accountability. *Digit. J.* **2015**, *3*, 398–415. [CrossRef]
49. Möller, J.; Trilling, D.; Helberger, N.; Van Es, B. Do not Blame it on the Algorithm: An Empirical Assessment of Multiple Recommender Systems and their Impact on Content Diversity. *Inf. Commun. Soc.* **2018**, *21*, 959–977. [CrossRef]
50. Diakopoulos, N.; Koliska, M. Algorithmic Transparency in the News Media. *Digit. J.* **2016**, *5*, 809–828. [CrossRef]
51. Kompan, M.; Bieliková, M. Content-Based News Recommendation. Available online: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.299.4505&rep=rep1&type=pdf (accessed on 25 September 2022).
52. Bozdag, E. Bias in Algorithmic Filtering and Personalisation. *Ethics Inf. Technol.* **2013**, *15*, 209–227. [CrossRef]
53. Lu, Z.; Dou, Z.; Lianz, J.; Xiez, X.; Yang, O. Content-Based Collaborative Filtering for News Topic Recommendation. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Available online: https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9611/9247 (accessed on 25 September 2022).
54. Darvishy, A.; Ibrahim, H.; Sidi, F.; Mustapha, A. *HYPNER: A Hybrid Approach for Personalized News Recommendation*; IEEE Access: Piscataway, NJ, USA, 2020.
55. Li, L.; Wang, D.D.; Zhu, S.; Li, T. Personalised news recommendation: A review and an experimental investigation. *J. Comput. Sci. Technol.* **2011**, *26*, 754–766. [CrossRef]
56. Feng, C.; Khan, M.; Rahman, A.U.; Ahmad, A. *News Recommendation Systems Accomplishments, Challenges & Future Directions*; IEEE Access: Piscataway, NJ, USA, 2020; Volume 8, pp. 16702–16725.
57. Lavie, T.; Sela, M.; Oppenheim, I.; Inbar, O.; Meyer, J. User Attitudes towards News Content Personalization. *Int. J. Hum.-Comput. Stud.* **2010**, *68*, 483–495. [CrossRef]
58. Smets, A.; Hendrickx, J.; Ballon, P. We're in This Together: A Multi-Stakeholder Approach for News Recommenders. *Digit. J.* **2022**, 1–19. [CrossRef]
59. Ekstrand, M.D.; Harper, F.M.; Willemsen, M.C.; Konstan, J.A. User perception of differences in recommender algorithms. In Proceedings of the 8th Conference on Recommender Systems, RecSys'14, Silicon Valley, CA, USA, 6–10 October 2014; pp. 161–168.
60. Reuver, R.; Mattis, N.; Sax, M.; Verberne, S.; Tintarev, N.; Helberger, N.; Atteveldt, W. Are We Human, or are We Users? The Role of Natural Language Processing in Human-centric News Recommenders that Nudge Users to Diverse Content. In Proceedings of the 1st Workshop on Nlp for Positive Impact, Online, 7 December 2022; pp. 47–59. [CrossRef]
61. Thaler, R.; Sunstein, C. *Nudge: Improving Decisions about Health, Wealth, and Happiness*, 1st ed.; Yale University Press: New Haven, CT, USA, 2008.
62. Vermeulen, J. To Nudge or Not to Nudge: News Recommendation as a Tool to Achieve Online Media Pluralism. Available online: https://www.tandfonline.com/doi/abs/10.1080/21670811.2022.2026796 (accessed on 25 September 2022).
63. Pasquale, F. *The Black Box Society: The Secret Algorithms that Control Money and Information*; Harvard University Press: Cambridge, MA, USA, 2015.
64. Diakopoulos, N. Towards a Design Orientation on Algorithms and Automation in News Production. *Digit. Journal.* **2019**, *7*, 1180–1184. [CrossRef]

65. Sandvig, C.; Hamilton, K.; Karahalios, K.; Langbort, C. Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms, Paper presented to "Data and Discrimination: Converting Critical Concerns into Productive Inquiry", a Preconference at the 64th Annual Meeting of the International Communication Association, Seattle, WA, USA, 22 May 2014. Available online: http://www-personal.umich.edu/~{}csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf (accessed on 25 September 2022).
66. Bucher, T. *If . . . then: Algorithmic Power and Politics*; Oxford University Press: Oxford, UK, 2018.
67. Splichal, S.; Dahlgren, P. Journalism between de-professionalisation and democratization. *Eur. J. Commun.* **2016**, *31*, 5–18. [CrossRef]
68. Moeller, J.; De Vreese, C. Spiral of Political Learning: The Reciprocal Relationship of News Media Use and Political Knowledge Among Adolescents. *Commun. Res.* **2019**, *46*, 1078–1094. [CrossRef]
69. Lundahl, O. Algorithmic meta-capital: Bourdieusian analysis of social power through algorithms in media consumption. *Inf. Commun. Soc.* **2022**, *25*, 1440–1455. [CrossRef]
70. Schudson, M. *Why Democrecies Need an Unlovable Press*; Polity: Cambridge, UK, 2008.
71. Stray, J.; Halevy, A.; Assar, P.; Hadfield-Menell, D.; Boutilier, C.; Ashar, A.; Beattie, L.; Ekstrand, M.; Leibowicz, C.; Moon Sehat, C.; et al. Building Human Values into Recommender Systems: An Interdisciplinary Synthesis. *arXiv* **2022**, arXiv:2207.10192. [CrossRef]