Where Are the Values? A Systematic Literature Review on News Recommender Systems

CHRISTINE BAUER, Paris Lodron University Salzburg, Salzburg, Austria
CHANDNI BAGCHI, Utrecht University, Utrecht, Netherlands
OLUSANMI A. HUNDOGAN, Utrecht University, Utrecht, Netherlands
KARIN VAN ES, Utrecht University, Utrecht, Netherlands

In the recommender systems field, it is increasingly recognized that focusing on accuracy measures is limiting and misguided. Unsurprisingly, in recent years, the field has witnessed more interest in the research of values “beyond accuracy.” This trend is particularly pronounced in the news domain where recommender systems perform parts of the editorial function, required to uphold journalistic values of news organizations. In the literature, various values and approaches have been proposed and evaluated. This article reviews the current state of the proposed news recommender systems (NRS). We perform a systematic literature review, analyzing 183 papers. The primary aim is to study the development, scope, and focus of value-aware NRS over time. In contrast to previous surveys, we are particularly interested in identifying the range of values discussed and evaluated in the context of NRS and embrace an interdisciplinary view. We identified a total of 40 values, categorized into five value groups. Most research on value-aware NRS has taken an algorithmic approach, whereas conceptual discussions are comparably scarce. Often, algorithms are evaluated by accuracy-based metrics, but the values are not evaluated with respective measures. Overall, our work identifies research gaps concerning values that have not received much attention. Values need to be targeted on a more fine-grained and specific level.

CCS Concepts: • General and reference → Surveys and overviews; • Information systems → Recommender systems; • Human-centered computing;

Additional Key Words and Phrases: Recommender systems, human values, systematic literature review, news recommendation

ACM Reference Format:

This publication has been supported in part by AI Labs Utrecht University in collaboration with DPG Media. Additionally, it has been supported by the EXDIGIT (Excellence in Digital Sciences and Interdisciplinary Technology) project, funded by Land Salzburg under grant number 20204-WISS/263/6-6022.

Authors’ addresses: C. Bauer (Corresponding author), Department of Artificial Intelligence and Human Interfaces, Paris Lodron University Salzburg, Salzburg, Austria; e-mail: christine.bauer@plus.ac.at; C. Bagchi and O. A. Hundogan, Utrecht University, Utrecht, Netherlands; e-mails: chandni.bagchi@gmail.com, o.hundogan@gmail.com; K. van Es, Department of Media and Culture, Utrecht University, Utrecht, Netherlands; e-mail: K.F.vanEs@uu.nl.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs International 4.0 License.

© 2024 Copyright held by the owner/author(s).
ACM 2770-6699/2024/06-ART23
https://doi.org/10.1145/3654805
1 INTRODUCTION

Recommender systems (RS) pervade our everyday life: Many online platforms integrate such systems to help users discover relevant items such as movies [46], fashion [45], jobs [111], or social matching [212]. Essentially, RS are a means to help users deal with information and choice overload [9] by recommending items that might be interesting to the user; often, such recommendations are personalized to the user [196, 198].

While the optimization of accuracy in RS has been a long-standing focus—thus, increasing a recommender algorithm’s performance in accurately predicting a user’s rating—there is growing awareness that relying solely on accuracy metrics is restrictive and misguided [75, 146]. In recent years, the RS field has witnessed more interest in research that goes “beyond accuracy” [3, 75, 103]. In fact, Kaminskas and Bridge [103] identified a shift in RS research towards including beyond-accuracy objectives. Their survey demonstrates that the most extensively studied and integrated beyond-accuracy objectives include diversity, serendipity, novelty, and coverage. These, we suggest, can thus be considered the “standard” beyond accuracy values in RS research. Then, later, the focus shifted beyond just those standard beyond-accuracy metrics towards a broader range of values. For instance, using the term “value-aware RS” [32], research has paid attention to the business value of RS [98]. Often, such works focus on optimizing the economic value of recommendations by balancing the interests of multiple stakeholders [1]. For a recent literature review on value-aware RS, see De Biasio et al. [32]. However, this strong business orientation does not necessarily embrace the wide spectrum of values beyond economic and utility aspects. Interestingly, works considering and investigating a wide range of values in RS rarely use the term “value-aware.” Instead, these works typically (only) specify those values they are concentrating on (e.g., privacy [99], fairness [61], trust [34]) or subsume some values under other overlapping umbrella terms and concepts (e.g., ethics [150, 225]).

Paying attention to values when designing systems is not restricted to RS and is not a recent idea at all, as it traces back to the 1980s and earlier. Algorithms are often perceived as objective procedures for solving problems [125]. However, adopting this technical perspective overlooks the fact that algorithmic systems are socio-technical in nature [206]. Culture and cultural nuances play an important role in how and why these systems function as they do [206]. In other words, technologies reflect the values of the cultures in which they are made [36, 143, 249]. This recognition in the 1980s and 1990s laid the foundation for the development of approaches such as Value Sensitive Design (VSD) [66–68]—a concept that was popularized in the information systems and human-computer interaction fields. VSD centers on the engagement and balancing of human values in the design process of technologies. In this context, the term “value” has been broadly defined as “what a person or group of people consider important in life” [68].

Within RS research, news is a specific domain in which values have received considerable attention. In part, this attention is because news recommender systems (NRS) are a part of the editorial function of news organizations and need to uphold journalistic values [218]. With the increasing spread of false and misleading information (“fake news”) [8, 262], the demand for considering and acknowledging journalistic values has become louder and more evident. This example illustrates that the importance of certain values is also domain-dependent. For instance, journalistic values are a crucial cornerstone in the news sector but are less relevant in other areas (e.g., in games). Moreover, (relevant) values are not only domain-specific but can also be specific to an organization or product. As Bastian et al. [14] point out,

“[...] value interpretations and prioritization can vary between news organizations and even individual practitioners. An important implication of this finding is that responsible, value-aware use of and implementation of [news recommender systems] require
news organizations to engage internally in an organization-wide process of identifying their core values with regard to news recommender [systems’] use, which can subsequently inform their strategies to achieve value-sensitive design.” [14, p. 855]

Given that values can vary by domain and may be specific to an organization or product, it would be expected that academic discourse and practice would encompass a broad range of values and variations of approaches in NRS.

As previous works (e.g., References [14, 25, 88, 90, 155, 214]) stress, values are essential in the journalistic process. Indeed, as we demonstrate in this article, a range of values and methodologies has been proposed and assessed within the academic discourse on NRS. However, while knowledge about a wide set of values expands, these are scattered across papers and research communities. We tackle this research gap through a systematic literature review, analyzing and synthesizing 183 papers on NRS. We seek to trace and reflect on the scale, research fields, and range of values in papers on recommender systems within the news domain.

The systematic review offers three key contributions: First, from our analyzed corpus, we identified a total of 40 values and developed a categorization scheme to group these values into five value groups. Second, our review synthesizes the body of research on value-aware NRS, tracing its development back to 1995. This includes an overview of the research approaches and metrics employed in this research, not only on a general level but also in relation to specific value groups and individual values. Third, we highlight the prolific authors and author teams on value-aware NRS. Our work’s novelty lies in its focus on value-aware NRS and embracing an interdisciplinary perspective.

The remainder of this article is organized as follows: In Section 2, we examine related work. In Section 3, we discuss how we selected and categorized existing research for consideration in our review. Section 4 provides insights into the development, scope, and focus of value-aware NRS over time. Section 5 offers a discussion of the identified trends. The conclusion section explores potential avenues for future research (Section 6).

2 RELATED WORK

In this section, we start by discussing the specifics of the news domain concerning the integration of values (Section 2.1). Subsequently, we provide a brief overview of research on NRS (Section 2.2) and discuss the motivation to target objectives “beyond accuracy” (Section 2.3).

2.1 Specifics of the News Domain

In the news domain, the role of values in RS has attracted considerable attention. This specific focus is largely due to the recognition that news plays a crucial role in supporting democratic functions. As such, algorithmic personalization has sparked much concern within this domain about so-called “filter bubbles” [174] and “echo chambers” [223]. Some raise concerns that news recommenders might exacerbate political divisions among individuals and potentially harm the development of an informed public. Evidence supporting the filter bubble hypothesis is, however, limited [31, 149]. For example, research by Nechushtai and Lewis [164] found that users from various states and political leanings were recommended similar news items, undermining the idea that algorithms necessarily create echo chambers. Nonetheless, they observed a high degree of homogeneity and concentration in the news recommendations, indicating that popular news providers are reinforced in popularity. This observation raises an important question about the desired role of NRS.

Bastian et al. [14] interviewed media practitioners (e.g., journalists, data scientists, and product managers) from quality newspapers in the Netherlands and Switzerland to gain insights into how they perceive algorithmic NRS and to understand what values they consider important in the design of these systems. The study revealed that media practitioners believe that NRS should not
be exempt from upholding journalistic values. For them, it is important that values such as transparency, diversity, editorial autonomy, a broad information offer, personal relevance, usability, and surprise are taken into consideration in how these systems are designed and implemented [14]. The news organization that Lu et al. [136] worked for participated in the research of Bastian et al. [14]. They conducted further research seeking to identify values that were both desirable and technically feasible to implement in these systems, identifying two: (i) timely and fresh content and (ii) surprising readers. The former was modeled as dynamism and the latter as serendipity. Importantly, Lu et al. [136] demonstrated that introducing dynamism into NRS can be achieved without sacrificing accuracy. This finding challenges the common assumption that incorporating values into such systems necessarily involves a tradeoff. Instead, their study suggests that pursuing multiple objectives simultaneously in NRS is feasible.

2.2 News Recommender Systems

NRS have been the subject of several review papers. Our work stands apart from these previous studies in several key aspects: (i) Focus on value-aware NRS: To date, no other literature review focuses on value-aware NRS specifically. (ii) Little overlap of references: The overlap of cited references with other systematic reviews of news recommenders is low: 16% of our references also appear in Karimi et al. [106], 14% in Raza and Ding [195], 12% in Mitova et al. [151], and 7% in Feng et al. [64]. The overlap of references with other survey papers is marginal (e.g., De Biasio et al. [52] 3%, Özgöbek et al. [171] 3%, Li and Wang [131] 2%, Qin and Lu [190] 2%, Borges and Lorena [26] < 1%, and Dwivedi and Arya [60] < 1%). Our review article features 158 references (i.e., 59% of our references) that have not been covered by any of the aforementioned survey papers. (iii) Interdisciplinary view: Our literature review covers references investigating NRS from different angles, including papers from computer science and journalism alike.

Most literature reviews on NRS specifically review the underlying algorithmic approaches [26, 128, 171]. Karimi et al. [106] additionally focus on empirical evaluation and the users’ perception of the systems. As Raza and Ding [195] point out, these reviews generally take the perspective of computer scientists (e.g., References [26, 60, 64, 106, 109, 131, 171]). A notable exception is a more recent review by Mitova et al. [151] that takes a political communication perspective on NRS, synthesizing findings concerning journalistic distribution and audience acquisition of political information for democracy and identifying research gaps. In contrast to these previous reviews, our review takes an interdisciplinary approach. Our primary goal is to examine the development, scope, and focus of value-aware NRS over time. We specifically aim to identify and scrutinize the spectrum of values discussed and incorporated in NRS research. Additionally, we investigate the metrics employed to optimize and assess these values within the systems.

Our study, which concentrates on elements beyond accuracy, aligns closely with surveys by Karimi et al. [106] and Raza and Ding [195]. Karimi et al. [106], in their survey of 140 papers published between 2005 and 2016, examine the general challenges, algorithmic approaches, and methodological issues related to the evaluation of NRS. They observe a steady increase in the number of papers on NRS throughout this period, suggesting that it has become an important subtopic within RS research. While their findings indicate that the primary optimization goal in NRS research is to accurately predict relevance for news readers, Karimi et al. [106] note that this approach is often not optimal and explain this by providing vivid examples:

“If, for example, a user is interested in politics and has shown interest in articles about an ongoing presidential election in the past, recommending more articles about this topic is probably a good choice. However, recommending solely articles about the election, or solely about politics, might be too monotonous for users and would probably
not lead to high user engagement in the future. In case of news aggregation site, it is furthermore important that the recommended news are not too similar to each other. Presenting three articles from three different sources about, e.g., the same plane accident might be of little value for users.” [106, p. 1209]

Given the limitations of an accuracy-centric perspective, it has become increasingly important to consider other quality aspects in NRS research [75, 106, 195]. To balance accuracy, it is crucial to consider quality aspects such as diversity, novelty, and serendipity alongside traditional accuracy metrics. These qualities are often discussed as beyond-accuracy aspects in the broader RS literature [75, 103]. Karimi et al. [106] note that from around 2011 onward, a growing number of NRS papers consider beyond-accuracy aspects. However, their work also emphasizes that much work still needs to be done [106]. Significantly, they raise an important critique of the work being done, “while some papers take aspects like diversity or novelty into consideration in the design process of their algorithm, they do not explicitly quantify any improvements w.r.t. these aspects with standard metrics in their experimental evaluation.” [106, p. 1214]

This highlights the importance of ensuring that the measures used in RS align with the intended goals [234].

Raza and Ding [195] seek to broaden this perspective by conducting a survey that not only examines the technical aspects of NRS but also investigates the effects of these systems on user behavior. Additionally, they explore the development and application of deep learning in the news domain. They also highlight the importance of using beyond-accuracy aspects in evaluating the quality of news recommenders, “typical accuracy-centric approaches may fail to consider other aspects of user experiences (such as choice satisfaction, perceived system effectiveness, better recommendations, and exposure to different points of view) when evaluating the recommendation quality.” [195, p. 3]

Their survey finds that accuracy remains a standard evaluation measure for the quality of NRS. Furthermore, they also conclude that although some research has been done on diversity [195, p. 16], a very limited number of works investigate novelty, coverage, and user experience.

In the following section, we discuss beyond accuracy more broadly in RS literature.

2.3 Beyond Accuracy

As mentioned, optimizing accuracy has commonly been the primary goal in RS research [98]. Typically, RS research has relied on a standard set of accuracy-based metrics, including Precision and Recall [21, 84], to evaluate a recommender systems’ success. Already in 2004, Herlocker et al. [92] wrote about matching the evaluation of RS to user needs. They postulated that recommendations should not just be accurate but also useful (e.g., recommending bananas to people in grocery stores is too obvious to be useful) and claim, “[w]e need comprehensive quality measures that combine accuracy with other [aspects such as] serendipity and coverage, so algorithm designers can make sensible trade-offs to serve users better” [92]. Importantly, there are many aspects of user satisfaction that accuracy-based metrics are unable to measure [267].

McNee et al. [146] later went as far as to claim that the narrow focus on improving accuracy in RS has actually hurt the field. They argue that having a high level of accuracy in an RS does not necessarily mean it is effectively aiding users in discovering items that genuinely interest them. They provide the need for a user-centric perspective that is pleasurable rather than helpful or simply accurate. Adamopoulos [3] underscores their plea, stating that many existing RS have focused “on providing more accurate rather than more useful recommendations.”
As outlined above, it is increasingly acknowledged that values other than accuracy play a significant role in improving the overall quality of an RS. Kaminskas and Bridge [103] survey the most widely discussed beyond-accuracy objectives: diversity, serendipity, novelty, and coverage. Diversity in recommendations ensures that the recommendations are not too similar. Kaminskas and Bridge [103] argue, with reference to the field of information retrieval, that diversifying retrieval results could potentially lead to increased user satisfaction. This is because an exclusive focus on maximizing retrieval accuracy may result in too-similar recommendations. They argue that sometimes accuracy needs to be sacrificed for increased user satisfaction. Serendipity in recommendations allows users to encounter in unplanned ways what they find interesting [22], whereas novelty denotes items previously unknown or new to the user. Finally, coverage concerns the extent to which recommendations cover the full range of available items in the catalog.

3 METHODS

With this study, we seek to understand which values have been addressed in RS research in the news domain and when and how they have been discussed. To do so, we rely on a systematic literature review [114].

The main motivations for conducting a systematic literature review on value-aware NRS are as follows: Papers on value-aware NRS are scattered across a wide variety of outlets with different aims, scopes, and target audiences across various research communities. A systematic literature review is a promising method for rigorously synthesizing the existing body of knowledge on a well-defined topic [116, 122, 230]. Furthermore, a systematic literature review’s explicit, rigorous, and reproducible procedure allows to reduce biases [19, 116, 230]. As such, a systematic literature review is a natural choice to target our research goal.

For the literature review, we systematically searched for papers on NRS, narrowed the scope to papers concerned with values, and analyzed the research landscape. Figure 1 illustrates the procedure, which we describe in the following subsections. First, we detail the literature search (Section 3.1) and the corresponding criteria for selection (Section 3.2), followed by an explanation of the coding process (Section 3.3).

3.1 Literature Search and Criteria

For the literature search, we followed the systematic literature review procedure according to the guidelines by Kitchenham et al. [114]. The search strategy to identify papers to be included in our sample consisted of several consecutive stages, illustrated in Figure 1.

First, we performed a scoping review of relevant published literature to develop an effective search strategy. From a comparison of search results using the databases Scopus, ACM Digital Library, Wiley Online Library, EBSCO, Web of Science, IEEE Xplore, and WorldCat, all of which contain papers relevant to technology and computer science, we concluded that the search results of Scopus also contain the papers from the other databases. Springer Link turned out to be inefficient for our research, because it mostly produced results that were outside our project’s scope, which was detrimental to the search; for this reason, it was omitted from the search. Beyond the technology and computer science angle (e.g., papers appearing in conference proceedings of RecSys or SIGIR), the search in Scopus also resulted in relevant papers taking a news and journalism perspective (e.g., papers in the journal Digital Journalism) and embracing a broader scope of digital sciences (e.g., papers appearing in the conference proceedings of CHI, CHIIR, or Hypertext). Accordingly, we sampled papers found in Scopus, where we searched in an unspecified time frame.

As our literature review explicitly focuses on NRS, we searched for papers indexed with the keywords news recommendation or news personalization considering spelling variations. Thus, we searched for the search terms news recommend*, news personal*, or personali* news in the title or
Where are the Values? A Systematic Literature Review on News Recommender Systems

Fig. 1. PRISMA diagram detailing the paper selection process.

the keywords. The search string was determined in a process of trial-and-error in which we tried various combinations. We compared the number of results and the relevancy of the results per search string. The search string that provided papers relevant to our research was finally selected. The query syntax looks as follows\(^1\):

\[
\begin{align*}
\text{( TITLE ("news recommend\*" OR "news personali\*" OR "personali\* news") OR}
\end{align*}
\]

\[
\begin{align*}
\text{KEY ("news recommend\*" OR "news personali\*" OR "personali\* news") )}
\end{align*}
\]

\[
\begin{align*}
\text{AND ( LIMIT-TO ( SRCTYPE, "p") OR LIMIT-TO ( SRCTYPE, "j") )}
\end{align*}
\]

\[
\begin{align*}
\text{AND ( LIMIT-TO ( DOCTYPE, "cp") OR LIMIT-TO ( DOCTYPE, "ar") )}
\end{align*}
\]

\[
\begin{align*}
\text{AND ( LIMIT-TO ( LANGUAGE, "English") )}
\end{align*}
\]

We chose to search for English-language conference papers and articles in journals and conference proceedings. As a result of our query on 19 April 2022, Scopus rendered 609 papers.

3.2 Data Cleansing and Selection of Papers for the Sample
As the aim of this review was to identify values beyond accuracy within the news domain, the four authors investigated the 609 retrieved papers and reviewed them against the inclusion and exclusion criteria described below.

\(^1\)DOCTYPE indicates the document type, which is conference paper (cp) and article (ar). SRCTYPE indicates the source, which are journal (j) and conference proceedings (p).
The criteria for inclusion were that the papers must report on values other than accuracy. A paper was considered out of scope (and, thus, excluded) if any of the following criteria was met (exclusion criteria):

- The paper’s core contribution was about the production of news articles.
- The paper’s core contribution was about evaluation methods of NRS.
- The document resembled a project proposal.
- The document was a dissertation.
- The document was a collection of conference papers (e.g., workshop proceedings).²

Reasons for these exclusion criteria reflect that our review focuses on news recommendations to news consumers rather than providing support for journalists and editors in the news article creation process. Furthermore, papers describing evaluation methods in general do not contribute to our work’s focus, namely, values in NRS. Moreover, we consider only peer-reviewed research papers as a quality criterion; this excludes research proposals and editorials (e.g., editorials to conference proceedings). This also refers to dissertations, which undergo a peer-review process similar to research papers in conferences and journals; the dissertations’ contents are frequently also published as research papers, which are part of the sample.

The four authors screened the retrieved 609 papers against these criteria by examining titles, abstracts, and of the main text, mainly the “Results” sections. For this task, the papers were divided among the authors, sorted in chronological order, and split into two halves; one half consisted of papers published before 2016 and the other half published after 2016. Two authors screened one half of the papers independently from each other, while the other two authors screened the other half independently. Disagreement about excluding certain papers was first resolved among the two authors assigned to the respective bulk of papers and subsequently discussed with all four authors to reach a unanimous consensus. This procedure led to the exclusion of 365 papers. The exclusion of many sources was based on the criterion that they did not pertain to values but instead concentrated solely on accuracy or click-through rate, which did not align with the focus of our research. After this process, 244 papers were left.

3.3 Review of the Selected Papers in Full Text (Coding)

The four authors reviewed the 244 papers in full text.

The following coding scheme was developed inductively from raw data:

- Type of paper (algorithmic work, conceptual, user experiment, interview(s), review)
- Additional details about the previous category (type of evaluation)
- Domain (financial, sports, etc.), if applicable
- Platform (social media, Twitter, video, mobile app, etc.)
- Problem statement
- Datasets
- Values
- Metrics

The four authors categorized the papers according to the coding system. Similar to the screening process mentioned above, we divided the papers among the authors. This time, the papers were sorted alphabetically by the surname of the paper’s first author before being divided into two halves. We altered the pairings of authors; two authors coded one half of the papers, and the other two authors the other half, again independently from each other. Subsequently, the coding was

²Note that proceedings and editorials to proceedings were excluded. Papers in such proceedings that were retrieved in our search—and met the inclusion and exclusion criteria—were included.
discussed, and conflicts were resolved to reach a unanimous decision. In this process, a duplicate was found, 3 papers were not accessible, and 55 were found irrelevant. (For example, Usher [232] reports on empirical research concerning how start-ups in the news domain differ from traditional journalism; thus, not focusing on recommender systems. Said et al. [203] describes a production recommender system infrastructure that allows research systems to be evaluated in situ, as an effort to move evaluation methodology forward.) Finally, 183 papers remained, making up our final corpus for analysis.

When coding the papers, we identified a total of 40 values. The coding process revealed that some values are closer to each other than others. For instance, some value codings required discussions among the raters to reach unanimous decisions. Based on this observation and given the large number of different values, it appeared adequate to aggregate values into categories. Subsequently, three authors aggregated the identified values into five categories in a joint iterative process. Each of the 40 identified values was written down on a sticky note. The authors sat together in this process so all had a fresh reminder and an overview of the identified values. The sticky notes were put next to each other on a big table so all were visible; we refer to this as the pool of values. Then, taking turns without a specific order, the team members relocated sticky notes to the table’s lower end to group them if they were considered similar.

While doing so, the team member explained why those values were considered similar to each other (or similar to the already grouped ones). The sticky notes were left grouped if the other team members (temporarily) agreed. If they disagreed, then an explanation for disagreement was provided, and the moved sticky notes went back to where they had been before or back to the pool of values if the group did not get support from at least two team members anymore. This process was iterative, with some sticky notes being moved back-and-forth between groups and the pool multiple times. Based on the grouping explanations, the team created a label for the respective group. The respective label was written down on a separate sticky note in a different color and put next to the sticky note group. This, too, was an iterative process, and several groups were relabeled several times. All sticky notes remained visible throughout the session. With some values and value groups, this led to many discussions, which—ultimately—also resulted in the emergence of subgroups. For example, there was heavy discussion about whether the values in the now-named value group “responsible agency” should form a separate value group or be listed among the values of the “responsibility” group. The unanimous decision was that the values “agency,” “autonomy,” and “future impact” should form a subgroup of the value group “responsibility,” as these have the underlying theme of giving a person agency. Furthermore, this process also led to the merging of values. For instance, “censorship” and “instrumentalization” and “propaganda” were merged into “objectivity”; “shifting user interests” and “interests over time” were merged into “temporality of interests.” In addition, the label for value group “editorial values” emerged after intense discussion. First, we tended to label this value group “journalistic values.” However, as “journalistic values” forms a distinct value within the very same value group and other authors (e.g., References [136, 218]) subsume several values under “editorial values,” we chose the “editorial values” for the group label. We detail the identified values in Section 4.3, where we also present the categorization scheme (Figure 4).

The coding scheme allowed for coding a paper for multiple attributes within a category. For instance, a paper may discuss five different values; hence, all were coded. Also, regarding the type of paper, multiple codings were possible. However, almost all papers were only one type (e.g., conceptual work). There were only 7 papers that have been double-assigned in terms of their paper types: Epure et al. [62] and Li et al. [129] are categorized as algorithmic work and analysis;

---

3We contacted the authors but have not received a response.
Viana and Soares [236] and Jain et al. [96] are characterized as both algorithmic work and user study; Bastian et al. [15] is categorized as both conceptual work and a review paper; Krebs et al. [117] is categorized as both conceptual and analytical work; Wang et al. [242] is algorithmic and conceptual work.

4 RESULTS

Before considering what values are discussed in the literature and how value-aware NRS are evaluated, we briefly review the number of articles on NRS and values specifically. Here, we also identify the types of papers produced and explore the prolific authors publishing on values in NRS.

4.1 General Overview

Figure 2 provides an overview of the number of papers published over time—on NRS in general and value-aware NRS in particular. The rhombus-shaped symbols (in red) represent the papers focused on NRS overall, and the circle-shaped symbols (in blue) refer to the subset of papers in which we identified values beyond accuracy. The rhombus-shaped (red) symbols in the graph reveal an increase in the total number of papers published on NRS from 1995 to 2022. There was an initial surge in the number of papers observed after 2008, followed by another sharp increase from 2016 onward. Raza and Ding [195] conjecture that the increase in the later years (from 2016 onward) might be credited to both the CLEF NEWSREEL Challenge,\(^4\) [30] providing resources for the evaluation and optimization of news recommenders, as well as the emergence and development of RS based on deep learning, which happened around that same time. Further, Raza and Ding [195] claim that the higher number of publications on NRS in 2021 is linked to the release of the benchmark dataset MIND (by Microsoft) [252].

While the absolute number of papers on value-aware NRS appears at first sight to remain relatively stable (Figure 2), having a closer look at the proportion of value-aware NRS papers compared to the overall number of published papers on NRS paints a different picture. Over the years, the percentage of value-aware NRS papers (compared to overall NRS papers) varies greatly. In short,

\(^4\)https://www.newsreelchallenge.org/
while the number of NRS papers has grown rapidly from 2008 onward, the number of value-aware NRS papers did not grow proportionally. As Figure 2 shows, the absolute number of value-aware NRS papers has increased in the past two decades. The first value-aware NRS paper in our corpus was published in 1995, after which there was a six-year gap until the next value-aware NRS paper. Value-aware NRS papers started picking up in 2010, with the number gradually increasing over time. A noticeable decline in the number of value-aware NRS papers occurred in 2018, which corresponds to a general stagnation in the publication of NRS papers during the same year.

Looking into the research approaches taken to investigate values in NRS, we found that most research takes an algorithmic approach (108 papers, 59%), focusing on the development of algorithms. Compared to this, only a few works take a conceptual (30 papers, 16.4%) or analytical approach (23 papers, 12.6%).

To clarify, in contrast to the algorithmic works, analytical works do not necessarily introduce new algorithms. Instead, they prioritize the evaluation and comparison of existing recommender approaches concerning specific values. Conceptual works, in contrast, neither implement the criteria algorithmically nor analyze existing approaches; instead, they reflect on values—their relevance, need, and conceptualization.

Compared to algorithmic, conceptual, and analytical works, user studies do not appear often (15 papers, 8.2%). Review papers (9 papers, 4.9%) and interview-based research (5 papers, 2.7%) are even rarer. While the user studies centered on news consumers, only one of the 5 interview-based works focused on news readers (i.e., Reference [85]). Instead, 4 of the 5 interview-based works were conducted through interviews with people employed or active in the news domain (e.g., References [14, 24, 59]) and with parliament members or party officials (e.g., Reference [80]). For values, we argue, it is crucial to listen and “check” with news readers and practitioners about their concepts of and experiences of these values rather than impose our own assumptions about them in the implementation of recommender systems. However, as indicated, this type of work is lagging behind.

Figure 3 depicts the temporal evolution of these approaches in the publications. Despite annual fluctuations, reflective in part of the dynamics inherent in academic publishing, it underscores the finding that algorithmic work predominantly characterizes the literature on NRS dealing with values. It furthermore demonstrates that reviews, analyses, and conceptual work gradually constitute a growing proportion of the overall output. The increased prominence of conceptual work is particularly notable here compared to the other two categories.

The distribution of these different research approaches raises questions about possible gaps in NRS research. Considering that the number of algorithmic work significantly outweighs conceptual work, the perhaps most fundamental question is the following: Do journalistic teams have a different understanding of these values than the tech teams responsible for the development of NRS? Additionally, the limited number of user studies and interviews raises questions such as: Do we know how users experience these values? Are they even aware of the values embedded into these systems? What are the specific expectations and needs of practitioners in the news domain? Are these expectations and needs being met?

4.2 Prolific Authors on Value-aware News Recommendation

To yield insights regarding who does research on values in NRS and the status of these publications within RS research, we ranked the top authors concerning their number of publications in our corpus (Table 1). We used five publications as the cutoff point.

---

5 Seven papers (i.e., References [15, 62, 96, 117, 129, 236, 242]) have double-assignments concerning type of paper; thus, $n = 190$. 
Regarding the question of who, we identified three clusters of collaborators in this ranking. The first cluster concerns the scholars Balaji Padmanabhan and Shankar Prawesh, who published on value-aware NRS in our corpus together—between 2011 and 2015—when they both worked at the Department of Information Systems and Decision Sciences at the University of South Florida, Muma College of Business, USA. These authors primarily consider manipulation-resistant NRS, discussing the problem of the self-reinforcing nature of “most popular” type lists.

The second cluster centers around Natali Helberger from the Institute of Information Law at the University of Amsterdam, The Netherlands. In our corpus, Helberger has several (co-)authored
publications from 2018 to the present. These works tackle the democratic role of NRS and reflect on values—mainly diversity. These publications are linked to Helberger’s PersoNews project\(^6\) (2015–2021) on the impact of personalized news for democracy, funded by the European Research Council. Mykola Makhortykh worked as a postdoctoral researcher in Data Science at the Amsterdam School of Communication Science. Makhortykh was connected to this project, too, studying algorithmic (un)fairness in news personalization systems. Finally, Mariella Bastian worked as a postdoctoral researcher at the Institute for Information Law on the PersoNews project.

Third, we observe a cluster with Jon Espen Ingvaldsen, Özlem Özgöbek, and Jon Atle Gulla. They all work at the Department of Computer and Information Science at the Norwegian University of Science and Technology, Norway. Together, they have published on context-aware, user-driven news recommendation, the intricacies of time in news recommenders, and user-controlled news recommenders. Ingvaldsen and Gulla published about NRS in relation to location awareness and geographical proximity, mostly around 2015. Gulla and Özgöbek published together on topics such as exploratory news recommendations and interactive mobile news recommenders. They also published on news recommenders with other co-authors.

By setting a cut-off point of five papers, we aim to spotlight academics who have consistently engaged with the subject matter, showing a continuing line of inquiry, and thereby distinguishing them from those who have contributed more occasionally. Earlier, we described that, despite an uptick in publications related to NRS, the proportion of studies focused on values has not kept up with this growth. We have also identified that only three groups of authors account for 20% (49 out of 244) of the papers on value-aware NRS within our dataset. Furthermore, these publications are connected to large research funding, indicating that research in this domain is still limited in the studied time frame. Collectively, these findings underscore the niche and ad hoc nature of the field. In a more mature field, the threshold for identifying the top researchers based on the number of publications would have been considerably higher.

4.3 Identified Values and Value Groups in News Recommendation

A core interest of this review is to bring to light the range of values discussed and considered in NRS research. In our study, we adopted a broad definition of values in accordance with Value Sensitive Design (VSD). This definition encompasses all factors deemed important. In our sample of 183 papers, we identified 40 values. In an iterative process (see Section 3.3), this multitude of values could be aggregated into five categories (value groups). Figure 4 presents an overview of the identified values and value groups, which we present and discuss in the following:

The first value group is termed standard values (90 occurrences). It embraces the values diversity, popularity, novelty, and coverage. These values are considered “standard,” because—as discussed earlier—they are the most discussed beyond-accuracy values. Rather unsurprisingly, these values are also widely discussed within the field of NRS.

The responsibility values (88 occurrences) embrace privacy, explainability, accountability, transparency, trust, fairness, and manipulation prevention. These values point to news providers’ responsibility towards their users and society. In addition, this value group includes responsibility values that specifically concern providing users the opportunity to act and, thus, form a subgroup of responsibility values. This subgroup, termed responsible agency, encompasses autonomy, agency, and future impact.

The user experience (UX) values (81 occurrences) refer to values concerned with aspects that primarily target how users experience the NRS. Our sample features the following UX values: temporality of interests, engagement, user satisfaction, curiosity, emotion, serendipity, fatigue, surprise,
and convenience. While these UX values are essentially not news-domain-specific, they are also relevant in the news domain. We note that serendipity is also one of the standard beyond-accuracy values widely discussed in the RS field in general [103]. Still, we hold with Smets et al. [215] that “serendipity should be understood as a user experience rather than a mere offline evaluation metric such as diversity or novelty.” Hence, our categorization scheme includes serendipity in UX values.

The editorial values (91 occurrences) reflect an organization-centered perspective and encompasses freedom, objectivity, authority, credibility, democracy, journalistic values, and editorial influence. These values are inherent in the news domain and strongly associated with journalistic and editorial values (for details, see Bastian et al. [14] and Lu et al. [136]). Lu et al. [136] emphasize that these values must be considered when implementing RS in the news domain. Similar to the responsibility values, a subgroup emerged for the editorial values. This subgroup termed recommender-specific encompasses context, location, recency, relevancy, and saliency. While these values embrace editorial values—and are, thus, integrated into this group—the recommender-specific ones form a subgroup, as these are specifically instrumental within the context of RS.

The technical values (15 occurrences) embrace values associated with the technical operation of an RS: scalability, real-time capabilities, personalization, utility, and tradeoff. We note that many values may conflict with technical values; thus, improving all of them is challenging or infeasible. When researchers recognize and acknowledge that several values must be considered despite potentially creating tradeoffs, optimizing such tradeoffs can be considered a value on its own.

Table 2 provides an overview of the total number of occurrences of values summed up per value group (2nd column) and the number of papers that address at least one value in the respective value group (3rd column). Overall, Table 2 indicates that editorial values are featured most often (91 occurrences), followed by standard and responsibility values (90 and 88 occurrences, respectively). UX values (81 occurrences) are featured only marginally less. In comparison, technical values are mentioned to a far lower extent (15 occurrences). As explained, it is unsurprising that the standard values are featured often.

The high number of occurrences for editorial values pinpoints that domain-specifics are important. As we show in Section 4.4 (particularly Figure 5), responsibility values gained particular attention from 2019 onward. By comparison, technical values receive very little attention in value-aware NRS research. There are several possible explanations for this. One possibility is that, in some cases, other values may take precedence over technical values, leading to their relatively lower occurrences.
lower emphasis. Additionally, technical values may not be questioned, as they are already part of a functioning system. Last, it is plausible that technical values are simply of greater importance in domains other than news.

While the second column in Table 2 presented the total number of occurrences of values summed up per value group, the third column presents the number of papers that address at least one value in the respective value group. Standard values and editorial values (75 papers, respectively) are also the most covered value groups in this regard, whereas technical values (13 papers) are addressed the least. For standard, UX, and editorial values, the number of papers is only slightly lower than the number of occurrences in the respective value group (standard values: 90 occurrences in 75 papers; UX values: 81 occurrences in 66 papers; editorial values: 91 occurrences in 75 papers). Technical values are mostly addressed individually—thus, one at a time (15 occurrences in 13 papers); rarely together in one paper. However, regarding the responsibility values, we see stark differences, because values in this group appear 88 times, yet across only 50 papers. This observation points out that the values in this group are often addressed together within one paper.

### 4.4 Value Groups in the Discourse on News Recommender Systems

Diving into the identified value groups (as described in Section 4.3), we see interesting publishing patterns over time (Figure 5). The value groups discussed from early publications onward are the editorial and the standard values. This observation is expected, because these values represent the basic tenets of the news domain and the RS field, respectively. From 2004 onward, papers about responsibility and UX values start surfacing. Increasingly, almost all value groups receive more attention in the early 2010s. The only exception is the technical values group, which remains stable throughout; we note that we identified a total of only 13 papers addressing those values, and due to this limited number, this observation is inconclusive. Notable is the rise in editorial values starting in 2013 and declining in 2018. Around the same time, a surge can be observed for UX values. Another interesting development is the increased discussion of responsibility values in 2019. More broadly, from 2018 onward, we observe a shift towards responsibility, standard, and UX values rather than editorial values. This trend corresponds to overall developments in research—also outside the news domain; for instance, with the establishment of the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT),\(^7\) launched that year, we witness a general acknowledgment of responsibility values in the research and development of algorithmic systems.

To get an overview of which values have been implemented in algorithmic approaches and which ones have been discussed on a conceptual level, Figure 6 shows the value groups per paper type. While all value groups feature in algorithmic work, responsibility values occur less often

---

\(^7\)https://facctconference.org. Note that, in 2018, the conference’s name was FAT*. The conference was affiliated with ACM in 2019. After the 2020 conference, the conference changed its name to ACM FAccT.
in algorithmic work (16 occurrences) compared to UX values (43 occurrences), editorial values (42 occurrences), and standard values (39 occurrences). Instead, in conceptual work, responsibility values are the most represented value group (19 occurrences). Standard values (14 occurrences) and editorial values (12 occurrences) range in the middle in conceptual works; other value groups appear in conceptual works to a far lower extent (6 occurrences and lower). From this observation, we infer that while responsibility values have been integrated, discussion on a conceptual level is still needed.

Unsurprisingly, the standard value group is featured in algorithmic work a lot (39 occurrences) because these are considered “standard” already. In addition, these values are discussed in conceptual works (14 occurrences), indicating ongoing research to tease out those values on a conceptual level.
Similar to the standard values (39 algorithmic papers and 14 conceptual works), editorial values are strongly featured in algorithmic work (42 occurrences), while there is an ongoing discussion in conceptual work (12 occurrences).

Further, we note that there are, in total, only 15 papers with user studies. Interestingly, though, these cover all value groups.

4.5 Values within Value Groups

Having discussed the values in the publications on NRS (Section 4.4), this section examines the values within the value groups in more detail. Figure 7 illustrates the number of occurrences of individual values per value group. This figure shows that diversity far outnumbers other values in terms of occurrences (62 occurrences). The second most popular value is recency (39 occurrences), followed by temporality of interests (26 occurrences).

Besides recency, location (19 occurrences) is featured substantially among the editorial values, followed by democracy, context, and objectivity (10, 7, and 6 occurrences, respectively). Still, the latter three values (i.e., democracy, context, and objectivity) are far less often considered than the overall most popular ones. We note that these are more abstract concepts than recency and location, which could explain why these are, in comparison, considered less.

Among the responsibility values, transparency and trust are considered most often (18 occurrences each). Interestingly, these two values concern features that are often discussed in the context of RS in general: A lack of trust and transparency are the aspects that make people less receptive to RS [169, 244]. Especially when it comes to NRS, these values are vital [102, 200] because—from a democratic perspective—trust in news is essential for the ideal of the informed citizen [50]; and the increasing spread of “fake news” [217] has an impact on people’s trust in media [108, 245].

Further, from Figure 7, we see that the high popularity of standard values is primarily due to diversity (62 occurrences), which is the most featured value overall. Aside from diversity, novelty (15 occurrences) and coverage (12 occurrences) are featured often, compared to diversity to a far lower extent. Coverage is a value that is relevant for the news provider, and novelty is what keeps news readers interested [136].

Within the UX values group, temporality of interests (26 occurrences), serendipity (20 occurrences), and emotion (15 occurrences) are tackled most often. In comparison, other UX values appear far less frequently. It is interesting to observe that serendipity is frequently considered in the context of NRS. This observation indicates that it is considered important that NRS do not only provide news readers with the news they want to read but also serve users in a way they do not necessarily expect. Shifting user interests is a topic that RS have to account for in general [58, 243, 257]. As the news domain is highly concerned with recency and (unexpected) real-life events, user interest shifts may follow different patterns in the news domain compared to other domains.

The technical values rarely occur in value-aware NRS papers (15 occurrences in 13 papers). Real-time capabilities and utility are addressed several times (5 occurrences, respectively), whereas tradeoff is only once.

As diversity is the most addressed value (62 occurrences), we detail what diversity embraces. In general, RS literature frequently calls for more diverse recommendations and suggests diversification approaches to ensure a certain level of diversity in the recommendations (e.g., References [4, 78])—however, without necessarily specifying how diversity should manifest. In other respects, some works address a particular type of diversity (e.g., Ziegler et al. [267] specifically addressed topic diversification in a book recommendation setting). Notably, various types of diversity are addressed in the news domain. Topic Diversity occurs the most (34 occurrences), which indicates

---

8For surveys on diversity in recommender systems, see Kaminskas and Bridge [103] and Kunaver and Požrl [121].
that NRS researchers are interested in showing news readers a variety of topics. The second diversity type is viewpoint (19 occurrences), which is very topical in the news domain. A valid concern around NRS is that they might show news readers only articles from one (political) viewpoint. Offering users viewpoint diversity may be a productive counter to that and address fears of, for instance, increased societal polarization. Compared to topic and viewpoint diversity, the other diversity types (i.e., diversity concerning sources, people, events, semantics, sentiments, authors,
Where are the Values? A Systematic Literature Review on News Recommender Systems

4.6 Evaluating Value-aware News Recommender Systems and Measuring Values

The most popular evaluation setup used in the value-aware NRS papers is offline evaluation (73), followed—with a large gap—by laboratory study (27), online evaluation (18), and simulation (8). It is important to note that not every kind of work includes an evaluation (e.g., most conceptual works and reviews). A large majority of papers on RS (in general) employ offline evaluation (see, e.g., References [20, 54]), which is similarly reflected in our sample. This indicates that the evaluation of value-aware NRS most often does not involve interaction with real users; instead, it relies on predicting preference or behavior based on historical data. While it has its eligibility and benefits as a controlled environment and for establishing baselines, it is limited to reflecting past rather than current and future behaviors and preferences. This is particularly disputable in the news domain, where—by definition—recency plays a critical role. Moreover, it uses simplified user models that do not capture the complexity of responses and the various factors influencing how these recommendations are experienced. Particularly when it comes to integrating values into RS, it is crucial to involve real users in the evaluation process. This is to avoid drawing and amplifying existing assumptions based on correlations. For instance, with regard to diversity: Do users perceive the implemented diversity?

With regard to the relation between evaluation types and value groups, there are some interesting patterns that warrant further research, beyond the scope of this article. For instance, offline evaluation is used to a far lower degree when responsibility values were involved (6 occurrences) compared to standard, UX, and editorial values (29, 29, and 28 occurrences, respectively). Moreover, only two papers used online evaluation when responsibility values were involved. Note that standard, UX, and editorial values were evaluated with online evaluation only slightly more (5, 7, and 9 occurrences, respectively). Interestingly, laboratory studies are on similar levels for standard, responsibility, UX, and editorial values (10, 12, 12, and 8 occurrences, respectively). Given the limited number of papers addressing technical values in our sample, it is not possible to draw conclusions in that regard. Additionally, the small number of instances (only 10) where simulation was used as an evaluation approach makes it difficult to derive meaningful observations regarding this method.

Accuracy-based metrics clearly dominate in our sample on value-aware NRS. Throughout the timeline, the accuracy family, which includes accuracy, precision, recall, and F1, is consistently used the most (in total, 115 occurrences). This observation that accuracy is the most-used measure in NRS has also been found in other reviews (e.g., Reference [195]). Also, click-through rate (CTR) and normalized discounted cumulative gain (nDCG) are frequently used (20 and 25, respectively). It is striking that diversity is heavily addressed in algorithmic work (note: 34 of the 109 algorithmic works address diversity). However, intra-list similarity (ILS) [267], which is the widely used measure for diversity [100], is relatively rarely used in our sample (11 papers). Beyond these, further metrics occur in our sample, yet scarcely—often only once—(e.g., hit rate (HR), mean reciprocal rank (MRR), distortion, Jaccard similarity, Gini coefficient, root mean square error (RMSE)).

To sum up, accuracy-based measures remain dominant in the field, even with regard to the news domain. CTR and nDCG gain attention from 2013 onward. Although used in the past few years, other measure types have not gained momentum. In this light, we also propose that, although a wide range of values is considered in papers on NRS, these values are not being evaluated (which is also in line with the observation by Karimi et al. [106, p. 1212]). Reflecting on this matter, van Es
et al. [234] point to the importance of aligning concept, design, and evaluation. For instance, Helberger et al. [91] explore how diversity can be conceptualized in very different ways. Each conceptualization implies a different operationalization and benchmarks and metrics. These evaluation metrics are, however, merely proxies in that they stand in for the concept it tries to capture [159, p. 4]. This means that there can be disagreement on what the right benchmarks and metrics should be. Ideally, relevant stakeholders (e.g., computer scientists, journalists, advertisers) find common ground in how alignment should be achieved.

5 DISCUSSION

In Section 4, we presented an overview of the literature on value-aware NRS. This section raises discussions prompted by these findings.

First, Raza and Ding [195, p. 16] found that there is some effort to introduce diversity in news recommendation but very limited work on novelty, coverage, and user experience. Our findings underscore that diversity is the most published beyond-accuracy value of all. As explained by Helberger and team [25, 89–91, 154], this value is of high concern within the news domain, as it is linked to policy objectives and normative ideals. However, this means other values are relatively understudied. Is this disproportional attention for diversity really warranted? Or is it a fairly “easy” value to implement into recommender systems? To make matters even more complex, as touched on earlier, does the incorporation of these values necessarily involve tradeoffs?

Second, there is tension between the abstractness (complexity) of values and their operationalization (simplification). There are indications that many values that are discussed in the literature are oversimplified and detached from their original meaning. Values are complex phenomena that need to be targeted on a more fine-grained and specific level to understand to which extent they are embedded in RS or what implications RS have on specific values. We have found that it is now being done in the case of diversity, where researchers look into specific sub-dimensions. However, here, most work is on topic and viewpoint diversity, glossing over many other forms of diversity.

Breaking down these values into more specific and actionable components is a task that needs to be undertaken for other values as well. Currently, some values are too broad or coarse, and a more detailed examination is required to make them practical and applicable. As a consequence, it is also difficult to measure those. For instance, explainability: What is explained (procedure or outcome)? To whom is it explained or explainable (a user or the developers)? In what level of detail is it explained? While there is work in this direction (see Zhang and Chen [263]), this is seemingly not happening in the news domain.

A third observation relates to the “stability” of the values that are being implemented. For instance, our review suggests that some values (e.g., editorial values and standard values) have frequently been integrated into algorithmic work, while it is also tackled on a conceptual level. Indeed, it is a well-studied phenomenon that values undergo changes over time [248], whereby these changes are not necessarily due to time effects, but rather emanate from time-invariant contextual influences [229]. In the context of our review, this suggests that early algorithmic works might not be capturing and integrating values in the same way as later work, because these values were conceptualized (on a deeper level) only later. This raises questions concerning the comparability of works, as conceptualizations of values may vary over time.

Fourth, as pointed out by Stray [218], lots of works focus on principles—thus, “written descriptions of the values that technical systems should uphold” [218]—for news recommenders rather than metrics, evaluation, datasets, and feedback. Our findings suggest that many of these value-ware NRS are evaluated by accuracy metrics. However, we need to use metrics that are
aligned with the goals. As Jannach and Bauer [97] suggest, the RS research community has fallen prey to what they call the McNamara fallacy: a focus on quantitative and easy-to-take measures in offline experiments. As such, the effectiveness of the algorithms in practice remains unknown.

Then, there is the lack of understanding whether users actually perceive the values in question. Does making recommendations more diverse indeed increase user satisfaction? Recent work outside the NRS field suggests a discrepancy between measures and human perception [100]. In the news domain, there is—to date—only one small study on this issue, suggesting that body text similarity is most representative of human perception (compared to, e.g., the similarity of authors or images) [216]. More research is required to understand whether users perceive specific values and whether that correlates to greater satisfaction.

Finally, designing (N)RS designed in a value-aware fashion raises ethical questions. Should we nudge users towards "healthier" news consumption? As Helberger et al. [91] explain,

> "influencing people’s choices, even for good and legitimate reasons, can sit at odds with users’ conceptions of personal autonomy, freedom from manipulation and privacy. This is even more so if diversity-sensitive design is used to realize more normative, societal objectives, such as serving democratic discourse rather than the interests of individual users." [91, p. 201]

Answering this requires reflective discussions about different stakeholder values and how these are balanced in designing these systems. The inability of users to detect certain biases in recommended news invites conversations about transparency, responsibility, accountability, and explainability. With traditional news outlets, the public is aware of their ideological slant, and their editors are typically willing and able to publicly discuss why certain editorial decisions were made. Studying users’ perceptions and needs concerning the role of NRS in this context requires more user surveys and interviews.

6 CONCLUSIONS

With this systematic literature review, we have traced and reflected on the scale, research fields, and range of values discussed and engaged with in the scientific discourse on recommender systems in the news domain. Our review suggests that value-aware NRS is still an under-researched area of interest, particularly within computer science. We observed that although values gain more attention in NRS research, it still constitutes a relatively small and ad hoc "field" and has not grown proportionally with the RS field as a whole. This concluding section summarizes the main findings regarding values and news recommender systems.

In our review, we found that most value-aware NRS research has taken an algorithmic approach. Conceptual papers, analytical works, review papers, user studies, and interview-based research are far rarer. This suggests a possible research gap concerning users’ experiences of values and alignment between editorial and tech staff on what these values mean.

Moreover, the driving force bringing attention to value-aware NRS seemingly comes from fields outside computer science (e.g., information systems and media studies) and is linked to collaborations on specific topics or within funded projects.

Further, our work identified and categorized values into value groups. In our corpus, we identified five different value groups: editorial values, responsibility values, standard values, technical values, and UX values. Within these value groups, diversity (standard values) far outnumbers other values in terms of occurrences. Most of the publications on diversity tend to deal with topic diversity, followed by viewpoint diversity. Second, in terms of occurrences in the value groups, is recency (editorial values). Both diversity and recency are very relevant to the news domain. The
former concerns the relationship between news and democracy, and the latter concerns the connection between news and the capture of unfolding developments. However, it does leave room and invite research on other values that are less investigated so far, but may improve recommender systems in this domain.

Furthermore, we observe that values and value group aspects are system-centered, whereas others are more user-centered. For instance, coverage, popularity, and explainability affect the system as a whole. In contrast, privacy and trust are values expressed by users. Future research could investigate whether certain values and value groups might be optimized or designed (or not), depending on whether these values (or value groups) concern the system or user.

Finally, we found that recommendations are often evaluated by accuracy-based metrics. Thus, although many principles for news recommenders have been developed, there is still much work to be done in aligning these principles with relevant metrics that evaluate success and exploring the matter of potential tradeoffs. As indicated, this task is not easy and requires more inter- and transdisciplinary exchanges that help translate abstract values into design principles. It also necessitates ongoing collaboration between industry and academia. This also relates to the observation that offline methods are currently the by far the most used evaluation methods. The benefit of an academia-industry collaboration would be that these algorithms are made and tested in practice, providing the context and constraints under which these systems and their makers need to function.

This survey is subject to several limitations. First, the corpus of publications was determined by a specific query and was limited to publications in English, introducing potential biases in the selection of literature. Research conducted in other languages or with different terminology may have been excluded. Second, the field of NRS has been rapidly evolving, and the survey’s cutoff date may have led to the omission of recent developments and papers. The dynamic nature of the field requires continuous updates to capture the latest research. Third, we systematically approached the selection and coding of papers by employing several coders, randomizing the assignment of papers to coders, and including reflexivity and dialogue within the research team. While this approach helps reduce systematic biases, a qualitative approach like ours may still retain some subjectivity. A similar limitation concerns categorizing values into value groups, where other research teams may result in different groupings and labels of value groups. Last, this survey represents a distant reading of NRS publications, where publications are treated as comparable units. However, the significance and aspirations of these publications may vary significantly. Individual papers may have different goals, methodologies, and impacts that are not captured by this bird’s-eye view.

These caveats notwithstanding, our systematic literature review contributes to the body of knowledge in several ways. We have identified a comprehensive set of values (total of 40 values) in our analyzed corpus, and we have developed a categorization scheme to group these values into five value groups. This provides a solid basis for future research to build upon. Future research may expand and refine the set of values and the categorization scheme. Moreover, our review synthesizes the body of research on value-aware NRS across disciplines and communities, tracing back to 1995. Among others, this synthesis clearly indicates the under-researched values that could be relevant to explore further. Furthermore, the analysis gives direction where values need to be targeted on a more fine-grained and specific level. Finally, our analysis suggests that the driving force bringing attention to value-aware NRS seemingly comes from fields outside computer science (e.g., information systems and media studies). To move forward, interdisciplinary and transdisciplinary research collaborations are strongly encouraged. While this is often advocated, the fundamental challenge is to put these collaborations into practice.
### Table 3. Overview of the Sample

<table>
<thead>
<tr>
<th>Reference</th>
<th>Journal/Conference Name</th>
<th>Journal or Conference</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdollahpouri et al. [2]</td>
<td>WWW 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Ashraf et al. [10]</td>
<td>ICSCE 2018</td>
<td>Conference</td>
<td>2018</td>
</tr>
<tr>
<td>Babanejad et al. [12]</td>
<td>INRA 2019</td>
<td>Conference</td>
<td>2020</td>
</tr>
<tr>
<td>Blanco et al. [23]</td>
<td>CIKM 2012</td>
<td>Conference</td>
<td>2012</td>
</tr>
<tr>
<td>Boutet et al. [27]</td>
<td>IPDPS 2013</td>
<td>Conference</td>
<td>2013</td>
</tr>
<tr>
<td>Bozdag and van de Poel [28]</td>
<td>PICMET 2013</td>
<td>Conference</td>
<td>2013</td>
</tr>
<tr>
<td>Carbone and Vlassov [37]</td>
<td>ICCAC 2015</td>
<td>Conference</td>
<td>2015</td>
</tr>
<tr>
<td>Chakraborty and Ganguly [38]</td>
<td>ASONAM 2018</td>
<td>Conference</td>
<td>2018</td>
</tr>
<tr>
<td>Chen et al. [42]</td>
<td>WI-IAT 2008</td>
<td>Conference</td>
<td>2008</td>
</tr>
<tr>
<td>Chesnais et al. [43]</td>
<td>International Workshop on Community Networking</td>
<td>Conference</td>
<td>1995</td>
</tr>
<tr>
<td>Cotter et al. [47]</td>
<td>CHI EA 2017</td>
<td>Conference</td>
<td>2017</td>
</tr>
<tr>
<td>Cui et al. [48]</td>
<td>SPML 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Dacon and Liu [49]</td>
<td>WWW 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Daneshi et al. [51]</td>
<td>ICMEW 2013</td>
<td>Conference</td>
<td>2013</td>
</tr>
<tr>
<td>Epure et al. [62]</td>
<td>RecSys 2017</td>
<td>Conference</td>
<td>2017</td>
</tr>
<tr>
<td>Gabrilovich et al. [69]</td>
<td>WWW 2004</td>
<td>Conference</td>
<td>2004</td>
</tr>
<tr>
<td>Gao et al. [70]</td>
<td>WI-IAT 2011</td>
<td>Conference</td>
<td>2011</td>
</tr>
<tr>
<td>Gao et al. [71]</td>
<td>SmartBlock 2020</td>
<td>Conference</td>
<td>2020</td>
</tr>
<tr>
<td>Garcin et al. [72]</td>
<td>RecSys 2013</td>
<td>Conference</td>
<td>2013</td>
</tr>
<tr>
<td>Garcin et al. [73]</td>
<td>RecSys 2014</td>
<td>Conference</td>
<td>2014</td>
</tr>
<tr>
<td>Garrido et al. [74]</td>
<td>SISY 2015</td>
<td>Conference</td>
<td>2015</td>
</tr>
<tr>
<td>Gebremeskel and de Vries [76]</td>
<td>CLEF 2015</td>
<td>Conference</td>
<td>2015</td>
</tr>
<tr>
<td>Gharaighhehi and Vens [77]</td>
<td>OHARS 2020</td>
<td>Conference</td>
<td>2020</td>
</tr>
<tr>
<td>Harambam et al. [85]</td>
<td>RecSys 2019</td>
<td>Conference</td>
<td>2019</td>
</tr>
<tr>
<td>Hassan and McCrickard [87]</td>
<td>WWW 2019</td>
<td>Conference</td>
<td>2019</td>
</tr>
<tr>
<td>Hu et al. [93]</td>
<td>HICSS 2012</td>
<td>Conference</td>
<td>2012</td>
</tr>
<tr>
<td>Ingvaldsen et al. [94]</td>
<td>IntRS@RecSys 2015</td>
<td>Conference</td>
<td>2015</td>
</tr>
<tr>
<td>Isilamouli et al. [95]</td>
<td>HUMAN 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Jain et al. [96]</td>
<td>HotMobile 2017</td>
<td>Conference</td>
<td>2017</td>
</tr>
<tr>
<td>Kang et al. [104]</td>
<td>ICACT 2014</td>
<td>Conference</td>
<td>2014</td>
</tr>
<tr>
<td>Karimi et al. [105]</td>
<td>INRA 2019</td>
<td>Conference</td>
<td>2020</td>
</tr>
<tr>
<td>Kazai et al. [110]</td>
<td>SIGIR 2016</td>
<td>Conference</td>
<td>2016</td>
</tr>
<tr>
<td>Khattar et al. [112]</td>
<td>ICDMW 2017</td>
<td>Conference</td>
<td>2017</td>
</tr>
<tr>
<td>Kille and Albayrak [113]</td>
<td>RecTemp 2017</td>
<td>Conference</td>
<td>2017</td>
</tr>
</tbody>
</table>

(Continued)
Table 3. Continued

<table>
<thead>
<tr>
<th>Reference</th>
<th>Journal/Conference Name</th>
<th>Journal or Conference</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Krebs et al. [117]</td>
<td>CHI EA 2019</td>
<td>Conference</td>
<td>2019</td>
</tr>
<tr>
<td>Kulkarni et al. [118]</td>
<td>ICCUBEA 2019</td>
<td>Conference</td>
<td>2019</td>
</tr>
<tr>
<td>Kumar et al. [119]</td>
<td>ICDMW 2017</td>
<td>Conference</td>
<td>2017</td>
</tr>
<tr>
<td>Kumar et al. [120]</td>
<td>ICDMW 2017</td>
<td>Conference</td>
<td>2017</td>
</tr>
<tr>
<td>Li et al. [127]</td>
<td>SIGIR 2011</td>
<td>Conference</td>
<td>2011</td>
</tr>
<tr>
<td>Li et al. [129]</td>
<td>RecSys 2011</td>
<td>Conference</td>
<td>2011</td>
</tr>
<tr>
<td>Liu et al. [133]</td>
<td>WWW 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Loecherbach et al. [134]</td>
<td>WWW 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Lommatzsch et al. [135]</td>
<td>WI 2017</td>
<td>Conference</td>
<td>2017</td>
</tr>
<tr>
<td>Lu et al. [136]</td>
<td>UMAP 2020</td>
<td>Conference</td>
<td>2020</td>
</tr>
<tr>
<td>Lu et al. [137]</td>
<td>SIGIR 2019</td>
<td>Conference</td>
<td>2019</td>
</tr>
<tr>
<td>Lv et al. [141]</td>
<td>WWW 2011</td>
<td>Conference</td>
<td>2011</td>
</tr>
<tr>
<td>Ma et al. [142]</td>
<td>WWW 2016</td>
<td>Conference</td>
<td>2016</td>
</tr>
<tr>
<td>Maksai et al. [145]</td>
<td>RecSys 2015</td>
<td>Conference</td>
<td>2015</td>
</tr>
<tr>
<td>Meguebli et al. [147]</td>
<td>KDIR 2014</td>
<td>Conference</td>
<td>2014</td>
</tr>
<tr>
<td>Mohallick and Özgöbek [153]</td>
<td>WI 2017</td>
<td>Conference</td>
<td>2017</td>
</tr>
<tr>
<td>Mulder et al. [158]</td>
<td>FAccT 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Nagaki et al. [161]</td>
<td>MOBIQUITOUS 2016</td>
<td>Conference</td>
<td>2016</td>
</tr>
<tr>
<td>Niu et al. [165]</td>
<td>CHI 2018</td>
<td>Conference</td>
<td>2018</td>
</tr>
<tr>
<td>Niu and Al-Doulat [166]</td>
<td>CHIIR 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Noh et al. [167]</td>
<td>BIGCOMP 2014</td>
<td>Conference</td>
<td>2014</td>
</tr>
<tr>
<td>O’Banion et al. [168]</td>
<td>RSWeb 2012</td>
<td>Conference</td>
<td>2012</td>
</tr>
<tr>
<td>Oh et al. [170]</td>
<td>ICACT 2014</td>
<td>Conference</td>
<td>2014</td>
</tr>
<tr>
<td>Özgöbek et al. [171]</td>
<td>WEBIST 2014</td>
<td>Conference</td>
<td>2014</td>
</tr>
<tr>
<td>Özgöbek et al. [172]</td>
<td>WEBIST 2015</td>
<td>Conference</td>
<td>2015</td>
</tr>
<tr>
<td>Panteli et al. [173]</td>
<td>INRA 2019</td>
<td>Conference</td>
<td>2020</td>
</tr>
<tr>
<td>Patankar et al. [176]</td>
<td>ICSC 2019</td>
<td>Conference</td>
<td>2019</td>
</tr>
<tr>
<td>Pfahler and Morik [177]</td>
<td>FATE/MM 2020</td>
<td>Conference</td>
<td>2020</td>
</tr>
<tr>
<td>Phelan et al. [178]</td>
<td>WWW 2011</td>
<td>Conference</td>
<td>2011</td>
</tr>
<tr>
<td>Pon et al. [179]</td>
<td>KDD 2007</td>
<td>Conference</td>
<td>2007</td>
</tr>
<tr>
<td>Pon et al. [180]</td>
<td>WIDM 2008</td>
<td>Conference</td>
<td>2008</td>
</tr>
<tr>
<td>Qi et al. [188]</td>
<td>ACL-IJCNLP 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Qi et al. [189]</td>
<td>EMNLP 2020</td>
<td>Conference</td>
<td>2020</td>
</tr>
<tr>
<td>Qin and Zhang [191]</td>
<td>CONF-CDS 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Raza and Ding [192]</td>
<td>Big Data 2019</td>
<td>Conference</td>
<td>2019</td>
</tr>
<tr>
<td>Raza and Ding [193]</td>
<td>Big Data 2020</td>
<td>Conference</td>
<td>2020</td>
</tr>
<tr>
<td>Raza and Ding [194]</td>
<td>Big Data 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Reuver and Mattis [197]</td>
<td>EACL 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
</tbody>
</table>

(Continued)
Table 3. Continued

<table>
<thead>
<tr>
<th>Reference</th>
<th>Journal/Conference Name</th>
<th>Journal or Conference</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robindro et al. [199]</td>
<td>ICCCA 2017</td>
<td>Conference</td>
<td>2017</td>
</tr>
<tr>
<td>Bathla et al. [17]</td>
<td>ICRITO 2015</td>
<td>Conference</td>
<td>2015</td>
</tr>
<tr>
<td>Sadhasivam et al. [201]</td>
<td>ICECCS 2014</td>
<td>Conference</td>
<td>2015</td>
</tr>
<tr>
<td>Saravia et al. [205]</td>
<td>TAAI 2017</td>
<td>Conference</td>
<td>2018</td>
</tr>
<tr>
<td>Sertkan et al. [208]</td>
<td>CBI 2019</td>
<td>Conference</td>
<td>2019</td>
</tr>
<tr>
<td>Shan et al. [209]</td>
<td>ICCC 2016</td>
<td>Conference</td>
<td>2017</td>
</tr>
<tr>
<td>Su et al. [220]</td>
<td>SMAP 2016</td>
<td>Conference</td>
<td>2016</td>
</tr>
<tr>
<td>Sullivan et al. [221]</td>
<td>ACM UMAP 2019 Adjunct</td>
<td>Conference</td>
<td>2019</td>
</tr>
<tr>
<td>Sun et al. [222]</td>
<td>SCC 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Suppasert et al. [224]</td>
<td>ICT-ISPC 2017</td>
<td>Conference</td>
<td>2017</td>
</tr>
<tr>
<td>Tavakolifard et al. [227]</td>
<td>WWW 2013 Companion</td>
<td>Conference</td>
<td>2013</td>
</tr>
<tr>
<td>Tintarev et al. [228]</td>
<td>UMAP 2018</td>
<td>Conference</td>
<td>2018</td>
</tr>
<tr>
<td>Verheij et al. [235]</td>
<td>WI 2012</td>
<td>Conference</td>
<td>2012</td>
</tr>
<tr>
<td>Vrijenhoeck et al. [237]</td>
<td>CHIIR 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Wang et al. [239]</td>
<td>CIKM 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Wang et al. [240]</td>
<td>SIGIR 2010</td>
<td>Conference</td>
<td>2010</td>
</tr>
<tr>
<td>Wang et al. [241]</td>
<td>ICDMW 2021</td>
<td>Conference</td>
<td>2021</td>
</tr>
<tr>
<td>Wang et al. [242]</td>
<td>ICDE 2015</td>
<td>Conference</td>
<td>2015</td>
</tr>
<tr>
<td>Chen et al. [41]</td>
<td>CMC 2009</td>
<td>Conference</td>
<td>2009</td>
</tr>
<tr>
<td>Werner and Lommatzsch [246]</td>
<td>CLEF 2014</td>
<td>Conference</td>
<td>2014</td>
</tr>
<tr>
<td>Wu et al. [251]</td>
<td>IJCAI 2020</td>
<td>Conference</td>
<td>2020</td>
</tr>
<tr>
<td>Wu et al. [253]</td>
<td>BigComp 2016</td>
<td>Conference</td>
<td>2016</td>
</tr>
<tr>
<td>Xie et al. [255]</td>
<td>CIKM 2013</td>
<td>Conference</td>
<td>2013</td>
</tr>
<tr>
<td>Xue et al. [256]</td>
<td>ETT and GRS 2008</td>
<td>Conference</td>
<td>2008</td>
</tr>
<tr>
<td>Yeung and Yang [258]</td>
<td>DeSE 2010</td>
<td>Conference</td>
<td>2010</td>
</tr>
<tr>
<td>Yeung et al. [259]</td>
<td>CICSyN 2010</td>
<td>Conference</td>
<td>2010</td>
</tr>
<tr>
<td>Zeleník and Bieliková [261]</td>
<td>WEBIST 2011</td>
<td>Conference</td>
<td>2011</td>
</tr>
<tr>
<td>Zhao et al. [264]</td>
<td>WI-IAT 2020</td>
<td>Conference</td>
<td>2020</td>
</tr>
<tr>
<td>Zhu et al. [265]</td>
<td>ICDM 2014</td>
<td>Conference</td>
<td>2014</td>
</tr>
<tr>
<td>Bastian et al. [14]</td>
<td>Digital Journalism</td>
<td>Journal</td>
<td>2021</td>
</tr>
<tr>
<td></td>
<td>Conflict Management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bastian et al. [16]</td>
<td>Internet Policy Review</td>
<td>Journal</td>
<td>2020</td>
</tr>
<tr>
<td>Bodó et al. [25]</td>
<td>Digital Journalism</td>
<td>Journal</td>
<td>2019</td>
</tr>
<tr>
<td>Briguez et al. [29]</td>
<td>International Journal on Artificial Intelligence Tools</td>
<td>Journal</td>
<td>2013</td>
</tr>
<tr>
<td>Burr et al. [33]</td>
<td>Minds and Machines</td>
<td>Journal</td>
<td>2018</td>
</tr>
<tr>
<td>Chakraborty et al. [39]</td>
<td>Information Retrieval Journal</td>
<td>Journal</td>
<td>2019</td>
</tr>
<tr>
<td>Chen et al. [40]</td>
<td>IEEE Access</td>
<td>Journal</td>
<td>2017</td>
</tr>
<tr>
<td>Descampe et al. [56]</td>
<td>AI &amp; Society</td>
<td>Journal</td>
<td>2022</td>
</tr>
<tr>
<td>Diaz et al. [57]</td>
<td>Online Information Review</td>
<td>Journal</td>
<td>2001</td>
</tr>
<tr>
<td>Dovbyshe et al. [59]</td>
<td>Digital Journalism</td>
<td>Journal</td>
<td>2022</td>
</tr>
<tr>
<td>Eskens [63]</td>
<td>International Data Privacy Law</td>
<td>Journal</td>
<td>2019</td>
</tr>
<tr>
<td>Feng et al. [64]</td>
<td>IEEE Access</td>
<td>Journal</td>
<td>2020</td>
</tr>
<tr>
<td>Feng et al. [65]</td>
<td>Journal of Web Engineering</td>
<td>Journal</td>
<td>2021</td>
</tr>
<tr>
<td>Gharahighechi and Vens [78]</td>
<td>Personal and Ubiquitous Computing</td>
<td>Journal</td>
<td>2021</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Reference</th>
<th>Journal/Conference Name</th>
<th>Journal or Conference Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gharahighehi et al. [79]</td>
<td>Information Processing &amp; Management</td>
<td>Journal 2021</td>
</tr>
<tr>
<td>Gu et al. [81]</td>
<td>Neural Computing and Applications</td>
<td>Journal 2016</td>
</tr>
<tr>
<td>Harambam et al. [86]</td>
<td>Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences</td>
<td>Journal 2018</td>
</tr>
<tr>
<td>Heitz et al. [88]</td>
<td>Digital Journalism</td>
<td>Journal 2022</td>
</tr>
<tr>
<td>Helberger [90]</td>
<td>Digital Journalism</td>
<td>Journal 2019</td>
</tr>
<tr>
<td>Joris et al. [101]</td>
<td>Digital Journalism</td>
<td>Journal 2021</td>
</tr>
<tr>
<td>Karimi et al. [106]</td>
<td>Information Processing &amp; Management</td>
<td>Journal 2018</td>
</tr>
<tr>
<td>Karimi et al. [107]</td>
<td>Journal of Information Science</td>
<td>Journal 2021</td>
</tr>
<tr>
<td>Koo et al. [115]</td>
<td>Knowledge and Information Systems</td>
<td>Journal 2021</td>
</tr>
<tr>
<td>Li et al. [126]</td>
<td>Journalism &amp; Mass Communication</td>
<td>Journal 2020 Quarterly</td>
</tr>
<tr>
<td>Li et al. [130]</td>
<td>Expert Systems with Applications</td>
<td>Journal 2014</td>
</tr>
<tr>
<td>Li and Wang [131]</td>
<td>IEEE Access</td>
<td>Journal 2019</td>
</tr>
<tr>
<td>Li et al. [132]</td>
<td>Information Sciences</td>
<td>Journal 2010</td>
</tr>
<tr>
<td>Lu et al. [139]</td>
<td>Journal of Systems and Software</td>
<td>Journal 2014</td>
</tr>
<tr>
<td>Makhortykh and Bastian [144]</td>
<td>Media, War &amp; Conflict</td>
<td>Journal 2022</td>
</tr>
<tr>
<td>Mizgajski and Morzy [152]</td>
<td>User Modeling and User-Adapted Interaction</td>
<td>Journal 2019</td>
</tr>
<tr>
<td>Møller [155]</td>
<td>Digital Journalism</td>
<td>Journal 2022</td>
</tr>
<tr>
<td>Montes-García et al. [156]</td>
<td>Expert Systems with Applications</td>
<td>Journal 2013</td>
</tr>
<tr>
<td>Monzer et al. [157]</td>
<td>Digital Journalism</td>
<td>Journal 2020</td>
</tr>
<tr>
<td>Nanas et al. [162]</td>
<td>Information Processing &amp; Management</td>
<td>Journal 2010</td>
</tr>
<tr>
<td>Portilla [181]</td>
<td>El Profesional de la Información</td>
<td>Journal 2018</td>
</tr>
<tr>
<td>Prawesh and Padmanabhan [187]</td>
<td>PLOS ONE</td>
<td>Journal 2021</td>
</tr>
<tr>
<td>Raza and Ding [195]</td>
<td>Artificial Intelligence Review</td>
<td>Journal 2022</td>
</tr>
<tr>
<td>Sagui et al. [202]</td>
<td>Inteligencia Artificial</td>
<td>Journal 2008</td>
</tr>
<tr>
<td>Saranya and Sudha Sadasivam [204]</td>
<td>Mobile Networks and Applications</td>
<td>Journal 2017</td>
</tr>
<tr>
<td>Semenov et al. [207]</td>
<td>Expert Systems with Applications</td>
<td>Journal 2022</td>
</tr>
<tr>
<td>Shin [211]</td>
<td>Journalism Studies</td>
<td>Journal 2021</td>
</tr>
<tr>
<td>Sivetc and Wijermars [213]</td>
<td>Media and Communication</td>
<td>Journal 2021</td>
</tr>
<tr>
<td>Smets et al. [214]</td>
<td>Digital Journalism</td>
<td>Journal 2022</td>
</tr>
<tr>
<td>van Drunen et al. [233]</td>
<td>International Data Privacy Law</td>
<td>Journal 2019</td>
</tr>
<tr>
<td>Viana and Soares [236]</td>
<td>International Journal on Artificial Intelligence Tools</td>
<td>Journal 2017</td>
</tr>
<tr>
<td>Wieland et al. [247]</td>
<td>Media and Communication</td>
<td>Journal 2021</td>
</tr>
<tr>
<td>Xiao et al. [254]</td>
<td>China Communications</td>
<td>Journal 2015</td>
</tr>
<tr>
<td>Yoon et al. [260]</td>
<td>Applied Mathematics and Information Sciences</td>
<td>Journal 2015</td>
</tr>
<tr>
<td>Zhu et al. [266]</td>
<td>IEEE Access</td>
<td>Journal 2018</td>
</tr>
</tbody>
</table>
REFERENCES


Where are the Values? A Systematic Literature Review on News Recommender Systems


Where are the Values? A Systematic Literature Review on News Recommender Systems


Received 22 August 2023; revised 1 February 2024; accepted 6 March 2024