Do We Care About Personalization and Explainability? An Interview Study with News Recommendation Engineers

Jasmin Kareem
University of Amsterdam
Amsterdam, Netherlands
Jheronimus Academy of Data Science
's-Hertogenbosch, Netherlands
j.kareem@tue.nl

Siddharth Mehrotra
Birla Institute of Technology & Science, Pilani
Pilani, India
siddharth.mehrotra@pilani.bits-pilani.ac.in

Martijn Willemsen
Jheronimus Academy of Data Science
's-Hertogenbosch, Netherlands
Eindhoven University of Technology
Eindhoven, Netherlands
m.c.willemsen@tue.nl

Maarten de Rijke
University of Amsterdam
Amsterdam, Netherlands
M.deRijke@uva.nl

Abstract

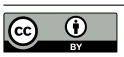
Research on explainability in recommender systems largely centers on end users, overlooking the perspectives of those who build and maintain these systems and their potential use cases such as model debugging. In this study, we examine how news engineers and related technical stakeholders perceive and implement personalization and explainability in practice. We conducted 15 semi-structured interviews across nine news organizations, spanning diverse regions in both public and private sectors, to investigate the challenges and motivations shaping their approaches. Our findings reveal that personalization is not always a straightforward or desirable choice for news organizations, as concerns around user tracking, editorial control, and resource constraints often limit its adoption. Even among organizations implementing personalized news recommender systems in production, explainability is rarely prioritized, with day-to-day operational demands frequently taking precedence over longer-term transparency goals. Definitions of explainability vary widely across organizations, though some demonstrate promising internal practices and visualization tools that facilitate communication between engineering teams and newsrooms. Based on our analysis, we provide actionable and practical guidelines for news engineers and researchers on how to adopt explainability methods within a news personalization pipeline.

CCS Concepts

• **Human-centered computing** → **User studies**; • **Computing methodologies** → **Machine learning**; • **Information systems** → **Personalization**; **Recommender systems**.

Keywords

News recommendation, Personalization, Explainability



RecSys '26, Minneapolis, MN, USA

ACM ISBN 979-8-4007-2284-4/26/09
https://doi.org/10.1145/3773078.3831744

ACM Reference Format:
Jasmin Kareem, Siddharth Mehrotra, Martijn Willemsen, and Maarten de Rijke. 2026. Do We Care About Personalization and Explainability? An Interview Study with News Recommendation Engineers. In *20th ACM Conference on Recommender Systems (RecSys '26), September 27–October 02, 2026, Minneapolis, MN, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3773078.3831744

1 Introduction

Personalization in news recommender systems has received growing attention from both academia and industry due to its potential to connect diverse audiences with relevant news content at the right time [63]. Reflecting this trend, research on algorithmic approaches to news recommender systems has expanded rapidly. This body of work has largely focused on optimizing technical performance, while more user-centered and "value-aware" perspectives remain comparatively underexplored [7]. This imbalance is concerning, as news recommender systems play a central role in shaping public discourse and supporting democratic processes [22].

Values in news recommender systems (NRSs). A growing line of research emphasizes the importance of embedding societal and editorial values into news recommendation [28, 43]. E.g., Vrijenhoek et al. [60] demonstrate how normative values can be operationalized in news recommender systems. Building on this, Lu et al. [32] incorporate editorial values into personalized news recommendations and evaluate their approach through user studies.

News recommender systems and explainability. Next to diversity, transparency has emerged as a key value in the context of news recommendation. Understanding the mechanisms behind recommendation algorithms is essential for maintaining trust in journalism and news organizations [14, 41]. Explainability techniques are commonly proposed as a means to increase transparency, traditionally focusing on helping readers understand why a particular news article was recommended to them [31, 40, 54]. However, explainability encompasses multiple definitions and goals; Tintarev and Masthoff [55] identify several explanatory goals, including transparency, which aims to explain how a system works, and scrutability, which allows users to indicate when the system is wrong. For

example, Sullivan et al. [51] show that explanations grounded in user profiles can support readers' sense of self-understanding and agency. However, end-users are not the only target group that require transparency, as there are multiple stakeholders that could benefit from explanations [34].

Explainability for news engineers. Recent work has begun to explore alternative roles for explainability, particularly as an internal tool for news organizations. Cools [13] examine the internal and external uses of explainability at the BBC, highlighting tensions between editorial goals and technical constraints. More broadly, there are further motivations for interpretability, including model debugging, bias detection, and trust-building [12]. Prior studies show that interpretability tools can support model debugging [8], yet even expert users such as data scientists may misunderstand or misuse explanations [27]. These findings underline the importance of examining how explainability is understood and applied by engineers, who are ultimately responsible for deploying recommender systems and explanation mechanisms in practice [4].

There is a clear need for qualitative research that captures the perspectives of engineers working with news recommender systems in real-world settings [7]. Interview-based studies have been conducted in adjacent domains, such as evaluation practices in news [58] and fairness and serendipity in music recommendation [9, 15], but work focusing specifically on explainability in news recommender systems remains limited. Our study differs from prior qualitative work [13, 39] by centering explainability within news recommender systems and by interviewing engineers across multiple news organizations in a variety of countries.

Research questions. We explore the current state of personalization and explainability practices in an internationally diverse set of news organizations, guided by the overarching research question: *Can explainability methods address the challenges that engineers face when building news recommender systems?* To answer this question, we investigate the following research questions:

- **RQ1** To what extent is personalization adopted by engineering teams at different news organizations?
- **RQ2** What are the challenges that these engineering teams face?
- **RQ3** To what extent have engineers adopted explainability for recommendations?

Contributions. First, we provide insights into the adoption of personalization practices across nine news organizations and identify key challenges faced by engineering teams. Second, we examine how explainability methods are currently used, if at all, within these organizations, and outline the reasons for their adoption or absence. Third, we offer practical guidelines for researchers and practitioners on how explainability methods can be more effectively integrated into news recommender development and deployment.

2 Related Work

We outline prior work on journalistic values in recommendation pipelines, explainability, and studies that investigate user needs of engineers for explainability.

2.1 Journalistic values in NRSs

News connects people to information on current events and directly affects the functioning of democratic societies [22]. As online news production accelerates, news recommender systems help filter information to the right person at the right time, using a variety of approaches explored over time [63]. NRS research has developed rapidly, yet the impact on improving experiences of recommendations for a user's every day life is not so clear [24], leading to disparities between what is researched and what is practised. Recent work shares lessons learned from *Der Spiegel* to inform the development of more responsible NRSs through collaborations between industry and academia [16]. Alaqabawy et al. [2] argue, based on interviews with journalists, that a multi-stakeholder approach to news recommender systems is needed, with journalists as first-class users. Smets et al. [47] conduct 11 interviews with professionals from media organizations, also finding that news recommender systems are formed through multi-stakeholder communication. A related study examines organizational dynamics in adopting news recommender systems across 10 media organizations, highlighting tensions between journalistic, market, and technical stakeholders [39]. Other work shows how journalistic values are embedded in NRSs through interviews with technical and non-technical stakeholders at two newspapers [41]. Diversity-aware approaches are another way of ensuring a large variety of opinions are shown to readers [21]. Michiels et al. [38] re-define filter bubbles as a decrease in any dimension of diversity. These diverse aspects can also be reflected in beyond accuracy metrics, with different metrics being relevant for different stakeholders [58] and transparency as an important driver [18]. In this work, we provide novel insights into how news organizations embed journalistic values in their recommendation pipelines.

2.2 Explainability in recommender systems

Explainability and transparency in recommender systems have a long history [23, 55], but initially focused on explanations for users (external explainability). More algorithmic approaches to explaining recommender system decisions have recently gained popularity [64], and thus the evaluation of explanation quality gained popularity too [25, 44]. Early methods in explainable AI, e.g., the post-hoc, local explanations LIME [45] and SHAP [35], have been adapted for recommender systems in a collaborative filtering setting [42, 65]. Counterfactual explanations are another post-hoc approach that focuses on explaining collaborative filtering-based recommendations [6, 56]. In contrast, Ariza-Casabona et al. [3] focus on generating textual explanations with LLMs.

Inherently explainable recommendation methods often leverage knowledge graphs to provide transparency. Key approaches include using reinforcement learning for path reasoning [53], hierarchical feature learning for interpretable modeling [17], and specialized knowledge reasoning for news [26]. In the content-based NRS domain, Liu et al. [31] use topic modeling to generate explanations and boost performance, while Möller and Padó [40] apply integrated gradients to attribute click behavior to a specific reading history. Additionally, Spišák et al. [48] employ sparse autoencoders to interpret article embeddings, facilitating editorial decision-making at *The Telegraph*.

Recent algorithmic advances [61] have spurred growth in user-centered explainability. Szymanski et al. [52] map explainability needs to stakeholder expertise, while Mehrotra et al. [37] demonstrate that integrity-based explanations foster appropriate trust.

Tran et al. [57] conducted a large-scale study on user needs across tourism and hospitality datasets, finding that in low-stakes scenarios, users prefer simple, positive explanations [1]. Conversely, in job recommendation, requirements vary significantly by stakeholder [46]. Finally, Waterschoot et al. [62] suggest that detailed explanations in group recommenders add little value if the system is already transparent. Our work extends this earlier work to the NRS domain and investigates if news organizations have considered or are using explanations, both internally (for engineers) and externally (for their users).

2.3 Explainability for engineers

Bhatt et al. [8] find that engineers primarily use explainability tools for debugging and model development across domains. However, Kaur et al. [27] note that data scientists often over-trust and misuse these tools, complicating their deployment. Addressing the development process, Sporsem [49] explores how requirements engineering strategies can elicit explainability needs to improve transparency. Building on this, Habiba et al. [20] interview 14 professionals to define explainability in AI-based systems and identify the practical hurdles practitioners face.

Expanding the scope to include both practitioners and end users, Liao et al. [30] investigate how different stakeholders engage with explainability. The appropriateness of explanations is highly dependent on the underlying questions being asked, and that practitioners often struggle to bridge the gap between algorithmic outputs and explanations that are meaningful and interpretable to humans. Balayn et al. [5] examine how multidisciplinary teams, including engineers, develop trust in their organization's LLM supply and Mehrotra et al. [36] examine the effect of different explanation types (text, visual, and hybrid) on building appropriate trust with police officers who develop a predictive policing system. From a user interface design perspective, Brachman et al. [10] explore a personalized, generative-model-based user interface for model debugging and find that even within a single user group of engineers, factors such as background and task goals influence the types of explanations required.

Within the domain of news recommender systems, research focused on analyzing the explainability needs of engineers is sparse. Cools [13] conducts 22 interviews with mostly technical stakeholders in a single organization (BBC) and finds that although transparency and explainability are formally articulated in internal policies and AI principles, their interpretation and operationalization varies across teams. Our study extends this line of inquiry by adopting a cross-organizational perspective among different international media organizations and for different internal roles (engineers and product owners). In doing so, we present what we view as both current gaps and best practices for explanations in the field.

3 Methods

We conducted 15 semi-structured interviews with engineers and product owners at varying levels of seniority across nine news organizations; see Table 1. Organizations were selected to have a broad international sample of institutions, i.e., private/public and news organization/aggregator. A thematic analysis was conducted on the transcribed interviews following guidelines outlined in [11].

Table 1: Organizations grouped by media type with team size (S/M or L). Team size is categorized as small-to-medium (S/M, ≤ 10) or large (L, > 10). Team size is approximate, as team definitions and responsibilities vary across organizations.

Media organization type	Organization (Team size)
Public broadcaster	NOS[1] (S/M), BBC[2] (L)
Newspaper/Magazine	LA Times[3] (S/M), DER SPIEGEL[4] (L)
Media conglomerate	JP Politiken[5] (S/M), DPG Media[6] (S/M), Schibsted[7] (S/M)
News aggregator	Ground News[8] (S/M), Cafeyn[9] (S/M)

The study was reviewed and approved by the Ethics Review Board of Eindhoven University of Technology. All participants received an informed consent form and provided consent prior to participation.

3.1 Interview setup

All interviews were conducted in English and semi-structured. Interview questions were initially based on the interviews conducted by [8, 33, 34]; we followed the guidelines in [30] on designing interview-based studies. All interviews were conducted by the same person in an online meeting format. Recordings were made using either Zoom or Google Meet. The duration of the interviews were planned for 30 minutes and varied between 22 and 56 minutes; see Table 2. The questions were structured into four phases. We started with warm-up questions. These included questions such as:

(1) Can you describe your role, how long you've been with the company, and what you're working on?
(2) What does your team do?

Next, we moved on to questions relating to the state of **personalization** at these organizations and **communicating with internal stakeholders**. These include:

(3) Do you incorporate personalized recommender systems? Why or Why not?
(4) If so, what does that look like?
(5) What is your recommender system development workflow?
(6) Can you explain what type of model it is? What data is it trained on or what data do you use?
(7) Which stakeholders do you need to communicate with within your organization?

The following stage included questions about the **challenges** that the individual and their team faces, including:

(8) What are your pain points in deploying recommender systems?
(9) Does your model ever fail or produce strange results? How do you identify these types of results?
(10) Think of an instance when you had to debug a model recently. What were some challenges you faced in doing this?
(11) How did you overcome these challenges?

Lastly, we asked questions on whether **explainability**, for internal or external use, could address these challenges, following the definition of internal and external explainability in [13]. These include:

[1] https://nos.nl/ [2] https://www.bbc.com/ [3] https://www.latimes.com/
[4] https://www.spiegel.de/ [5] https://jppol.dk/en/
[6] https://www.dpgmediagroup.com/ [7] https://schibsted.com/
[8] https://ground.news/ [9] https://www.cafeyn.co/

Table 2: Participants grouped and labelled by team with interview durations (in min) in parentheses.

Team	Identifier (Duration in mins)
Data Science	D1 (30), D5 (38), D7 (56), D11 (38)
Product	P2 (28), P3 (28), P10 (22), P14 (30), P15 (27)
Engineering	E4 (42), E6 (31), E8 (27), E9 (27), E12 (30), E13 (33)

(12) How do you try to understand your model? What do you use to try to understand your model?
(13) If so, were these methods useful to you? Why or why not?
(14) How familiar are you with Explainable AI research and tools?
(15) If so, have you considered using these tools? Why/Why not?

Our goal was to obtain a broad overview of the challenges faced by news engineers. Accordingly, questions were adapted to organizational contexts, including current uses of LLMs in news and recommendation systems, as well as related applications such as article generation.

3.2 Participants

We recruited interviewees through our existing research network via email, LinkedIn and used snowball sampling to identify further contacts. We conducted interviews until thematic saturation was reached [19], ensuring coverage of diverse organizations, including public and private news institutions and news aggregators. Table 2 presents anonymized participant identifiers and team specifications. Team categories are derived from self-reported job titles (e.g., "Data Scientist" → "Data Science"), and abstracted team information is shown to protect participant anonymity.

3.3 Thematic analysis

We conducted a thematic analysis in MAXQDA 2024 [59]. One researcher first cleaned and coded all transcripts, after which a second researcher independently recoded them using the same code system. The initial agreement was 84.38% (Cohen's Kappa $\kappa = 0.80$) [29], prompting a discussion that resolved disagreements in 81 coded segments and increased agreement to 96.2%. We then used axial coding to develop themes [50], iteratively grouping codes into themes and categories guided by our research questions.

4 Results

In this section, we report the results to our stated research questions. Participants are referred to using the identifier codes in Table 2.

4.1 To what extent is personalization adopted?

The first part of the interviews pertained to understanding if and how personalization exists in NRSs. We group organizations into those that use personalization in news recommendation and those that do not. We then discuss reasons why organizations are adopting or avoiding personalization. Lastly, we analyze our results from a cross-organizational perspective, using the aforementioned groupings to define personalization levels in Figure 1 (left).

Organizations using personalization. Most organizations, seven out of the nine interviewed, stated that personalization is a part of the wider user experience, on the website, in an app, or in a personalized newsletter. Implementations of personalization pipelines differ. E.g., P3 indicated their organization has a unified approach: "*we can use the same recommendation engine on the front page as we can in a box underneath an article in a widget.*" Other organizations have a segmented approach "*For every purpose there is a different algorithm*" (D7). In between these two perspectives, three participants mentioned that their personalized news pipelines mainly contain two or three types of approaches: content-based news recommendation, collaborative filtering, and/or content-to-content based recommendation which is non-personalized. E.g., D1, P2 and E8 mentioned a personalized approach and a non-personalized approach, with E8 stating "*We currently have two flavors of personalization. We have the content-based personalization [...] And then we have a collaborative filtering flavour.*" Similarly, D1 noted that "*one is personalized and one is not personalized, but it's based on what is performing well, and showing that content to more users. We've had really good success with that latter bucket.*" Interestingly, P2 also mentioned non-personalized approaches already having a positive impact for the organization: "*And then a lot of 'read more' at the bottom of an article [...] But even that was a big win in automating that for the news brands.*" In the case of D11, the recommender system is outsourced, using "*out of the box solution*" that is"*very much content-based with some business rules for re-ranking and filtering.*"

Organizations not using personalization. A smaller group of organizations indicated, at the time of the interview, not to incorporate any personalization in their pipeline. E.g., similarly to the content-to-content method mentioned by D1 and P2, P10 stated that "*when a user opens a certain article, we're trying to find articles that are similar to that article that they're reading right now [...] So, recommendations are also the same for everyone who opens that article. It's not personalized.*" Another organization recently moved away from personalization, with E9 stating "*We did that for six months. We actually recently just removed it [...] That is why news recommendation is a very interesting challenge, or a product for news because it goes against the ethos of our organization in a way.*" As an alternative, E9 mentioned a larger focus on categorizing articles and allowing users to filter based on their interests and that "*everybody sees it as a newspaper. Everybody gets the same treatment.*"

Reasons for moving towards personalization. Participants' responses frequently prompted follow-up questions about the rationale behind organizational decisions and whether investments in personalization would continue. E6 and D7 noted that their organizations are moving forward with automating the front page. D7 described how a previously manual newsletter, where "*everyone got the same*" became partially automated after experimenting with personalization. This was particularly effective for topics the editorial team "*simply stayed out*" of, such as cryptocurrency, which was not considered relevant for most users but "*nice to get*" for a minority, marking "*the first step of replacing humans*" (D7). Other organizations fall somewhere in between automation and editorial control (P2, P3, D11). P3 eluded to the balance between editorial and personalized news with "*we're trying to check off sort of all the possible contexts while still safeguarding the journalistic mission because we aren't YouTube, we aren't Instagram, we aren't TikTok.*

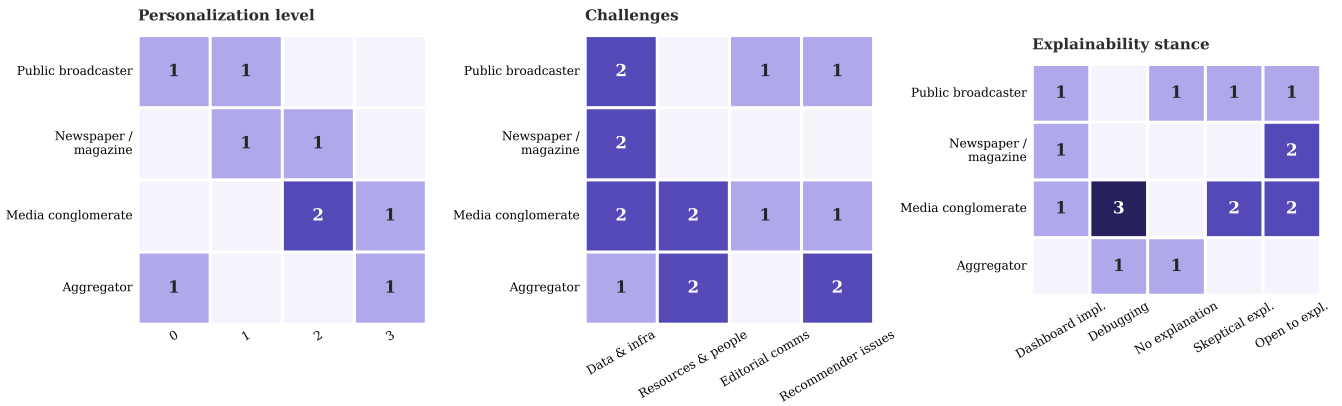


Figure 1: Cross-organizational overview of key findings across RQ1–3. Each cell reports the number of organizations of a given type exhibiting that characteristic. (Left) Personalization levels range from 0 (no personalization) to 3 (pro-personalization). (Middle) Challenges reflect infrastructural and organizational barriers to recommender system deployment. (Right) Explainability stances capture how organizations approach or resist algorithmic transparency.

We are serving news to the public and still have editors who at any time decide what's most important that the public should know now."

Reasons for moving away from personalization. Whilst most organizations have personalized approaches incorporated in their products, some deliberately choose to avoid personalization. P10 stated "*one of the reasons is we're just not a tech company*", and continued with "*The other reason why we're keeping it simple is because we don't want to gather a lot of personalized data.*" E9 cited the issue of echo chambers as a reason not to personalize, however, mentioning that they may bring it back "*Maybe in the future recommendations will come back [. . .] For news, it's kind of a slippery slope.*" Even for organizations that fall somewhere in between this spectrum, there are questions on how much to invest in personalized experiences. D1 mentioned "*We're also at a bit of a fork in the road of whether we're going to stick with that [. . .] do we want to really invest a lot more in this system and make those kinds of improvements?*" Another organization reported prioritizing personalization for its video platform, with news taking a back seat. D5 explained that users may view the platform similarly to a Netflix subscription, where payment implies personalized value, making personalization "*one of the top priorities*" for their video platform.

Zooming out with Figure 1 (left), we find that personalization in news is less widely adopted by public broadcasters than by other types of media organizations. This is unsurprising, as public broadcasters typically lack access to user data comparable to that of privately funded organizations, where logins are often tied to subscription-based funding models. More surprising, however, is the variation within privately funded organizations: one aggregator and one newspaper/magazine fall at the lower end of personalization, while the others exhibit higher levels of personalization.

4.2 What are the challenges?

Drawing on questions 8–11 from Section 3, we identify four categories of challenges related to personalization and discuss their impact on different media types.

Data, backend and infrastructure. When asking participants about their challenges, the overwhelming majority pointed to data, backend or infrastructure as a key concern (P3, E4, D5, E6, D7, P10, D11, P15). E.g., D7 said that "*most of the time your backend land is pretty disjoint from your data land*" and D11 stated that these challenges made it "*a little bit hard to debug the system.*" E4 indicated that transitioning from one infrastructure to another causes issues, with their recsys team becoming "*a centralised unit*" while building "*a huge new data structure.*" P10 mentioned being bottlenecked by old infrastructure. P3 emphasized access to data and subscriber retention, noting that "*the GDPR makes it really hard to get good data that would benefit society. We believe that it would benefit society if more people would have a newspaper subscription.*" Improving retention is also a concern for P14: "*our biggest pain point is that we want to use recommendations to keep people on the platform, and we can optimize and measure relatively short term results like clicks and plays.*"

Resources and people. Another stated challenge by participants (P3, E4, D7, E9) was access to resources and people. E9 mentions the "*big computational challenge*" of running recommender systems and D7 stated that their"*current data science team, it's four people. We don't have this in-depth knowledge. We don't have a specialist. Because at this point, we need generalists*[10] *who can make changes on the product and not only focus on optimising specific parts.*"

Communicating with editorial teams. Communication and talking to the newsroom often came up as a challenge (P2, D5, P10, P14). E.g., P2 referred to informing newsrooms about articles that were generated and how that is difficult because they are "*quite a big company with also a lot of silo work within those news brands, not everyone is always up to date.*" P14 mentioned editorial reviewing changes, which can also slow things down, and P10 mentioned they "*always want to make sure that people like our colleagues who are editors know that we are not planning to replace their work with AI.*"

Issues with the recommender system. Lastly, several participants reported pain points with how the recommender system actually worked (D5, E9, E12, P14). Some argued that different products require different recommender systems (D5) and indicated that

[10] Hiring constraints force a focus on generalists, even though new technology requires specialists.

they are currently not using "*cross-product interactions*." In contrast, E9 mentioned diversifying recommendations and how this is particularly challenging with news where recommendations should not only focus on what is trending. E12 mentioned issues with news categories like sports where there are either "*too many*" or "*too few*" articles for different types of user. P14 found complex business rules challenging, stating that "*we've got lots of business rules to constrain our recsys outputs. So, for instance, if you liked a politics show isn't a thing we'd ever show because if you like, a certain political persuasion isn't what we want to promote. So we'd have a business rule. And if you liked, well, that would say, you know, here are the recommendations. Now strip out any political titles, but then a different recommender might have a very similar rule, but it will be written in a different way. Not one is right or wrong, but just different teams at different points.*" The cold start problem in news came up with some participants as a key challenge (D7, E8, P14). Other participants mentioned solutions such as LLMs (P3) and using non-personalized bandit algorithms for exploratory recommendations (P15).

From a cross-organizational perspective (Figure 1, middle), data, backend, and infrastructure emerge as universal challenges for engineers. Resource and staffing constraints might be expected to be widespread, but they are reported only by aggregators and media conglomerates. Challenges related to recommender systems are most prevalent among organizations with higher levels of personalization.

4.3 To what extent is explainability adopted?

This study assessed how engineers adopt explanations, allowing participants to define the concept in their own terms. No single organization explicitly stated to use explainability or interpretability algorithms for any purpose. This does not mean that organizations do not have a means to debug models, to communicate with the newsroom or to be transparent on their websites or apps. We identified five categories of explanation (related) practices and vision (see Figure 1 (right)): two focus on internal explainability (dashboards and debugging), the other three about external explanability (no explainability, skeptical about explainability, and open to explainability in the future).

Dashboards for the newsroom. No organization explicitly uses explainability tools but three mentioned having a dashboard for internal use, specifically to communicate and monitor changes to the recommender system. P3 mentioned "*the editors that we bring in the projects, they get to see everything. We try to explain how the model works, the feature significance. We have different tools where we can look up articles and say alright I understand that you would think that this article would perform much better but when we look at the data and the features on this we see that it didn't really attract the young users as you had expected.*" P15 and D5 interact with editorial teams through metrics and dashboards: "*We use a combination of metrics [...] including diversity metrics, coverage metric [...] but very often we have an internal review with bespoke visualisation tools, which are used to show recommendations to editorial. And they're very helpful to spot inappropriate pairings*" (D5). P15 noted "*now I can run a dashboard and a report, and I can show you that, in fact, we are not doing what you asked and that we are representing our values inside of the system with real data, with real dashboards. This was very compelling.*"

Debugging approaches. When asked questions 12–16, some participants mentioned debugging approaches and how they understand their model. E6 mentioned explainability for "*Just for the other IT people*" and that "*we make an effort like in the end we want to transform all the scores so they're between 0 and 1.*[11] *Because then when you see the raw data, you can tell what a high score & a low score is. So that's one type of explainability. It's just useful for the other IT people.*" E8 and E9 do manual checks to understand where things went wrong, such as "*we just select a few users and we do some checks with the models*" (E8) and "*one of the ways we debug is to look at what was the previous user's behavior and what other users who are closest to this user, on what their behaviors are.*" When referring to engineers using feature attribution for debugging in their team, P3 said "*they probably use everything*;" usage is currently unknown, but engineers are free to use the tools as needed.

No external explainability. Two organizations stated that they would not use explanations externally, although for very different reasons. D7 stated "*I think we discuss it twice a year. I'm not fully convinced this is the thing that our users want. [...] This is shared by the product team. So we do discuss and it does come up, I know the techniques are there. We could do something about it. But it's not a way that we can differentiate [...] from a business perspective. But I also think that there's only a small group of users who actually would get the benefit from it.*" The motivation for not incorporating explanations is linked to the history in recommender systems with user controllability and customizability, further mentioning "*Every news site tried it out. Every news site disbanded it, that users get this full control. Because the moment that you give them control, it means that it makes it much harder for you. If you want to take control, to push something that you think is important. Either from a journalistic perspective or from something that needs to be pushed. It's the same here.*" In contrast, E13's reason for not including explanations is that if you avoid personalization, you also avoid black box scenarios where explanations are necessary "*I think by designing it the way we did now, explainability is not like a hard requirement.*"

Skeptical about external explainability. Other engineers acknowledged that explanations have some concerns (P2, E4, E6, P10, E12). P2 mentioned that there may be fear from newsrooms with complaints from readers, stating "*So it's like don't poke the bear if there's no reason to consider people being unhappy with the lists*" where lists are the algorithmic outputs such as ranked news feeds or content suggestions. E4 and P10 questioned whether this is what users even want: "*Within the first year [of deploying the recommender system] we had one person contact us. And that was some student that was interested in it. So, I'm not sure how many people actually care about it*" (E4) and "*information overload if you place an explanation with every recommendation*" (P10). E6 and E12 mentioned risks with explanations stating "*It can always be done badly*" (E6) and "*it can be manipulated a lot*" (E12).

Open to external explainability in the future. The last group of participants seemed open to the idea of incorporating explanations to readers in the future (D1, E6, D11, P14). D1 mentioned a potential

[11] Here, the 0 and 1 refer to the normalized bounds of a model's output score.

"*next project is like a specific module that's like 'because you read'.*" Again, we see a connection with personalization and a growing need for explanations, with E6 stating "*as you take more and more [of the front page], this explainability is in focus, both for our own sakes so we can know what's going on, but also for the end user.*" For P14, generic explanations have already been implemented, such as "*based in the area*," but more personalized explanations have not been tested. For D11 there is a lot of interest within their team, mentioning a researcher that "*is very interested in transparency. So communicating the fact that it is personalised to the users. So we will do some experiments with this on the website this year. And we are also interested in a controllability. And that is a little bit related to explainability. It depends a little bit on your definition of explainability.*" For most participants, a common reason why explainability has not yet been incorporated, even if they are positive about it, is a prioritization for current projects, e.g., optimizing current recommender systems (D1, E6, D11, P14), and internal challenges, e.g., data infrastructure (D11, P14) and lack of time (E6).

A cross-organizational view of explainability adoption (Figure 1, right) reveals broad receptiveness, with only a subset of organizations skeptical of its utility. A clear pattern emerges among media conglomerates: all cite debugging as a key lens for interpreting recommendation results, reinforcing the earlier observation that personalized recommendation pipelines are prevalent in this sector.

5 Overarching Themes

Three themes emerged: newsroom influence on personalization, a lack of stakeholder explanations, and unexpected findings outside of the original scope.

5.1 Editorial control

In all organizations except news aggregators, editors play a central role in defining brand values, shaping NRS design, and influencing system transparency. Even the two aggregators we interviewed apply business rules grounded in their organizational values. Communicating these editorial values to the newsroom can be challenging for engineers (see Section 4.2). Despite this, news engineers realize that personalization in news has valid constraints due to the journalistic values (P3: "*Editors decide what the public should know*") and privacy concerns of the organizations (P10: "*don't want to gather personal data*"), which are expressed by editorial teams. Because of these constraints, personalization may not need to be as developed as in other domains such as e-commerce and movie recommendation (P14: "*we rely on editorial stakeholders for a lot*").

Values that are embedded into the fabric of the organization are reflected through the business rules (D11: "*we have developed them together with the editors*"). Almost all participants mentioned them as a crucial part of developing their recommender systems (D1, P2, P3, D5, E6, D7, E9, D11, E12, E13, P14, P15). An example of editorial control through business rules is a temporal cutoff (D1: "*So weather stories [...] they have a very clear expiration. And we do have mechanisms for kind of editorial control over that*"). Another example of business rules in recommendations are sensitive news topics. D5 mentioned an anecdote on issues with the editorial and a content-based news recommender system: "*[there were] arson attacks at a synagogue in New York [...] the recommender linked that article immediately with something about Holocaust Memorial Day. But I don't want to link everything that's related to Jewish people to the Holocaust. And so editorial told us, well, this might be inappropriate.*"

Business rules are not just a part of the recommender design, but they clearly tie in with the results on explanations in Section 4.3, where dashboards in the newsroom can be used to monitor and possibly develop business rules (P3, D5, P15). Identifying problems or bugs in recommendations can also motivate and inform new business rules. For example, E12 mentions where they are "*playing around with the 'most read' list because there is actually a lot of things that you can optimize on [...] usually they just say you should count [the last] 24 hours, but if you take at 4 [pm] today, then the most popular article is the one that was popular yesterday.*"

Another sub-theme across participants was the importance of diversity in recommendation (D1, E4, D7, E9, D11, E12, E13, P14, P15). Diversity emerges as a challenge, a business rule, an evaluation criterion, and a core objective of the recommender system, with multiple dimensions shaping what counts as a diverse recommendation. E.g., E13, who does not personalize their NRS, said that "*even if it's not the perfect recommendation, it's still kind of fine because we generally also just want to present the user with more content that they might be interested in*". P15 related a lack of diversity to news fatigue, i.e., too many articles about wars may lead to "*a decline in the visits per user.*" There could also be more diversity in article stances on the same topic or from different sources (D7, E9).

5.2 Explainability vs. Understanding

Even without formal explainability tools, engineers still seek to understand and communicate how their recommender works. As Section 4.3 shows, most debugging is manual and unsystematic. More analytic monitoring could involve feature attribution, counterfactual explanations, or examining article embedding spaces, an approach some engineers explicitly described when trying to understand the high-dimensional space of article text (D1, P2, E9, P10, D11, E12). E12 said "*sport articles are very easy if you take the embeddings and then you visualize it, then sports is very easy to classify but everything else is kind of all over the place.*" Engineers do not always find manual checking problematic, but some are more eager to improve: "*I don't think we have a method to entirely capture all things that could go wrong. And I'm not sure if it even exists. Because if it does, it would be great to know about it*" (P10).

Interestingly, when talking about explainable AI, many engineers pivoted to the 'old school' recommender systems explanations for end-users that could be found on Netflix (E6: "*Sort of like Netflix [...] we have talked about that*") or Facebook (E12: "*explain it by 'your friends are also looking at'*") [55]. Surprisingly, algorithmic approaches were not favored, despite aligning better with technical stakeholders. We see from discussions on bugs in recommendations (Section 4.2) and investigations into the article embedding space that there is a desire for tools that can better equip them to understand their models better and not necessarily to explain to users.

5.3 LLM Adoption

While outside our original research questions on personalization and explainability, LLM adoption emerged consistently enough across interviews to warrant separate note: many organizations are adopting LLMs, for article generation or editing (P2, E4, E6,

D7, E13), generating meta-data (D1, D7) or as an internal tool (P2, E4, E6, E12). Regarding LLMs as an internal tool, E4 said for now "*the target customers are the journalists and the editors.*" E12 was skeptical of the tool for model debugging: "*Everybody says that we should use the chatbot and I really tried and it took me on so many goose chases [...] it made mistakes that took me longer time to find than it probably would have done if I hadn't tried to get help in the first place.*"

LLMs are used to generate articles, reports, and weather summaries (P2, E4). While boosting efficiency, e.g., in automated newsletters (D7), this has also led to job displacement. E6 mentioned that "*we have made our own sort of ChatGPT which has resulted in some proofreaders which have been fired. This was not very popular.*" With article text generation, there are concerns about their alignment with editorial values. E.g., E13 stated that their organization was "*very hesitant with using the word genocide, regarding what's going on in Israel and Gaza and there's lots of critique. So sometimes, if you ask them [LLMs] to give a certain summary and it uses a word that's not the word that you would want to be responsible for. [...] we as an organisation are responsible for what the LLM writes at the end of the day and it's easier to hold people accountable than it is with a chatbot.*" E13 mentioned using Model Context Protocol (MCP) servers to improve the article position on services such as ChatGPT: "*so besides search engine optimization, we also have some sort of chatbot search optimization.*"

6 Discussion

Key insights & implications. Our interviews reveal a complex landscape in which personalization and explainability are unevenly adopted and difficult to implement. Three key insights stand out: (i) Contrary to much algorithmic research, personalization is neither universally desirable nor always feasible; it clashes with editorial control, user privacy, and journalistic missions. (ii) Interviews exposed substantial operational and infrastructure challenges, including limited team sizes (D7: "*four people … we don't have a specialist*"), and competing priorities that constrain what organizations can implement. This creates a significant gap between academic methods and what news organizations can practically deploy. (iii) When organizations invest in transparency, they mostly build internal dashboards rather than reader-facing explanations. Explainability tools thus mainly support internal debugging, echoing [8] and extending it by suggesting that explainability research should target multi-stakeholder communication, not only end-user transparency.

Relation to prior research. Our insights align with Cools [13]'s finding at the BBC that explainability remains aspirational rather than operational. Our cross-organizational perspective reveals that these challenges are *systemic*, not organization-specific. Further examining this from a cross-organizational perspective, we find that, overall, the differences between public and private organizations are not that large. There is a spectrum, with public organizations valuing data privacy more and some private organizations having more financial incentives for automating jobs. But both types of organization embed their values in similar ways and have similar views on explainability. Consistent with [39], we observe tensions among journalistic, market, and technical stakeholders, and extend it by showing how these tensions manifest in explainability decisions. Our findings also extend [8, 20], revealing an additional critical function in news: bridging engineering and editorial teams.

Guidelines. Returning to the wider question of **whether explainability methods could address the challenges that engineers face when building news recommender systems**, we believe they could and should play a role in communication with the newsroom and with monitoring bugs in a system (i.e., explainability-for-maintainers rather than explainability-for-end-users, per the scrutibility definition above). But there are organizational challenges that cannot be solved in this way, like access to better data, moving to new infrastructures, and a lack of financial resources and people. Based on our analysis, we recommend: (**i**) Start with internal explainability for editorial communication through dashboards showing recommendation behavior and diversity metrics (as demonstrated by P3, D5, P15), rather than complex end-user explanations, given their uncertain benefits (E4: "*only one user inquiry in a year*") and potential risks (E6: "*can be done badly*"). (**ii**) Adopt more systematic debugging approaches instead of score normalization, manually testing user profiles and embedding visualizations, instead of investing in more sophisticated methods. And (**iii**) Develop explainability requirements alongside business rules, creating traceability for when rules trigger.

Potential impact. Our work can have significant impact across research, industry, and policy. For researchers, it bridges the gap between explainability research and industrial practice by revealing what actually works in production environments and identifying real-world barriers that can guide future tool design. For practitioners, it enables organizations to learn from peers facing similar challenges, offers actionable guidelines that can reduce implementation costs, and helps teams make informed decisions about investing in explainability methods. Beyond immediate applications, this research can inform regulatory discussions around transparency requirements by grounding them in practical realities, influence industry best practices for responsible NRSs deployment, and affect millions of news consumers by improving how personalized systems are developed and made transparent.

Limitations. Our sample focuses on European and North American organizations under specific regulatory frameworks; findings may not generalize globally. We primarily captured technical perspectives; future work should incorporate editorial and reader views. Cross-stakeholder research on how engineers, editors, journalists, readers, and regulators perceive and value explainability could inform more nuanced design decisions.

7 Conclusion

We have conducted 15 semi-structured interviews with news engineers and product owners to investigate the extent to which personalization and explainability are currently developed in different news organizations. Through these interviews, we have found that personalization adoption is neither universal nor straightforward. Organizations make deliberate choices to limit or avoid it based on editorial values, privacy concerns, and resource constraints. Seven out of nine organizations use some personalization, but most employ hybrid approaches balancing algorithmic recommendation with editorial control. No organization explicitly uses algorithmic

explainability tools, despite interest from some teams. Where explainability exists, it focuses on internal stakeholders rather than end users, with external explanations viewed as unnecessary, risky, or low-priority.

For explainability to realize its potential in news recommendation and personalization, the RecSys community has both an opportunity and a responsibility to design methods that recognize constraints and provide lightweight, production-ready solutions that integrate into existing workflows. This work represents a step toward that goal, but much work remains to be done. A particularly important next step in this direction would be to expand the study to journalistic, market, editorial and end-user stakeholders.

Acknowledgments

We thank all participants for their time and valuable insights. We are also grateful to everyone who assisted with participant recruitment, and thank Hannes Cools, Maria Heuss and the AI, Media and Democracy Lab for their feedback. This research was (partially) supported by the Dutch Research Council (NWO), under project numbers 024.004.022, NWA.1389.20.183, and KICH3.LTP.20.006, and the European Union under grant agreement No. 101201510 (UNITE). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of their respective employers, funders and/or granting authorities.

References

[1] Juan Ahmad, Jonas Hellgren, and Alan Said. 2025. Tell Me the Good Stuff: User Preferences in Movie Recommendation Explanations. In *Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization*. 103–108.

[2] Nihal Alaqabawy, Aishwarya Satwani, Amy Voida, and Robin Burke. 2023. "It's just a robot that looks at numbers"': Restoring Journalistic Voice in News Recommendation. In *INRA@RecSys*.

[3] Alejandro Ariza-Casabona, Ludovico Boratto, and Maria Salamó. 2024. A Comparative Analysis of Text-based Explainable Recommender Systems. In *Proceedings of the 18th ACM Conference on Recommender Systems*. 105–115.

[4] Leif Azzopardi, Charles L.A. Clarke, Paul Kantor, Bhaskar Mitra, Johanne R Trippas, Zhaochun Ren, Mohammad Aliannejadi, Negar Arabzadeh, Raman Chandrasekar, Maarten de Rijke, et al. 2024. Report on the Search Futures Workshop at ECIR 2024. In *ACM SIGIR Forum*, Vol. 58. ACM New York, NY, USA, 1–41.

[5] Agathe Balayn, Mireia Yurrita, Fanny Rancourt, Fabio Casati, and Ujwal Gadiraju. 2025. Unpacking Trust Dynamics in the LLM Supply Chain: An Empirical Exploration to Foster Trustworthy LLM Production & Use. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1103, 20 pages. doi:10.1145/3706598.3713787

[6] Oren Barkan, Veronika Bogina, Liya Gurevitch, Yuval Asher, and Noam Koenigstein. 2024. A Counterfactual Framework for Learning and Evaluating Explanations for Recommender Systems. In *Proceedings of the ACM Web Conference 2024*. 3723–3733.

[7] Christine Bauer, Chandni Bagchi, Olusanmi A. Hundogan, and Karin van Es. 2024. Where Are the Values? A Systematic Literature Review on News Recommender Systems. *ACM Trans. Recomm. Syst.* 2, 3, Article 23 (June 2024), 40 pages. doi:10.1145/3654805

[8] Umang Bhatt, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, José M.F. Moura, and Peter Eckersley. 2020. Explainable Machine Learning in Deployment. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 648–657.

[9] Brett Binst, Lien Michiels, and Annelien Smets. 2025. What Is Serendipity? An Interview Study to Conceptualize Experienced Serendipity in Recommender Systems. In *Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization (UMAP '25)*. Association for Computing Machinery, New York, NY, USA, 243–252. doi:10.1145/3699682.3728325

[10] Michelle Brachman, Arielle Goldberg, Andrew Anderson, Stephanie Houde, Michael Muller, and Justin D Weisz. 2025. Towards Personalized and Contextualized Code Explanations. In *Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization*. 120–125.

[11] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[12] Andrea Brennen. 2020. What Do People Really Want When They Say They Want "Explainable AI?" We Asked 60 Stakeholders.. In *Extended abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–7.

[13] Hannes Cools. 2025. Navigating the Responsible AI Landscape: Unraveling the Principles-to-Practices Gap of Transparency and Explainability at the BBC. *Information, Communication & Society* (2025), 1–21.

[14] Nicholas Diakopoulos and Michael Koliska. 2017. Algorithmic Transparency in the News Media. *Digital Journalism* 5, 7 (2017), 809–828.

[15] 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.

[16] Sophia Egbert and Henning Bumann. 2026. Bridging Academia and Industry: A Collaborative Approach for Developing a Responsible News Recommender System. *ACM Transactions on Recommender Systems* (Feb. 2026). doi:10.1145/3797027 Just Accepted.

[17] Jingyue Gao, Xiting Wang, Yasha Wang, and Xing Xie. 2019. Explainable Recommendation through Attentive Multi-view Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 3622–3629.

[18] Andreas Grün and Xenija Neufeld. 2023. Transparently Serving the Public: Enhancing Public Service Media Values through Exploration. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 1045–1048.

[19] Greg Guest, Arwen Bunce, and Laura Johnson. 2006. How Many Interviews are Enough? An Experiment with Data Saturation and Variability. *Field methods* 18, 1 (2006), 59–82.

[20] Umm-e Habiba, Mohammad Kasra Habib, Justus Bogner, Jonas Fritzsch, and Stefan Wagner. 2025. How Do ML Practitioners Perceive Explainability? An Interview Study of Practices and Challenges. *Empirical Software Engineering* 30, 1 (2025), 18.

[21] 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.

[22] Natali Helberger. 2021. On the Democratic Role of News Recommenders. In *Algorithms, Automation, and News*. Routledge, 14–33.

[23] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining Collaborative Filtering Recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (Philadelphia, Pennsylvania, USA) *(CSCW '00)*. Association for Computing Machinery, New York, NY, USA, 241–250. doi:10.1145/358916.358995

[24] Karl Higley, Robin Burke, Michael D. Ekstrand, and Bart P. Knijnenburg. 2025. What News Recommendation Research Did (But Mostly Didn't) Teach Us About Building A News Recommender. *CEUR workshop proceedings* 4063.

[25] Oana Inel, Nicolas Mattis, Milda Norkute, Alessandro Piscopo, Timothée Schmude, Sanne Vrijenhoek, and Krisztian Balog. 2023. QUARE: 2nd Workshop on Measuring the Quality of Explanations in Recommender Systems. In *Proceedings of the 17th ACM Conference on Recommender Systems* (Singapore, Singapore) *(RecSys '23)*. Association for Computing Machinery, New York, NY, USA, 1241–1243. doi:10.1145/3604915.3608754

[26] Hao Jiang, Chuanzhen Li, Juanjuan Cai, and Jingling Wang. 2023. RCENR: A Reinforced and Contrastive Heterogeneous Network Reasoning Model for Explainable News Recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Taipei, Taiwan) *(SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 1710–1720. doi:10.1145/3539618.3591753

[27] Harmanpreet Kaur, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. 2020. Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. In *Proceedings of the 2020 CHI conference on Human Factors in Computing Systems*. 1–14.

[28] Johannes Kruse, Kasper Lindskow, Saikishore Kalloori, Marco Polignano, Claudio Pomo, Abhishek Srivastava, Anshuk Uppal, Michael Riis Andersen, and Jes Frellsen. 2024. RecSys Challenge 2024: Balancing Accuracy and Editorial Values in News Recommendations. In *Proceedings of the 18th ACM Conference on Recommender Systems* (Bari, Italy) *(RecSys '24)*. Association for Computing Machinery, New York, NY, USA, 1195–1199. doi:10.1145/3640457.3687164

[29] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* (1977), 159–174.

[30] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3313831.3376590

[31] Dairui Liu, Derek Greene, Irene Li, Xuefei Jiang, and Ruihai Dong. 2024. Topic-Centric Explanations for News Recommendation. *ACM Trans. Recomm. Syst.* 3, 2, Article 17 (Nov. 2024), 25 pages. doi:10.1145/3680295

[32] Feng Lu, Anca Dumitrache, and David Graus. 2020. Beyond Optimizing for Clicks: Incorporating Editorial Values in News Recommendation. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) *(UMAP '20)*. Association for Computing Machinery, New York, NY, USA, 145–153. doi:10.1145/3340631.3394864

[33] Ana Lucic, Sheeraz Ahmad, Amanda Furtado Brinhosa, Q. Vera Liao, Himani Agrawal, Umang Bhatt, Krishnaram Kenthapadi, Alice Xiang, Maarten de Rijke, and Nicholas Drabowski. 2022. Towards the Use of Saliency Maps for Explaining Low-Quality Electrocardiograms to End Users. In *Proceedings of the 39th International Conference on Machine Learning: Workshop on Interpretable Machine Learning in Healthcare (Proceedings of Machine Learning Research, Vol. 162)*. PMLR, Baltimore, Maryland, USA.

[34] Ana Lucic, Madhulika Srikumar, Umang Bhatt, Alice Xiang, Ankur Taly, Q Vera Liao, and Maarten de Rijke. 2021. A Multistakeholder Approach towards Evaluating AI Transparency Mechanisms. *arXiv preprint arXiv:2103.14976* (2021).

[35] Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems* 30 (2017).

[36] Siddharth Mehrotra, Ujwal Gadiraju, Eva Bittner, Folkert van Delden, Catholijn M. Jonker, and Myrthe L. Tielman. 2025. "Even explanations will not help in trusting [this] fundamentally biased system": A Predictive Policing Case-Study. In *Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization*. 51–62.

[37] Siddharth Mehrotra, Carolina Centeio Jorge, Catholijn M. Jonker, and Myrthe L. Tielman. 2024. Integrity-based Explanations for Fostering Appropriate Trust in AI Agents. *ACM Trans. Interact. Intell. Syst.* 14, 1, Article 4 (Jan. 2024), 36 pages. doi:10.1145/3610578

[38] Lien Michiels, Jens Leysen, Annelien Smets, and Bart Goethals. 2022. What are Filter Bubbles Really? A Review of the Conceptual and Empirical Work. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. 274–279.

[39] Eliza Mitova, Sina Blassnig, Edina Strikovic, Aleksandra Urman, Claes de Vreese, and Frank Esser. 2023. When Worlds Collide: Journalistic, Market, and Tech Logics in the Adoption of News Recommender Systems. *Journalism Studies* 24, 16 (2023), 1957–1976. doi:10.1080/1461670X.2023.2260504

[40] Lucas Möller and Sebastian Padó. 2025. Explaining Neural News Recommendation with Attributions onto Reading Histories. *ACM Trans. Intell. Syst. Technol.* 16, 1, Article 7 (2025), 25 pages. https://doi.org/10.1145/3673233

[41] Lynge Asbjørn Møller. 2024. Designing Algorithmic Editors: How Newspapers Embed and Encode Journalistic Values into News Recommender Systems. *Digital Journalism* 12, 7 (2024), 926–944.

[42] Caio Nóbrega and Leandro Marinho. 2019. Towards Explaining Recommendations through Local Surrogate Models. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*. 1671–1678.

[43] Maria Panteli, Alessandro Piscopo, Adam Harland, Jonathan Tutcher, and Felix Mercer Moss. 2019. Recommendation Systems for News Articles at the BBC.. In *Proceedings of the 7th International Workshop on News Recommendation and Analytics (INRA@RecSys 2019)*. 44–52.

[44] Alessandro Piscopo, Oana Inel, Sanne Vrijenhoek, Martijn Millecamp, and Krisztian Balog. 2023. Report on the 1st Workshop on Measuring the Quality of Explanations in Recommender Systems (QUARE 2022) at SIGIR 2022. In *ACM SIGIR Forum*, Vol. 56. ACM New York, NY, USA, 1–16.

[45] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why should I trust you?" Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1135–1144.

[46] Roan Schellingerhout, Francesco Barile, and Nava Tintarev. 2023. A Co-design Study for Multi-stakeholder Job Recommender System Explanations. In *World Conference on Explainable Artificial Intelligence*. Springer, 597–620.

[47] 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.

[48] Martin Spišák, Rodrigo Alves, Thomas Kelleher, Jason Sheppard, Ondřej Fiedler, Ergys Kosovrasti, Vojtěch Vančura, Petr Kasalický, and Pavel Kordík. 2025. Segment-Aware Analytics for Real-Time Editorial Support in Media Groups: Lessons from The Telegraph. In *Proceedings of the 13th International Workshop on News Recommendation and Analytics (INRA 2025) (CEUR Workshop Proceedings, Vol. 4056)*, Andreea Iana, Célina Treuillier, Vandana Yadav, Benjamin Kille, Andreas Lommatzsch, and Özlem Özgöbek (Eds.). CEUR-WS.org, Prague, Czech Republic. https://ceur-ws.org/Vol-4056/short1.pdf

[49] Tor Sporsem. 2024. Discovering Explainability Requirements in ML-Based Software. In *Proceedings of the 2024 IEEE/ACM 46th International Conference on Software Engineering: Companion Proceedings* (Lisbon, Portugal) *(ICSE-Companion '24)*. Association for Computing Machinery, New York, NY, USA, 170–172. doi:10.1145/3639478.3639807

[50] Anselm Strauss and Juliet Corbin. 1998. *Basics of Qualitative Research Techniques*. Thousand Oaks, CA: Sage Publications.

[51] Emily Sullivan, Dimitrios Bountouridis, Jaron Harambam, Shabnam Najafian, Felicia Loecherbach, Mykola Makhortykh, Domokos Kelen, Daricia Wilkinson, David Graus, and Nava Tintarev. 2019. Reading News with a Purpose: Explaining User Profiles for Self-Actualization. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization* (Larnaca, Cyprus) *(UMAP'19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 241–245. doi:10.1145/3314183.3323456

[52] Maxwell Szymanski, Vero Vanden Abeele, and Katrien Verbert. 2025. Disentangling Stakeholder Role and Expertise in User-Centered Explainable AI. In *Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization*. 32–39.

[53] Chang-You Tai, Liang-Ying Huang, Chien-Kun Huang, and Lun-Wei Ku. 2021. User-Centric Path Reasoning towards Explainable Recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, Canada) *(SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 879–889. doi:10.1145/3404835.3462847

[54] Maartje ter Hoeve, Mathieu Heruer, Daan Odijk, Anne Schuth, and Maarten de Rijke. 2017. Do News Consumers Want Explanations for Personalized News Rankings?. In *FATREC Workshop on Responsible Recommendation Proceedings*.

[55] Nava Tintarev and Judith Masthoff. 2015. Explaining Recommendations: Design and Evaluation. In *Recommender Systems Handbook*. Springer, 353–382.

[56] Khanh Hiep Tran, Azin Ghazimatin, and Rishiraj Saha Roy. 2021. Counterfactual Explanations for Neural Recommenders. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1627–1631.

[57] Thi Ngoc Trang Tran, Alexander Felfernig, Viet Man Le, Thi Minh Ngoc Chau, and Thu Giang Mai. 2023. User Needs for Explanations of Recommendations: In-depth Analyses of the Role of Item Domain and Personal Characteristics. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*. 54–65.

[58] Hanne Vandenbroucke and Annelien Smets. 2024. It's (not) all about that CTR: A Multi-Stakeholder Perspective on News Recommender Metrics. In *Proceedings of the 18th ACM Conference on Recommender Systems* (Bari, Italy) *(RecSys '24)*. Association for Computing Machinery, New York, NY, USA, 999–1003. doi:10.1145/3640457.3688183

[59] VERBI Software. 2024. *MAXQDA 2024*. Berlin, Germany. https://www.maxqda.com

[60] Sanne Vrijenhoek, Gabriel Bénédict, Mateo Gutierrez Granada, Daan Odijk, and Maarten de Rijke. 2022. RADio – Rank-Aware Divergence Metrics to Measure Normative Diversity in News Recommendations. In *Proceedings of the 16th ACM Conference on Recommender Systems* (Seattle, WA, USA) *(RecSys '22)*. Association for Computing Machinery, New York, NY, USA, 208–219. doi:10.1145/3523227.3546780

[61] Kathrin Wardatzky, Oana Inel, Luca Rossetto, and Abraham Bernstein. 2025. Whom do Explanations Serve? A Systematic Literature Survey of User Characteristics in Explainable Recommender Systems Evaluation. *ACM Transanctions on Recommender Systems* 3, 4, Article 49 (2025), 35 pages. doi:10.1145/3716394

[62] Cedric Waterschoot, Raciel Yera Toledo, Nava Tintarev, and Francesco Barile. 2025. With Friends Like These, Who Needs Explanations? Evaluating User Understanding of Group Recommendations. In *Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization*. 253–262.

[63] Chuhan Wu, Fangzhao Wu, Yongfeng Huang, and Xing Xie. 2023. Personalized News Recommendation: Methods and Challenges. *ACM Trans. Inf. Syst.* 41, 1, Article 24 (Jan. 2023), 50 pages. doi:10.1145/3530257

[64] Yongfeng Zhang and Xu Chen. 2020. Explainable Recommendation: A Survey and New Perspectives. *Found. Trends Inf. Retr.* 14, 1 (March 2020), 1–101. doi:10.1561/1500000066

[65] Jinfeng Zhong and Elsa Negre. 2022. Shap-enhanced Counterfactual Explanations for Recommendations. In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*. 1365–1372.