Diversity of What? On the Different Conceptualizations of Diversity in Recommender Systems

SANNE VRIJENHOEK, AI, Media & Democracy Lab and Institute for Information Law, University of Amsterdam, The Netherlands

SAVVINA DANIIL, Centrum Wiskunde & Informatica, The Netherlands

JORDEN SANDEL, University of Amsterdam, The Netherlands

LAURA HOLLINK, Centrum Wiskunde & Informatica, The Netherlands

Diversity is a commonly known principle in the design of recommender systems, but also ambiguous in its conceptualization. Through semi-structured interviews we explore how practitioners at three different public service media organizations in the Netherlands conceptualize diversity within the scope of their recommender systems. We provide an overview of the goals that they have with diversity in their systems, which aspects are relevant, and how recommendations should be diversified. We show that even within this limited domain, conceptualization of diversity greatly varies, and argue that it is unlikely that a standardized conceptualization will be achieved. Instead, we should focus on effective communication of what diversity in this particular system means, thus allowing for operationalizations of diversity that are capable of expressing the nuances and requirements of that particular domain.

1 INTRODUCTION

The concept of diversity is at the same time omnipresent and very ambiguous [13]. In popular discourse, diversity usually refers to the variation of human characteristics, often related to a notion of identity politics [3]; in biological research, as a qualifier to the health of an ecosystem [47]; and in media studies, as a concept adjacent to pluralism, expressing whether a news selection contains a plurality of sources, voices and opinions [25]. While all interpretations are valid and intuitively seem to reflect similar concepts, they differ in their operationalization in a way that is unique to their domain. At the same time, diversity is consistently noted as a social value that is beneficial to pursue, though in different capacities.

In the recommender systems domain accounting for the diversity of a recommendation can help avoid monotony [6, 51]. This follows from the assumption that a recommender system that is purely optimized on predicted relevance will result in a feedback loop and thus prioritize similar content, leading to 'more of the same'. There is as such also a business case to be made for diversity. While it has been challenging to show empirically, diversity may lead to higher user satisfaction and retention, increasing revenue in the process [12, 23]. This argument can be extended for the news domain, where worries over filter bubbles that reinforce existing beliefs by only showing content that align with a user's preferences are especially prevalent [33, 54]. Having a news recommender system that pursues diversity could help expose users to things different than they are used to or expect seeing, and tailor to their specific information needs [20, 21]. For machine learning in general, it is also important to account for diversity in a social context. Since many algorithms are trained on datasets that are not representative of all groups in society, not accounting for diversity may further "amplify stereotypes, alienate users, and further entrench rigid social expectations" [34].

The many different interpretations of diversity are a fundamental challenge to the practical development of recommender systems. Evaluating the performance of a recommender system requires objectively measurable properties, and the field strives to do so in a standardized manner [52]. Standardization allows for comparability and reproducibility of results and algorithms[1], and is achieved through, among others, the use of benchmark datasets, sampling techniques and evaluation metrics such as MRR or NDCG [50], but has not yet been achieved for diversity. Loecherbach et al. [31], who did a literature review of existing measures of diversity in media studies, note that "there is little to no overlap between concepts and operationalizations used in the different fields interested in media diversity". Lawrence [29], executing a conceptual analysis of the term 'diverse books', notes that the topic is inherently political, and that "we have yet to arrive at a clear consensus on what the modifier 'diverse' means in this and other instances".

The conceptual unclarity around diversity makes that not only may we implement it differently, we may also *mean* something completely different when we use the term. One could argue that diversity is, in fact, an essentially contested concept [8, 16], meaning that it is open for discussion and debate and that we are unlikely to reach consensus on its meaning. As such, striving for agreement or a clear definition may lead to a standstill and hinder progress, as it is unclear what a good operationalization or implementation would look like. It may cause the conceptualization of diversity to be driven by the information that is available or more easily suitable for quantification, rather than what is desirable; furthermore, blindly trusting 'objective' measures "may obscure the fact that the conceptualization of diversity is, eventually, a normative choice" [26]. Instead, what we need may be a more flexible operationalization that is capable of reflecting the nuances and requirements of the domain it is deployed in.

The aim of this paper is to explore the dimensions in which industry practitioners within a limited domain conceptualize diversity. To this end we conduct a series of semi-structured interviews with practitioners from three different public service media organizations in the Netherlands: a broadcaster, a news organization, and a library. The choice of these organizations is deliberate; though the medium through which they do it differs, they all play an active and important role in the dissemination and curation of information and ideas. As such, we expect there to be a fairly established conceptualization of diversity within the organization, and comparatively more overlap between them than with, for example, a commercial music recommendation service.

Through analysis of the interviews, and guided by the 'components' of diversity formalization defined in Vrijenhoek et al. [48], we aim to answer the following questions: what *goals* do the organizations aim to achieve with the recommendations, which *aspects* do they consider relevant for diversity, and through which *tactics* are the recommendations diversified. We find that even within the limited domain of public service media recommendation, there is a wide variety of conceptualizations of diversity. With this, we underline that a standardized definition of diversity in recommender systems is likely not achievable. However, this empirical categorisation, albeit non-exhaustive, can assist in the process of conceptualizing and implementing diversity in a particular concrete context.

2 RELATED WORK

Recommender systems have long moved from evaluating recommendations solely based on accuracy-related metrics. Other metrics such as novelty, serendipity and diversity have become a common part of evaluation practices [6]. Diversity in recommendation is a current topic, and multiple diversification methods have been proposed in recent years [15, 28]. In recommender systems research, diversity is often viewed as the opposite of similarity; a list of recommended

---

[1]Recent strands of research have noted that this perceived notion of standardization is often false. Even with commonly accepted principles there are often significant differences in implementation, and thus output [42, 45]

items is diverse if the items are sufficiently different between them along a set of axes [53]. However, Jesse et al. [24] point to a gap between human perceptions on diversity and intra-list similarity (ILS) commonly used to assess diversity in offline recommender systems experiments. Through a user study, they find that while ILS can be a good proxy, the details of the implementation matter and require validation in a given domain and application.

When diversity is seen as a social value that needs to be incorporated in a system, standardizing its definition and operationalization becomes even more challenging. Mitchell et al. [34] define diversity in a subset selection task (e.g., the construction of a list of items to recommend) as "variety in the representation of individuals in an instance or set of instances, with respect to sociopolitical power differentials (gender, race, etc.)". As such, they view diversity as a concept with inherently sociopolitical connotations in contrast to the more general term 'heterogeneity'.

In the context of media recommendations, diversity is often closely associated with pluralism. According to Karppinen [26], the interpretation of pluralism and diversity in media is dependent on the political and normative understanding of the role of media in society. Additionally, defining media diversity is further complicated by its often contradictory aspects and interpretations. Van Cuilenburg [46] point out an antithesis between the normative media diversity frameworks of *reflection* and *openness*; one requires content distribution to proportionally reflect societal distributions of relevance, while the other corresponds to perfectly equal representation and attention to all people and ideas in society. Similar contradictions have been laid out in the domain of library and information studies [29]. Based on a systematic literature review on media diversity, Loecherbach et al. [31], also referenced in the introduction on the little overlap between different fields' conceptualization of diversity, conclude that relevant research should be guided by an interdisciplinary effort to define and benchmark the concept of media diversity, along with its different sub-dimensions. Other work studying diversity in recommender systems would typically acknowledge the complexity of the concept, yet model it following existing technical standards; either as a distance to a user's reading history [18], political leaning [19] or as pair-wise distance between the items in a recommendation, for example in topics, categories and/or tone [35]. In this work, we aim to contribute to this effort by attempting to map the dimensions of diversity in media recommender systems.

Diversity in Public Service Media

Diversity in recommendation is especially relevant for public service media (PSM), whose role and societal responsibilities call for a careful consideration of the content they produce and give exposure to. Many media organizations underline the potential of recommender systems and the importance of reflecting editorial values such as diversity therein [5, 17, 27, 36]. PSM are often required to offer diverse content, which might be at odds with the primarily commercial use of recommender systems that mainly intend to increase media consumption, and therefore, profit [22]. Even if we assume societal agreement on the importance of providing diverse content for PSM, the practical implications are harder to lay out. Helberger [20] outlines four models for the normative conceptualization of diversity in recommender systems that serve different purposes: the liberal, the deliberative, the participatory, and the critical. While each of the models is in its own way relevant for PSM, the work suggests that the critical perspective, which focuses on the visibility of minority voices that are often disadvantaged in public platforms, is less likely to be encountered in commercial applications, and therefore could be partly served by PSM. Regardless, few have succeeded in concrete implementations that reflect diversity as a normative value, in part due to the gap between journalistic values and recommender system evaluation metrics [49]. Møller [37] suggests that "the abstract nature of journalistic values make them hard to account for computationally", and that "[h]uman journalists have an important role to play in these processes not only to help

conceptualise the values themselves but also as part of new algorithmic news practices." Translating values into concrete algorithmic practices is thus as challenging as it is necessary.

To bridge the gap between theory and practice, the perspectives of practitioners in a given domain can assist, orient, and ground research. On that note, researchers have conducted interviews with practitioners on the interaction between emerging algorithmic systems and norms and values. Sørensen [44] interviewed developers, data scientists, and project leaders from nine European PSM organizations on the topic of implementing recommender systems. They report that, while interviewees believe diversity to be an essential aspect of their catalogue, they are reluctant to depend on an algorithmic implementation instead of the traditional editorial control. Additionally, they attribute the general lack of formal definition of diversity to the different understanding between politicians, practitioners, and users. Bastian et al. [2] interviewed practitioners from different departments of two newspapers regarding the impact of algorithmic news recommenders on their organization's norms and mission, as well as how to integrate them in the design of news recommenders. They found that, while the interviewees attach varying degrees of importance to different values, diversity is perceived as one of the core values for news recommender design and implementation. During the interviews we conducted, we specifically focused on the conception and implementation of diversity, which allowed for an in-depth outline of the perspectives and practices of the interviewed professionals.

3 METHOD

For this study we conducted a series of semi-structured interviews with three public service media organizations in the Netherlands: a broadcaster, a news organization, and a library. Interviews were carried out in-person and on site in the offices of the interviewees, and took place over a span of four months, between December 2022 and March 2023. At this time each of these organizations were, at different stages of completion, (exploring the possibility of) developing a recommender system to effectively serve content to their users, which also means that decisions about incorporating diversity in the recommendations were actively being made.

3.1 Candidate Selection

Potential candidates were approached through a snowball sampling technique: after initial contacts with the organization were established, interviewees were asked for recommendations of colleagues to interview in the next round. We actively tried to find a set of participants that reflected the composition of the organisations and the relevant figures in the design, deployment and use of the recommender system. This process yielded twelve participants in total: six at the broadcaster, four at the library, and two at the news organization. Following Smets et al. [43], our goal was to have a good spread of different types of stakeholders, and sought participants with roles related to Business (four participants), Curation (two participants), Product owner (three participants) and Technology developer (three participants). During the snowball sampling we would explicitly ask participants whether they knew of potential interviewees with a role we had not covered yet. Finding all different roles did not always succeed, and we acknowledge this as a limitation of our work. This is also why we are adamant about not making normative claims about the definition of diversity per domain, but rather present it as an exploratory study. Participants were predominantly male (9 out of 12) and white (11 out of 12). This is reflective of the workforce but a caveat for generalization, further outlined in Section 3.4. Disclosing too many details about individual participants and their roles would undermine the promised anonymity, and thus each participant is assigned a code reflecting only their organization.

3.2 Interview structure

The interviews consisted of four parts. In the first part, interviewees introduced themselves and their role within the organization. In the second part interviewees were asked about their general conceptions of diversity, and how these conceptions related to their organization. Here, the goal was to understand the mission of the organization in question, and the role diversity plays in that mission. In the third part interviewees described their knowledge of the recommender systems that were in use by their organization, either in planning or in production. This served as a check that participants were sufficiently informed to be included in the analysis, and to prime the interviewees for the fourth part, in which they were specifically asked about the role of diversity in the recommender systems of their respective organization. The aim here was to go beyond what currently exists, and instead focus on what the recommender system should look like in an ideal situation. This part of the interview also contained a small experiment, in which the interviewees were asked to take on the role of a recommender system and rank a set of items while keeping diversity in mind. During the experiment participants were asked to voice their thought process by thinking out loud [7]. Our primary interest was not the final recommendation generated, but which aspects of the items participants considered before (not) including an item. For each organization we prepared a set of 15 candidate items based on the organization's (potential) catalogue, and included the metadata a user would see when interacting with the system. We attempted to ensure (to the best of our abilities) that there would be enough diversity in this candidate list present. To cover our own blind spots we would ask participants at the end of the experiment whether there was a type of content that they were missing. We acknowledge that our own conception here steers the type of diversity that could potentially be found (see also Section 3.4).

For each of the organizations, the metadata always included the items' title, summary and a cover photo. For the library, we also included the authors' name and a set of keywords about the book; for the broadcaster, the title and description of the series the item was part of (if applicable); and for the news organization, the time of publication and the first few paragraphs of the news item. Based on this candidate set, we asked participants to make a diverse selection of 5 items. In order to not influence people's thought processes in what could potentially be relevant aspects of diversity, we deliberately left instructions vague, and did not include a profile of a user to create a recommendation for in the instructions. Instead, we considered the participants' questions for clarification as part of what they deemed relevant for diverse recommendation.

3.3 Coding and analysis

We executed an inductive thematic analysis on each of the interviews, guided by the research questions posed in Section 1: what the organizations aimed to achieve with a (diverse) recommendation, which aspects of the items they would consider during recommendation, and which tactics they would employ to diversify the recommendation. The first two authors created a coding scheme based on half of the interviews, which after completion were discussed and merged. With the resulting coding scheme, the authors annotated the half of the interviews they had not previously seen, and extended the coding scheme when gaps were identified. Additionally, after the respective coding schemes were merged, the two authors independently coded the same interview and proceeded to compare their outputs. This step helped ensure practical consensus, as the authors were able to align on the specific nuanced interpretation of each of the codes, as well as how it applies in the context of the interviews. Based on the resulting coding scheme, we created a diversity 'map' of relevant aspects and how they relate to each other, which will be discussed in the next section. The coding

scheme is included in Appendix A. After the first draft of the paper was finished, we reached out to all participants. We shared with them which quotes had been attributed to them, and asked them to reach out in case of misunderstandings.

3.4 Limitations

There are a number of important caveats to consider in light of this method and experiment. By presenting the participants with a list of items to recommend, we inadvertently steer the type of diversity they are likely to mention. For example, if our sample did not contain any mention of politics, the participants might also be less likely to mention this as a relevant aspect. Simultaneously, participants may be influenced by what is commonly considered a relevant aspect of diversity, or what interpretations are currently feasible rather than ideal. As such, we cannot draw conclusions on the importance of a certain aspect.

Another difficulty when running this experiment was accounting for the recency of items, which is extremely relevant for the news organization and the broadcaster, where the first will never want to recommend old news, but the latter may sometimes want to include older content. This was especially an issue given that interviews took place on different days, sometimes weeks apart. For the broadcaster we mixed a stable set of older content with content that was at the day of the interview popular in the system. This has the important caveat that the results between broadcaster interviewees are not fully comparable. For the news organization, we opted for a fixed date a few months in the past. While the interviews are comparable in this case, there is a risk that important contexts are forgotten or mixed up with current events.

Lastly, while diversity in and of itself is already a complicated and multi-faceted concept, the concepts related to it are too. People may talk about 'different ethnicities', and can mean, among others, different skin colors, nationalities or cultures. By extension, our findings are largely influenced by the people that participated in the interviews. Organizations are not a monolith, and many of the interviewees were very explicit about not representing the opinion of every member of the organization they belong to. Furthermore, many of the responses are influenced by the background of the participants themselves, and the fact that this research was conducted in the Netherlands. The Netherlands has a strong public service media system with a focus on representing different groups in society [9]. That being said, participants were overwhelmingly white (11 out of 12) and male (9 out of 12). While this is representative of the general demographic working on recommender systems in the Netherlands, people are not as acutely aware of the needs of groups they are not, themselves, a part of [4, 32]. As such, this is a suitable group for a descriptive study ('what is') into the conceptualizations of diverse recommendations in public service media organisations; however, a complete normative formalization of diversity ('what should be') would require the participation of people from different backgrounds and perspectives in order to provide a complete overview. All limitations considered, the results of this study cannot be used as a definition of diversity. Rather, we aim to show that, even within this limited domain, 'diversity' indeed may refer to a wide variety of things.

4 RESULTS

We expect that how organizations operationalize diversity is dependent on what they aim to achieve with their recommendation. We first identify this goal in the following section (Section 4.1). Then, we analyze both the different aspects (Section 4.2) and tactics (Section 4.3) mentioned by the participants.

4.1 Goal of recommendation

Being a public organization means that the primary goal of the organization is to provide services that benefit society [38]. As such, each organization's goals with building a recommender system extend beyond selling ads and optimizing for clicks, as is the norm for most commercial organizations. This difference between being a public or a commercial organization is explicitly mentioned by most interviewees (L2,L3,L4,B2,B3,B4,B5,N1,N2), and awareness of this distinction can be considered central in day-to-day operations. In the absence of optimizing for monetary gains, the goal of recommendation is different for each organization, and is strongly linked to its mission.

4.1.1 News. Both the News organization and the Broadcaster highlight that they are "for and from everyone". For the News organization, this is fairly straightforward: to *"reach the widest possible audience, [and] enable people to be well-informed"* (N2). This means that the news should be accessible to anyone, regardless of skill level, and that journalistic quality takes precedence over personal interest (N1,N2). Each article has a non-personalized 'more like this' section which is automatically populated by a recommender system, but can be supplemented or turned off by the editorial team. Furthermore, there is a personalized recommender system under development that aims to connect readers to "important news they have missed" (N1). For this, the organization is experimenting with how behavioral data and editorial selections can complement each other, in order to compete against the commercial players (N1).

The News organization notes a very strong collaboration between the technical and editorial teams. When opening the app, the users will always land on the editorially curated front page listing the most important news of that moment. At any time, the editorial team has the power to turn off the recommender systems. While diversity is an important concept in the news organization, it is not something that is currently explicitly built into the recommender system. Rather, it is seen as a procedural thing, to be considered at every step of content creation. This goes from choosing which stories and events to cover, to which people are doing the reporting, to writing about events in a neutral way covering multiple perspectives. This control results in a set of items to recommend from that *"have to be told from a diverse standpoint in the first [place]"* (N1). The organization sees the UX design of the recommender system, including where it is placed within the app and accompanied by which headers and explanations as more impactful (N1). They do see potential in personalization of style, telling the news through the user's desired medium (text, video, audio) or in a language complexity level suitable to the reader (N2).

4.1.2 Broadcaster. For the Broadcaster, the mission to be 'from and for everyone' translates a little differently. The organization is effectively an umbrella for multiple smaller broadcasters, each representing a different section of Dutch society. They are the ones pitching ideas and creating the content, though the public broadcaster may request a certain type of program if they feel a particular perspective is missing (B3). The broadcaster then tries to balance on the one hand helping their users recognize themselves in the content they have on offer, while also fostering understanding and knowledge of people, ideas and groups that are 'different' (B1,B2,B4,B5). For this, they see a clear purpose for personalized and diverse recommendation: *"[I]f someone believes or thinks or feels in X, and they look at our platform, that person is also confronted with Y. And that Y is slightly different from what the person had in mind with X."* (B2). Diversity plays a large role in achieving this, but the split in goals between recognition and broadening causes a great deal of conceptual unclarity. As one of the interviewees notes, *"[i]f you are looking at personalization, then we don't want people to all get the same thing recommended. That's what I tend to see as diversity. [...] [P]luralism in recommendations would be [that] we'd like to look into [a] topic from different perspectives. And I think if you talk to different people within this organisation, those two things kind of mix."* (B6).

The current platform is largely manually curated, with separate sections (displayed relatively low on the landing page) for algorithmic recommendations. These algorithmic recommendations balance personal relevance, based on a user's past viewing behavior, with a so-called 'public value score'. This score is an aggregate, obtained through a daily survey sent to users, in which they are asked to score the programs they watched recently on things such as the presence of certain population groups or multiple perspectives. The broadcaster is working on the development of a new platform, which should have a much stronger algorithmic focus. They hold a uniformly strong position on user control: while they, as the broadcaster, should provide a diverse offering, the user is eventually always in charge of what they do and do not watch.

4.1.3 Library. The Library hosts a vast collection containing every book published in or about the Netherlands (L2). They hold a unique position among the cultural institutes, as they have the added role of coordinating 138 other public libraries. There is high interest in digitization and innovation in general, which manifests in creation of proof-of-concept applications that the other libraries can choose to adopt or not (L3). Among other services and initiatives, the Library has the dual goal of having *more people read* and *people read more* (L4). The Library suspects that the decline of reading among its users is in part caused by a general lack of available time. Currently, there is a recommender system feature in the mobile application hosted by the Library, but its functionality is not satisfactory as *"people cannot find something they might be interested in"* (L4), an issue partly caused by the organization of metadata. Improved personalization can assist with the goal of attracting more patrons, as it increases accessibility to the collection for different types of people. At the same time, the Library is conducting research on the ethics around recommender systems, with topics like bias and privacy being central (L3). In this sense, personalization is not sufficient. There are active efforts to reduce bias that historically existed in the collections, in order to facilitate *"every citizen to be able to participate in society" and "make society [as a whole] better, smarter and more creative"* (L2).

Overall, the Library aims to deploy a well-functioning recommender system to satisfy their readers while remaining inclusive and enacting bias correction. However, building such a system is not trivial, and there are many questions about how development should be approached.

4.2 Aspects of diversity

During the recommendation exercise, and the questions leading up to it, participants would mention aspects of the content or the user beyond personal relevance that would lead them to include that item in the diverse recommendation or not. Figure 1 provides a schematic overview of these aspects, which can be divided into Item-, Human- and World aspects.

4.2.1 Item aspects. The first dimension considers aspects of the items that are to be recommended. For the Library these would be books, news articles for the news agency, and videos for the Broadcaster. The first group of aspects revolves around *topicality*, or what is actually discussed in the items. The dimensions mentioned by most interviewees are **category**, or sometimes genre, and **topic**. All interviewees but one mentioned balancing different categories and topics in the recommendation, to avoid saturation and to keep a user engaged. This is reflective of how diversity is currently most often conceptualized in technical recommender system literature; see [6].

Related but still separate from the topic are the **geographic location** the item centers on, and the **time period** it discusses. The Library may want to recommend books that discuss a topic through the lens of different time periods (L1), whereas the News organization aims to ensure that they not only cover news from densely populated urban areas, but also from rural areas (N2). Secondly, recommendations may vary on accounts of *stylistic* properties. This relates
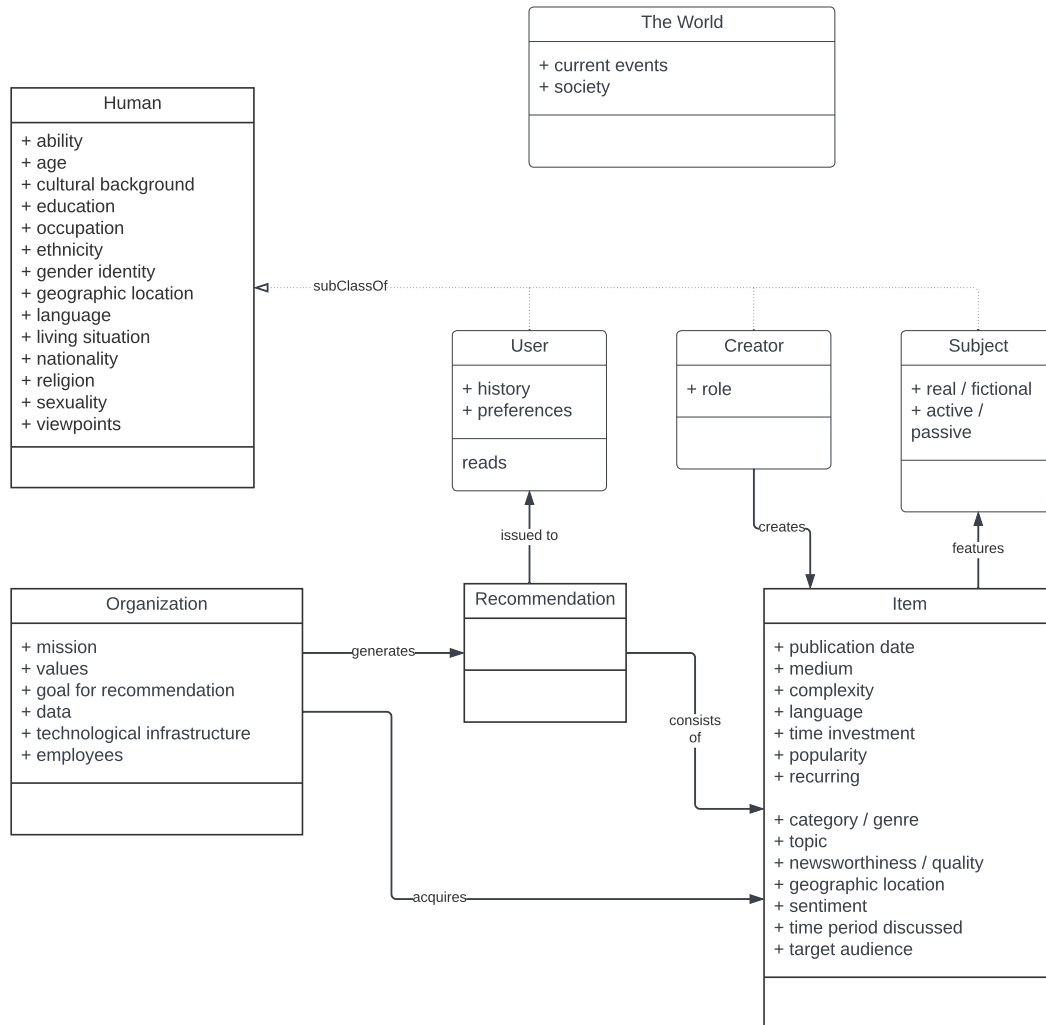
Fig. 1. Schematic overview of the identified aspects of diversity and how they interact with each other. The 'World' class is unconnected in the graph, but in truth encapsulates and informs everything: from the content that is being produced, to what user wants to read, to what constitutes a 'viewpoint' or a 'minority'.

to the way in which a message is communicated, and whether it is easily accessible for the user. Examples are the **complexity** of the content, the amount of **time investment** a user is required (and willing) to make, the **language** it is written in, the **target audience** of the item, and its **medium** (text, video, audio). Many of these properties are symmetric with characteristics of the user: a certain item complexity level or language requires a certain amount of skill from the user.

9

Thirdly, there are a range of other item properties which may be prioritized. An important one is an item's **publication date**. While this is of primary importance to the news agency (one would not want to recommend a news item from two months ago), it is much less so for the broadcaster and the library: while there may be some preference for new publications, they also want to help their users discover less popular content. Another important dimension is the perceived **quality** of the item. While the news organization would explicitly optimize for (editorial) quality, opinions differ between participants from the library: while some would find it acceptable if readers only read comic books and manga, others wanted them to be pushed towards more high-brow literature (L4). More differences could be observed in terms of an item's **popularity**: some said that popular items are likely to be items that readers are looking for and should therefore be recommended (L4,B1), while others explicitly distanced themselves from it: *"I do know that out of this selection, [sensationalist article] would have been clicked on more than [serious article]. [...] That's why what we're doing with those recommendations is designed the way we designed it, which is that the journalistic selection we make takes precedence."* (N2). Additionally, items that had a high production **cost** may be prioritized (B3,B4).

Last, but definitely not least, interviewees referred to the people involved in the content. These can be split into the **Creators**, such as producers or authors, and the people that were **Subject** in the items, either through active participation or being discussed by others. Think of guests at a talkshow, politicians discussed in the news or a novel's protagonist. These are further discussed in Section 4.2.2.

4.2.2 Human aspects. While item category is the aspect of diversity mentioned by most interviewees, it is closely followed by notions of the diversity in the context of *people*. While there are a number of attributes that all humans share, there are three 'subclasses' that can be derived with specific roles within the system. These are Users, Creators and Subjects. Users are the people that receive the recommendations, and have a set of unique properties that the other types do not have: their **history** and **preferences**. Creators are, as the name implies, in some way or another involved in the creation of the items; the authors of books, journalists or producers of shows. Lastly, there are the people that are the Subject of the item's content, such as protagonists of books or people appearing on talk shows. These subjects can be fictional or real, and be active or passive agents within the Item by either speaking on their own behalf or by being described or mentioned by others.

The general Human aspects, such as **age**, **gender identity**, etc, are applicable to each type. A frequently mentioned aspect of diversity is relating to a person's background. **Culture**, **ethnicity**, **nationality** and sometimes **religion** (for example 'Muslims') are often used interchangeably, and the distinction or relation between them is not always clear. For example, B5 notes: *"I think about [...] representing different societal groups. So ethnic minorities. More black and white. Maybe a bit more foreign language [...]"*, while L2 says: *" I don't know if this persona is from a specific country or has a specific background but [author] is an American author."* Identifying the presence of different ethnic and/or cultural groups is notoriously difficult, partly because the data required to do so is often lacking. In each of the domains described above, textual data was the only type of data available, and often no additional information was present beside a person's name[2]. In some cases, data enrichment might be possible (B3). Yet, this requires building a dataset on attributes that are often considered sensitive, which gets even more problematic when considering the fact that, to 'expand horizons' or 'represent', a similar type of dataset would be necessary about the users, which would then lead to issues of privacy (L2); see also Papageorgiou and Mougan [39]. Similar safety concerns can be raised for aspects like **gender identity** and **sexuality**; see Pinney et al. [40].

---

[2]Some approaches have attempted to predict a person's ethnicity based on their name, but this process is not generally accurate and can suffer from misrecognition bias [30]

Culture- and **viewpoint** diversity are sometimes mentioned in the same breath: *"for me, diversity means having multiple colors, multiple opinions, multiple cultures, multiple points of view about something."* (B2). We make a distinction between the two, and denote viewpoint diversity as a person's perspective on events. Viewpoint diversity then often becomes linked to politics, or opinions about current events. A rich body of work exists around so-called viewpoint or stance detection, aiming to algorithmically extract these from text [1, 11, 41]. Even when successful, it is often unclear what a diversity of viewpoints would look effectively like, which is further discussed in Section 4.3.

4.2.3   World aspects. Aspects related to the World are beyond the direct control of a recommender system or user. They influence which content is created, and as such the pool of items that a recommender system can make its selection from. They also interact with a user's preferences to determine what type of content a user is interested in, looking for, or should be reading. **Current events** determine what is newsworthy, and what news a user needs to consume to fulfill their information needs. Naturally this aspect is of high importance to the News organization (N1,N2) and to some extent the Broadcaster (B3, B5), but much less so to the Library. In contrast, **society** represents the world as it currently is, including its biases and existing power distributions. An organization may choose to either reflect or counter these existing structures (see Section 4.3.4). For example, the Library acknowledges that much of their catalogue consists of white, American male authors, and wants to account for this in their recommendations (L1).

4.3   Tactics for diversification

During the experiment, the participants considered and deployed different tactics to produce a diverse recommendation. Some participants articulated the tactic they deployed explicitly, while others did not clarify it in as much detail.

4.3.1   Diversity between items. Diversity between items is perhaps the most intuitive interpretation of diversity, and refers to ensuring that all items within a list of recommendations are sufficiently different between them. Following this tactic requires choosing one or more appropriate axes to diversify over. In that sense, constructing a diverse list also entails exclusion, since some aspects have to be deemed less relevant. This issue is particularly important for public organizations; selecting meaningful aspects for which diversity needs to be safeguarded is a crucial decision that must comply with the values and mission of the organization. Multiple interviewees constructed a recommended list by considering diversity between items in the process (B2,B3,B4,B5,B6,N2,L1,L2,L3,L4), especially to recommend items that are diverse between them in terms of category. In particular, the Broadcaster may aim for a *"balance between [e]ntertainment and information"* (B2), given their wide range of offerings and mission *"to inform, educate, and entertain"* (B2). For the Library, item category translates to book genre, and is commonly taken into account when creating a diverse list (L2,L3,L4).

Instead of generally diversifying between items, a somewhat adapted tactic is to recommend a set of items that are different between them in some perspective, but engage with the same topic or theme. For example, one interviewee opted for selecting *"five books that give [...] an insight on [...] LGBTQ [...] [,h]ow that works or is around the world in different cultures and different times"* (L1). Furthermore, in case of a socially relevant emerging topic, the Broadcaster can create *"a highlight lane and then offer a lot of different opinions that people can scroll through"* (B5).

Finally, interviewees noted as a result of the experiment that *"to combine several aspects [is] really hard"* (L2). It might be that a set of items is diverse over one important axis, but not over a different equally important one. In this context the act of creating a diverse recommendation can be seen as an optimization process that an algorithm can contribute to. Regardless, ensuring diversity between items requires some sort of aspect prioritization, as well as an appropriate justification for it.

4.3.2 Diversity as a within-item measure. A different tactic for ensuring diversity is recommending items that consist of diverse perspectives or types of people within the item itself. For example, according to a participant from the Broadcaster, a travel series that allows the viewer to *"see multiple worlds [...] fits very well into diversity"* (B2). This also pertains to episodes of political programs where *"people from a lot of the major parties [appear], which captures a big part of the political spectrum"* (B4). The News organization notes that their inventory consists of *"very good stories which have to be told from a diverse standpoint in the first [place]"* (N1). From this perspective, guaranteeing item diversity may be a part of the production/acquisition process or an explicit step of the recommendation. During the exercise, one interviewee (B4) suggested combining within-item and between items, by composing a list of items on different topics where each item contains multiple perspectives on the respective topic. According to the interviewee, pursuing diversity in recommendation can also be seen as a process of achieving maximum diversity. This tactic requires that the media organization's catalogue consists of items that can facilitate the maximization process.

4.3.3 Diversity considering the user. User-specific diversity is a personalized form of diversity, where the user's history and/or preferences are taken into account when composing a diverse list of items to recommend. The first way to operationalize this tactic is to cater to a user's specific needs that potentially diverge from the norm. This can manifest in assisting the user with finding uniquely niche content. Based on this outlook, diversity is about *"ensuring that [...] everyone feels that there is something for [them]"* (B1) in the catalogue. The Library and News organization also mentioned literacy in this context (L4,N1), as *"how [to] help those with difficulty reading [...] [is] also a diversity topic in a way"* (N1). Additionally, for the Library diversity in accordance with the user can help them *"recognize themselves in the author or in the main characters or the topics [such as] age and physical ability and sexual orientation"* (L2). The second way to apply user-specific diversity is to recommend to a user content that extends further from their current interests, and that they might not be aware of. In other words, an organization can also recommend items so as to *"give people [...] a broader perspective of what is available"* (L3) and to *"broaden the user's horizons"* (B5). This tactic can be seen as a way to ensure that users do not *"stay in their own bubble"* (L2) by *"educating people [that] there's more than [their] bubble"* (B5).

4.3.4 Diversity considering the world. Diversity considering the world entails recommending items that in some way reflect or diverge from the norms that exist in the world, leading to 'similar to world' and 'different from world' diversification tactics. When reflecting the world, one could think of having a good spread of topics that are representative of the important news of that day (N2), or, by representing political parties in a way similar to their distribution in government (B4). With a 'different from world' diversification tactic, the aim is to counter existing power structures. With this interpretation, an item can be considered diverse on its own merit. This can be because the content represents a minority culture, an "outsider [...] in a political sense" (B4), a *"very different part of the world that [the society] knows too little about"* (B2), or even a theme or topic that *"you do not find that many [items] in the catalogue"* (L4). In this case, very popular items may be deemed inherently not diverse, and might not need to receive further exposure: *"[I]t's going to be on the website probably anyway"* (B6).

What constitutes a minority was often implicitly assumed by the interviewees, potentially due to the social context that their organization operates in. For example, multiple interviewees referred to LGBTQ-themed media as inherently diverse (B4,B5,B6,L2), which also prompted their inclusion in a diverse recommendation as part of the experiment. The same can be said for media that gives exposure to people with a minority cultural background or engages with topics not related to the Randstad, a dominant urban area in the Netherlands (N2). From this perspective, diversity

considering the world and user-based diversity (in a broadening-horizons way) can overlap when a user is assumed to be the typical or default representation of the majority culture in the given context.

5 DISCUSSION

The wide range of interpretations highlighted in Section 4 show that, even within this limited domain, conceptualizations of diversity among participants vary greatly. It is therefore unlikely that a standardized definition, encompassing all potential goals, aspects and tactics, can be attained; however, that does not mean that it is impossible to implement a meaningful form of diversity into a recommender system. In this section, we outline the implications the findings of this study have for the implementation of diversity-aware recommender systems.

5.1 Implementing diversity: a normative process

Diversity has traditionally been seen as something we always want *more* of: more viewpoints, more topics, more different nationalities. However, during the interview participants often mentioned instances in which they would actually want *less*. An interviewee from the Library wanted to recommend items that fit the amount of time a user was willing to invest, while the News organization only wanted to recommend items of a certain quality. In these cases, meaningful diversity could be achieved by controlling for one aspect and diversifying over another; for example, when recommending different viewpoints for one controlled topic, or vice versa, by showing the opinions of one particular group over a wide range of topics. The same dynamic also occurs in the way interviewees spoke about diversification tactics. In some cases, a recommendation similar to a user's preferences or history would be desirable, in order for them to recognize themselves in the content on offer. In others, it should be different, so as to expose them to new perspectives. This shows that different applications may not only prioritize different aspects of diversity, they may also have different expectations on whether a recommendation should have high or low diversity.

Participants would often mention that recommendations should be 'diverse enough.' This requires an underlying model that not only determines which items are similar and different, but also informs the system whether sufficient diversity has been attained. This is non-trivial, especially in a domain such as news recommendation, where contexts change rapidly. One could imagine using an external source as a reference point: for example, the presence and size of political parties in government, or a country's composition in terms of cultural groups as determined by a national statistics agency. However, this yet again relies on a conscious decision of what is the 'right' model to use and reflect through the recommendations, and each choice will have pros and cons. There is therefore a normative choice to be made about the type of diversity that is desired, with different aspects to consider, different diversification tactics, and different levels of diversity. The wide variety in the answers given by participants is also an indication that this normative choice is not one that can be taken lightly, and requires a good deal of internal discussion and alignment with all the relevant stakeholders involved before it can be satisfactorily modeled and implemented.

5.2 Generalizing to domains beyond public service media

While the results are not directly generalizable to other domains, there are likely elements of the aspects and tactics identified here that would also be applicable. For example, while music recommendation might put less stock in the diversity of people discussed in an item's content, they would be interested in the diversity of Creators [14]; similarly, while in-item diversity would not be applicable, they could be interested in countering existing biases through different-from-world recommendation tactics [10]. We believe that the goals, aspects and tactics of this paper can still serve as a starting point for discussion in other domains, which may make it easier to identify gaps and unique challenges.

5.3 Exploring versus defining diversity

Participants varied greatly in which diversity aspects and diversification tactics they mentioned. Some of these differences can be traced to the different ways participants speak: some people are more verbose, more inclined to stick to a specific set of examples, or already have a more developed concept in mind than others. Furthermore, the three organizations were at different stages of developing strategies towards diversity, which may have an impact on said differences between participants. However, they are all working on or considering the implementation of a recommender system. Our participants are thus also actively making decisions about how diversity would be conceptualized and implemented. They are as such representative of the 'real' world, rather than the 'ideal' world, and therefore suitable candidates for an exploration of the dimensions along which diversity is currently conceptualized. For these reasons, we refrain from making claims about whether certain aspects or tactics are 'more important' than others, nor do we make claims about the 'correct' definition of diversity. Our hope is that the overviews of goals, aspects and tactics presented in this study will facilitate building a common understanding and vocabulary between stakeholders with a different background, making it easier to find common ground and establish priorities that reflect the requirements of that particular implementation.

6 CONCLUSION

Through interviews with participants from public service media organizations, we identified a range of different interpretations of diversity within a recommender system. We grouped these into different goals (e.g. broadly informing the public or allowing a user to recognize themselves), aspects (e.g. the topic of the content, or the cultural background of an author), and diversification tactics (e.g. ensuring diversity within a single item or countering biases in society). Given the great variety found in the conceptualization of diversity in recommender systems, even within a limited domain, we find that it is unlikely that a standardized definition can be attained. Instead, rather than striving for this standardization, we argue that it should be conceptualized on a case-by-case basis. We hope that our mapping of goals, aspects, and tactics can contribute in effectively communicating what diversity entails within a specific application.

STATEMENTS

Ethical considerations statement

A main ethical concern in interview-based research is breaching the anonymity of the interviewees. To support anonymity, we refer to the organizations by their general functionality instead of naming them. To the interviewees themselves we refer by their organization's functionality as well as a random identifier. Even though we did consider their position in the organization when selecting them, we do not include it when quoting them, as an extra measure for anonymity. Additionally, we asked the participants to sign an informed consent document. We explicitly asked

for permission to record the interviews. Finally, we forwarded to them the conclusions of our research and quotes attributed to them to ensure we correctly interpreted and contextualized their words.

Researcher positionality statement

The authors of this work are all European citizens who operate in the academic circles of the Netherlands. This shaped the work as we interviewed practitioners from the Netherlands who all work for organizations that we are familiar with in a professional and personal capacity, and some are in turn partly familiar with our work, which rendered communication with them easier. Three out of the four authors of this paper work in the computer science field, and one has joint expertise in computer science and communication science. For this reason we also received help (see Acknowledgments) from communication scientists, in particular when it comes to structuring the interview and devising a coding scheme. One author works closely with cultural institutes, one with media institutes, and one with both, which helps us contextualize the statements of the interviewees.

Adverse impact statement

Despite our efforts to ensure anonymity of the individual interviewees, it might be that readers familiar with the media landscape of the Netherlands can deduce which organizations we refer to in this paper. Additionally, readers of this paper might generalize the personal statements of the interviewees as completely representative of the entire organization that they work in. Finally, we do not intend our mapping to serve as a final and conclusive categorization of media diversity, but it might be interpreted as such by readers.

REFERENCES

[1] Abeer AlDayel and Walid Magdy. 2021. Stance detection on social media: State of the art and trends. *Information Processing & Management* 58, 4 (2021), 102597.

[2] Mariella Bastian, Natali Helberger, and Mykola Makhortykh. 2021. Safeguarding the journalistic DNA: Attitudes towards the role of professional values in algorithmic news recommender designs. *Digital Journalism* 9, 6 (2021), 835–863.

[3] Mary Bernstein. 2005. Identity politics. *Annu. Rev. Sociol.* 31 (2005), 47–74.

[4] Abeba Birhane, Elayne Ruane, Thomas Laurent, Matthew S. Brown, Johnathan Flowers, Anthony Ventresque, and Christopher L. Dancy. 2022. The Forgotten Margins of AI Ethics. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 948–958. https://doi.org/10.1145/3531146.3533157

[5] Christina Boididou, Di Sheng, Felix J Mercer Moss, and Alessandro Piscopo. 2021. Building public service recommenders: Logbook of a journey. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 538–540.

[6] Pablo Castells, Neil Hurley, and Saul Vargas. 2021. Novelty and diversity in recommender systems. In *Recommender systems handbook*. Springer, 603–646.

[7] Elizabeth Charters. 2003. The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal* 12, 2 (2003).

[8] David Collier, Fernando Daniel Hidalgo, and Andra Olivia Maciuceanu. 2006. Essentially contested concepts: Debates and applications. *Journal of political ideologies* 11, 3 (2006), 211–246.

[9] Joop Daalmeijer. 2004. Public service broadcasting in the Netherlands. *Trends in Communication* 12, 1 (2004), 33–45.

[10] Karlijn Dinnissen and Christine Bauer. 2023. Amplifying Artists' Voices: Item Provider Perspectives on Influence and Fairness of Music Streaming Platforms. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*. 238–249.

[11] Tim Draws, Oana Inel, Nava Tintarev, Christian Baden, and Benjamin Timmermans. 2022. Comprehensive viewpoint representations for a deeper understanding of user interactions with debated topics. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*. 135–145.

[12] Michael D Ekstrand, F Maxwell Harper, Martijn C Willemsen, and Joseph A Konstan. 2014. User perception of differences in recommender algorithms. In *Proceedings of the 8th ACM Conference on Recommender systems*. 161–168.

[13] Sina Fazelpour and Maria De-Arteaga. 2022. Diversity in sociotechnical machine learning systems. *Big Data & Society* 9, 1 (2022), 20539517221082027.

[14] Andres Ferraro, Xavier Serra, and Christine Bauer. 2021. Break the loop: Gender imbalance in music recommenders. In *Proceedings of the 2021 conference on human information interaction and retrieval*. 249–254.

[15] Dan Friedman and Adji Bousso Dieng. 2023. The vendi score: A diversity evaluation metric for machine learning. *Transactions on Machine Learning Research* (2023).

[16] Walter Bryce Gallie. 1955. Essentially contested concepts. In *Proceedings of the Aristotelian society*, Vol. 56. JSTOR, 167–198.

[17] Andreas Grün and Xenija Neufeld. 2021. Challenges Experienced in Public Service Media Recommendation Systems. In *Proceedings of the 15th ACM Conference on Recommender Systems* (Amsterdam, Netherlands) *(RecSys '21)*. Association for Computing Machinery, New York, NY, USA, 541–544. https://doi.org/10.1145/3460231.3474618

[18] Jaron Harambam, Dimitrios Bountouridis, Mykola Makhortykh, and Joris Van Hoboken. 2019. Designing for the better by taking users into account: A qualitative evaluation of user control mechanisms in (news) recommender systems. In *Proceedings of the 13th ACM conference on recommender systems*. 69–77.

[19] Lucien Heitz, Juliane A Lischka, Alena Birrer, Bibek Paudel, Suzanne Tolmeijer, Laura Laugwitz, and Abraham Bernstein. 2022. Benefits of diverse news recommendations for democracy: A user study. *Digital Journalism* 10, 10 (2022), 1710–1730.

[20] Natali Helberger. 2019. On the Democratic Role of News Recommenders. *Digital Journalism* 7, 8 (2019), 993–1012. https://doi.org/10.1080/21670811.2019.1623700 arXiv:https://doi.org/10.1080/21670811.2019.1623700

[21] Natali Helberger, Kari Karppinen, and Lucia D'acunto. 2018. Exposure diversity as a design principle for recommender systems. *Information, Communication & Society* 21, 2 (2018), 191–207.

[22] Jockum Hildén. 2022. The public service approach to recommender systems: Filtering to cultivate. *Television & New Media* 23, 7 (2022), 777–796.

[23] Dietmar Jannach and Michael Jugovac. 2019. Measuring the business value of recommender systems. *ACM Transactions on Management Information Systems (TMIS)* 10, 4 (2019), 1–23.

[24] Mathias Jesse, Christine Bauer, and Dietmar Jannach. 2023. Intra-list similarity and human diversity perceptions of recommendations: the details matter. *User Modeling and User-Adapted Interaction* 33, 4 (2023), 769–802.

[25] Kari Karppinen. 2013. *Rethinking media pluralism*. Fordham Univ Press.

[26] Kari Karppinen. 2015. The limits of empirical indicators: Media pluralism as an essentially contested concept. In *Media pluralism and diversity: Concepts, risks and global trends*. Springer, 287–296.

[27] Johannes Kruse, Kasper Lindskow, Michael Riis Andersen, and Jes Frellsen. 2023. Creating the next generation of news experience on ekstrabladet. dk with recommender systems. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 1067–1070.

[28] Matevž Kunaver and Tomaž Požrl. 2017. Diversity in recommender systems–A survey. *Knowledge-based systems* 123 (2017), 154–162.

[29] EE Lawrence. 2020. The trouble with diverse books, part I: on the limits of conceptual analysis for political negotiation in Library & Information Science. *Journal of Documentation* 76, 6 (2020), 1473–1491.

[30] Jeffrey W Lockhart, Molly M King, and Christin Munsch. 2023. Name-based demographic inference and the unequal distribution of misrecognition. *Nature Human Behaviour* (2023), 1–12.

[31] Felicia Loecherbach, Judith Moeller, Damian Trilling, and Wouter van Atteveldt. 2020. The unified framework of media diversity: A systematic literature review. *Digital Journalism* 8, 5 (2020), 605–642.

[32] Nora McDonald and Shimei Pan. 2020. Intersectional AI: A Study of How Information Science Students Think about Ethics and Their Impact. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 147 (oct 2020), 19 pages. https://doi.org/10.1145/3415218

[33] Lien Michiels, Jorre Vannieuwenhuyze, Jens Leysen, Robin Verachtert, Annelien Smets, and Bart Goethals. 2023. How Should We Measure Filter Bubbles? A Regression Model and Evidence for Online News. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 640–651.

[34] Margaret Mitchell, Dylan Baker, Nyalleng Moorosi, Emily Denton, Ben Hutchinson, Alex Hanna, Timnit Gebru, and Jamie Morgenstern. 2020. Diversity and inclusion metrics in subset selection. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. 117–123.

[35] Judith Möller, Damian Trilling, Natali Helberger, and Bram van Es. 2020. Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. In *Digital media, political polarization and challenges to democracy*. Routledge, 45–63.

[36] Lynge Asbjørn Møller. 2022. Recommended for you: how newspapers normalise algorithmic news recommendation to fit their gatekeeping role. *Journalism Studies* 23, 7 (2022), 800–817.

[37] Lynge Asbjørn Møller. 2023. Designing Algorithmic Editors: How Newspapers Embed and Encode Journalistic Values into News Recommender Systems. *Digital Journalism* (2023), 1–19.

[38] Christian S Nissen et al. 2006. Public service media in the information society.

[39] Ioanna Papageorgiou and Carlos Mougan. 2023. Necessity of Processing Sensitive Data for Bias Detection and Monitoring: A Techno-Legal Exploration. In *NeurIPS 2023 Workshop on Regulatable ML*.

[40] Christine Pinney, Amifa Raj, Alex Hanna, and Michael D Ekstrand. 2023. Much Ado About Gender: Current Practices and Future Recommendations for Appropriate Gender-Aware Information Access. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*. 269–279.

[41] Alisa Rieger, Tim Draws, Mariët Theune, and Nava Tintarev. 2021. This item might reinforce your opinion: Obfuscation and labeling of search results to mitigate confirmation bias. In *Proceedings of the 32nd ACM conference on hypertext and social media*. 189–199.

[42] Faisal Shehzad and Dietmar Jannach. 2023. Everyone'sa Winner! On Hyperparameter Tuning of Recommendation Models. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 652–657.

[43] Annelien Smets, Jonathan Hendrickx, and Pieter Ballon. 2022. We're in this together: a multi-stakeholder approach for news recommenders. *Digital Journalism* 10, 10 (2022), 1813–1831.

[44] Jannick Kirk Sørensen. 2019. Public service media, diversity and algorithmic recommendation: Tensions between editorial principles and algorithms in European PSM organizations. In *CEUR workshop proceedings*, Vol. 2554. CEUR Workshop Proceedings, 6–11.

[45] Yan-Martin Tamm, Rinchin Damdinov, and Alexey Vasilev. 2021. Quality metrics in recommender systems: Do we calculate metrics consistently?. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 708–713.

[46] Jan Van Cuilenburg. 1999. On competition, access and diversity in media, old and new: Some remarks for communications policy in the information age. *New media & society* 1, 2 (1999), 183–207.

[47] Alje van Dam. 2019. Diversity and its decomposition into variety, balance and disparity. *Royal Society open science* 6, 7 (2019), 190452.

[48] Sanne Vrijenhoek, Gabriel Benedict, Mateo Gutierez Granada, Daan Odijk, and Maarten de Rijke. 2022. RADio – Rank-Aware Divergence Metrics to Measure Normative Diversity in News Recommendations. In *accepted for RecSys 2022*.

[49] Sanne Vrijenhoek, Mesut Kaya, Nadia Metoui, Judith Möller, Daan Odijk, and Natali Helberger. 2021. Recommenders with a Mission: Assessing Diversity in News Recommendations. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia) *(CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 173–183. https://doi.org/10.1145/3406522.3446019

[50] Eva Zangerle and Christine Bauer. 2022. Evaluating recommender systems: survey and framework. *Comput. Surveys* 55, 8 (2022), 1–38.

[51] Mi Zhang and Neil Hurley. 2008. Avoiding Monotony: Improving the Diversity of Recommendation Lists. In *Proceedings of the 2008 ACM Conference on Recommender Systems* (Lausanne, Switzerland) *(RecSys '08)*. Association for Computing Machinery, New York, NY, USA, 123–130. https://doi.org/10.1145/1454008.1454030

[52] Jieming Zhu, Quanyu Dai, Liangcai Su, Rong Ma, Jinyang Liu, Guohao Cai, Xi Xiao, and Rui Zhang. 2022. Bars: Towards open benchmarking for recommender systems. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2912–2923.

[53] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving Recommendation Lists through Topic Diversification. In *Proceedings of the 14th International Conference on World Wide Web* (Chiba, Japan) *(WWW '05)*. Association for Computing Machinery, New York, NY, USA, 22–32. https://doi.org/10.1145/1060745.1060754

[54] Frederik Zuiderveen Borgesius, Damian Trilling, Judith Möller, Balázs Bodó, Claes H De Vreese, and Natali Helberger. 2016. Should we worry about filter bubbles? *Internet Policy Review. Journal on Internet Regulation* 5, 1 (2016).

A CODING SCHEME AND MENTIONS

Table 1. Final coding scheme, containing both aspects and tactics. Aspects are divided into Item, Person and World aspects.

Aspect	Feature	B1	B2	B3	B4	B5	B6	L1	L2	L3	L4	N1	N2
Item	category / genre	x	x	x	x	x	x	x	x	x	x	x	
	complexity		x	x	x						x	x	
	cost				x								
	creator (person)	x		x	x	x		x	x	x	x		
	geographic location	x	x	x	x	x	x	x		x		x	x
	language		x	x				x					
	newsworthiness / quality	x	x	x	x	x		x			x	x	x
	medium												x
	subject (person)	x	x	x	x	x	x		x		x		x
	popularity	x	x	x	x	x	x	x		x		x	x
	publication date	x		x	x			x				x	x
	recurring		x									x	
	relevance	x	x	x	x	x		x		x		x	x
	sentiment					x							
	target audience		x	x									x
	time investment									x	x		
	time period discussed				x	x		x					
	topic	x	x	x	x	x	x	x	x	x	x	x	x
Person	ability								x	x			
	age	x	x	x	x			x	x	x			
	cultural background		x		x	x	x	x	x				
	education				x							x	x
	ethnicity		x	x	x				x				
	gender identity		x	x	x	x		x	x	x			
	geographic location	x	x	x	x		x	x		x		x	
	living situation			x	x	x	x		x			x	
	nationality		x	x	x	x			x	x	x	x	
	religion	x	x	x	x	x	x			x			
	sexuality	x			x	x	x	x	x	x	x		
	viewpoint	x	x	x	x	x	x	x	x	x		x	
	user; history	x	x	x	x				x	x		x	
	user; preferences	x		x	x	x			x	x		x	
World	current events				x		x					x	x
	society				x	x				x		x	
Tactics													
	different in item		x		x	x	x					x	
	different in list	x	x	x	x	x	x	x	x	x	x		x
	similar to user	x		x		x	x	x	x	x	x	x	x
	different from user	x	x			x	x		x	x			
	different/similar to world		x		x	x	x	x			x		