Welcome to the Metrics Jungle: Organizational Stakeholder Perspectives on Evaluation of News Recommender Systems in Industry

HANNE VANDENBROUCKE, imec-SMIT, Vrije Universiteit Brussel, Vrije Universiteit Brussel, Brussels, Belgium
LIEN MICHIELS, imec-SMIT, Vrije Universiteit Brussel, Vrije Universiteit Brussel, Brussels, Belgium and Adrem Data Lab, Universiteit Antwerpen, Antwerp, Belgium
ANNELIEN SMETS, imec-SMIT, Vrije Universiteit Brussel, Vrije Universiteit Brussel, Brussels, Belgium

News recommender systems (NRSs) are increasingly implemented in the news industry to enhance digital distribution, yet their development and evaluation are shaped by a complex interplay of organizational stakeholders with diverse objectives. This study investigates the real-world complexities of NRS implementation, considering the multitude of objectives on the NRS provider side by exploring how different stakeholder groups —from journalistic, commercial, and tech logic— conceptualize NRS goals and assess their performance. Through 22 semi-structured expert interviews across four news organizations in Europe, we uncover a "jungle of metrics" where an overabundance of data insights does not necessarily lead to better decision-making. While traditional accuracy-based metrics like click-through rate (CTR) remain widely used, they fail to capture the broader strategic objectives that guide NRS adoption in industry. This article identifies a critical gap between current academic evaluations of NRS and the business key performance indicators (KPIs) used in practice. Rather than evaluating NRS in isolation, stakeholders assess their contribution to overarching organizational goals, including reach, engagement, conversion, and retention. Our findings highlight the need for a more nuanced, multi-stakeholder approach to NRS evaluation that aligns with both business imperatives and journalistic values.

CCS Concepts: • **Information systems** → **Data analytics**;

Additional Key Words and Phrases: News recommender systems, evaluation metrics, multi-stakeholder, personalization

ACM Reference Format:
Hanne Vandenbroucke, Lien Michiels, and Annelien Smets. 2026. Welcome to the Metrics Jungle: Organizational Stakeholder Perspectives on Evaluation of News Recommender Systems in Industry. *ACM Trans. Recomm. Syst.* 4, 4, Article 54 (July 2026), 42 pages. https://doi.org/10.1145/3778173

This work was partially supported by the **Research Foundation Flanders** (**FWO**) under grant number S006323N.
Authors' Contact Information: Hanne Vandenbroucke (corresponding author), imec-SMIT, Vrije Universiteit Brussel, Vrije Universiteit Brussel, Brussels, Brussels, Belgium; e-mail: hanne.vandenbroucke@vub.be; Lien Michiels, imec-SMIT, Vrije Universiteit Brussel, Vrije Universiteit Brussel, Brussels, Brussels, Belgium and Adrem Data Lab, Universiteit Antwerpen, Antwerp, Flanders, Belgium; e-mail: lien.michiels@vub.be; Annelien Smets, imec-SMIT, Vrije Universiteit Brussel, Vrije Universiteit Brussel, Brussels, Belgium; e-mail: annelien.smets@vub.be.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM 2770-6699/2026/07-ART54
https://doi.org/10.1145/3778173

1 Introduction

News recommender systems (**NRSs**) are increasingly being implemented on news websites and mobile applications to achieve a wide range of purposes [8, 66, 71]. While the primary function of NRSs—matching users to relevant items—aligns with **recommender systems** (**RS**) in other domains, the complexity of the news ecosystem introduces additional challenges [17, 59, 69]. Beyond that core recommendation task, NRSs are intended to align with broader objectives such as upholding journalistic values [5, 6] and supporting business strategies [8, 20, 35]. This points to the various stakeholders who are involved in NRS development, each with their own perspective and set of objectives. For instance, the newsroom focuses on maintaining journalistic values and diversity of content, while commercial teams prioritize monetization and subscriber retention. Additionally, technologists must keep in mind the computational demands and requirements linked to the NRS development [33, 82].

The emergence of a multi-stakeholder approach in research has been instrumental to better understand the development of multi-objective (news) recommender system ((N)RS) designs [1, 32]. Despite the growing academic attention on NRSs, much of the research relies on easily implementable evaluation metrics [6, 70], such as **click-through rates** (**CTR**) [41, 57], page views [93], or dwell time [44]. These metrics, while convenient for evaluating immediate and observable user behavior, do not fully capture the diverse objectives that guide NRS development in news organizations in practice [55].

There has been an emerging call for research that addresses this disconnect between the evaluation metrics commonly studied in academic research and the broader strategic goals of news organizations in practice [52, 55]. This work contributes to this line of research by exploring how different stakeholders—guided by distinct institutional logics [51]—envision and assess the role of NRSs in contributing to those high-level organizational goals.

More specifically, this article addresses two gaps between the academic research and the real-world complexity of NRSs. First, much of the literature to date has focused on consumer-oriented metrics, such as recommendation diversity or user satisfaction [6], while metrics related to business performance have received comparatively less attention. To address this first research gap, our work focuses on the provider-side [34], where various internal organizational stakeholders are around the table to decide on the NRS design and implementation. Second, from this provider-side perspective, there is a need to develop more comprehensive and informative metrics that better align with the diverse NRS objectives as defined by organizational stakeholders. Although few works exist that have considered the business value of recommenders and point out metrics such as CTR and sales [34, 35], there remains a gap in identifying metrics that offer meaningful insights into real-world NRS impact on broader strategic goals and KPIs.

This work addresses these research gaps by investigating the evaluation metrics valued by different stakeholders involved in NRS decision-making, using insights from expert interviews. Drawing on interviews with 22 stakeholders across four profit-driven news organizations in Northwestern Europe, we examine how such metrics are interpreted, operationalized, and deemed relevant for achieving diverse organizational goals. In doing so, we contribute to the existing body of research, responding to the scarcity of interview-based studies noted in the systematic review of NRSs by Bauer et al. [6].

This article extends our previous work, *It's (Not) All About That CTR: A Multi-Stakeholder Perspective on News Recommender Metrics* [83], by two significant additions. First, we elaborate more on our methodological approach that connects social science literature on institutional logics and organizational stakeholder dynamics with existing literature on RS. This interdisciplinary approach allows us to bridge the gap between technical optimization and the broader

organizational considerations that shape NRS development through qualitative research. Second, this article presents findings from 11 additional interviews, expanding our dataset to include insights from two Scandinavian news organizations. While our previous study covered two Western European news organizations (operating in Flanders and the Netherlands), this expanded dataset also includes two news organizations from the Northern European media market (operating in Scandinavia), substantially expanding the scope of our research. The contribution of this article lies in helping to understand the differing perspectives on NRS objectives and related metrics in light of the inherent institutional logic of various organizational stakeholders involved.

The remainder of this article is structured as follows: First, we provide an overview of related works on multi-objective NRS design and evaluation metrics. Next, we outline the conceptual framework in Section 2.4, which informed the methodological setup of this research. In Section 3, we detail the news media organizations studied, our set of respondents, the interview procedure and the analytical approach. This is followed by a comprehensive overview of our findings in Section 4. Section 5 includes a discussion of the implications of our results. This is followed by section 7, in which we outline the limitations of this work and provide suggestions for further research.

2 Background

This background section situates our study in the current line of multi-stakeholder research on NRS. First, we review the distinct stakeholder logics that inform the design of multi-objective NRSs in real-world settings. Next, we examine how different perspectives on NRS purposes are translated into evaluation metrics, drawing on research from both computational and media studies. Finally, we clarify our approach to NRSs within the broader digital strategy, which allows to comprehend the challenges linked to the adoption of innovative technology within news organizations.

2.1 Multi-objective NRS Design and Stakeholder Logics

RSs are useful tools to provide accurate item predictions tailored to each user. However, there is growing awareness that optimizing only for accuracy is often insufficient in real-world applications. Instead, a multitude of—often competing—objectives must be pursued, reflecting the diverse goals of stakeholders involved in the design and use of RS [32, 34]. This article adds to the growing body of research investigating the complexities of multi-objective (N)RS in real-world settings from a multi-stakeholder perspective. Foundational work by Jannach and Adomavicius [33] distinguishes between users and providers as two distinct stakeholder groups. Next, Jannach and Bauer [34] emphasize that even within one stakeholder group, multiple goals coexist. Throughout this article, we discuss various types of aims, and to ensure clarity and consistency, we outline our conceptualization, inspired by Jannach and Adomavicius [33] and Smets et al. [71]: "Goals" denote high-level strategic aims, while "objectives" refer to specific **key performance indicators** (**KPIs**) that operationalize those organizational goals for particular teams. Those objectives are pursued through various actions, among which is NRS implementation. The specific, concrete uses or tasks for which a NRS is developed and optimized are referred to as NRS "purposes" in this work.

Given the focus of our study on the complexity on the NRS provider side, we draw upon Smets et al. [71], who adapted the multi-stakeholder framework to the news industry. Their interview-based study demonstrates that NRS design emerges from an ongoing negotiation among editors, business executives, and software developers, and so on. [71]. Other empirical studies that conducted qualitative, interview-based research in Western European news media organizations [7, 8, 51, 82] found similar negotiation dynamics between different actors within news companies. Building upon these works, we group the different actors involved in NRS decision-making into three distinct stakeholder groups: newsroom, commercial department, and technological team.

Each stakeholder group brings its own perspective to the table. We argue that each stakeholder group operates according to a distinct institutional logic [46], which entails a set of values, goals, and assumptions shaped by professional identities and organizational subcultures [42, 51].

Drawing on the work of Mitova et al. [51], we distinguish three stakeholder logics: First, the *journalistic logic*, including both journalists and editors, which emphasizes editorial goals such as maintaining news quality and journalistic integrity. Concretely, editors and journalists may want to ensure that recommendations align with journalistic values, such as maintaining news quality, promoting diversity, and upholding the democratic mission of journalism [18, 28, 90], as the design and operation of NRS may have a significant impact on users' news consumption. Scholars have argued that NRSs hold significant potential to positively influence reading behavior by promoting diversity within the news catalog, unlocking the "long tail" of content rather than reinforcing filter bubbles and routine behaviors [28, 90]. On the contrary, the adoption of NRSs has been linked to several negative outcomes, such as filter bubbles [49], algorithmic bias [6], lack of transparency [79], and over-optimization for engagement leading to clickbait or poor quality of journalism [81].

From a business point of view, NRSs have proven to be valuable in increasing revenue streams for news organizations [47, 64]. This is related to the second stakeholder logic, the *commercial logic*, which prioritizes business-oriented objectives such as maximizing ad revenue and boosting subscriptions. News organizations rely on advertising revenues and/or subscription sales, often employing a combined or hybrid business model [2]. In this study, the news organizations in our sample rely on both advertisements and subscriptions. Therefore, the commercial departments want NRS design to maximize profitability, contributing either positively to advertising revenues or increasing subscription sales.

As pointed out by Smets et al. [71], editors play a dual role in the NRS development, bridging between both logics. On the one hand, they safeguard journalistic values [5], ensuring that recommended content aligns with editorial standards and mitigating undesired outcomes of tailored news recommendations. On the other hand, editors are embedded within the broader business strategy of news organizations, requiring them to align NRS decision-making with strategic considerations [71].

The third logic is the closest linked to the NRS development process, the so-called *tech logic* focuses on the technical implementation and optimization of the algorithm, including concerns about scalability and operational functionalities [7, 71]. Nonetheless, the tech logic extends beyond NRS development and encompasses all who deal with the broader digital infrastructure, including data architecture, analytics pipelines, development and maintenance of online tools, and the management of hardware and software systems.

These three logics shape the sets of values, perceptions, and objectives of each stakeholder group. Accordingly, each stakeholder group articulates different expectations for how the NRS should support organizational strategic goals, leading to divergent interpretations of the specific purposes for which the NRS should be designed. This leads to our first research question: **RQ1: What are the envisioned purposes for NRSs from each stakeholder's perspective?**

Our interviews aim at understanding the interplay of these various interests and how they translate into concrete objectives to which a RS should contribute according to different stakeholders, while taking into account the needed computational considerations to ensure a technically feasible system [71, 82]. The recent studies cited throughout this section examine the NRS development process from a multi-stakeholder approach. However, they only briefly address the link between differing, competing goals and evaluation criteria. For instance, Smets et al. [71] focused on the preconditions for translating objectives into metrics, without elaborating on the translation process. In Mitova et al. [51] and Blassnig et al. [7], audience metrics cited by respondents are included throughout the reporting of findings, but they are not the focus of the study. Lastly, van Es and

Nguyen [82] point to a paradoxical attitude among stakeholders toward existing metrics as too limited, while simultaneously being very interesting. Their work, however, does not explore which metrics are tied to which specific KPIs or objectives, nor how these relate to high-level strategic goals. The contribution of this article lies in the focus on the link between objectives and metrics directly connects to a broader body of work on metrics for NRS development, which we review in the next section.

2.2 Metrics for NRS Development

Building on the understanding of the different envisioned NRS purposes and their role in supporting high-level strategic objectives, we now turn to how these purposes are operationalized into concrete metrics that guide system development evaluation in practice. To do so, our literature review draws on two complementary strands of research. The first comes from the computational RS field, where system-level evaluation dominates. The second, reflecting this article's focus on real-world settings, comes from media studies, which situates NRS evaluation within the broader strategic and organizational objectives of news publishers.

2.2.1 NRS Metrics in Computational Research. There is a large body of computational research on RS evaluation. Most academic works in this domain seek to improve functionalities using either offline or online evaluation methods. For a complete overview of dominant metrics employed, we refer to Karimi et al. [38], Silveira et al. [70], Fayyaz et al. [20], and Tamm et al. [75].

A study of these works points to the frequent recurrence of typical accuracy-centric metrics such as precision, recall, nDGC, MRR and CTR in research. In real-world NRS settings, online experimentation using robust A/B testing, remains critical. In this context, Jeunen et al. [36] provide a valuable overview of metrics for A/B testing, offering practical tools for online experimentation. However, their work does not delve into how these metrics align with broader organizational objectives. In recent years, there has been growing awareness that an exclusive focus on providing accurate item predictions for individual users is insufficient. Beyond-accuracy measures have gained interest. Specifically in the news context, values such as diversity, novelty, serendipity and fairness have received more academic attention, see, for example, the work of Bauer et al. [6]. Their literature review reflects the emergence of a normative approach in RS research. This normative lens stems from increasing recognition that traditional accuracy metrics do not adequately measure the impact of personalized recommendations on relevant stakeholders. This results in a growing body of studies that focus on the norms and values underlying (N)RS design [72]. Similarly, the work of Vrijenhoek et al. [88, 89] conceptualizes the value of diversity within European news media systems, translating different dimensions of diversity into concrete metrics. However, to the best of our knowledge, no prior work has investigated whether these metrics used in academic works are also adopted in a real-world NRS setting. Therefore, rather than reiterating accuracy (e.g., recall and precision) and beyond- accuracy (e.g., diversity, novelty) metrics that have already been discussed in RS literature (e.g., [6, 20, 70, 89]), this article investigates which evaluation metrics are used in practice to measure the multiple objectives of NRSs. Despite the large body of work on system-level NRS evaluation, little is known about how existing metrics relate to actual NRS evaluation practices in the industry. To this end, we examine recent works on the evaluation of multi-objective NRS within news organizations from media studies.

2.2.2 NRS Metrics in Media Studies. As outlined in our earlier work [83], organizational stakeholders refer to RS as useful tools to achieve goals along different stages of the user journey, such as increasing engagement by enhancing audience experiences and contributing to business outcomes through maximizing conversion. As customary in marketing discourse, different stages of the user journey are conceptualized as a funnel. Unlike a linear pipeline, the funnel represents the gradual

filtering of users, where the broad base of potential readers becomes smaller as users pass through the stages of acquisition, engagement, conversion and retention [27, 67]. Our first set of interviews already showed that all stakeholders, regardless of logic, discuss the NRS goals in line with the four stages of the reach-engagement-conversion-retention funnel [83]. This funnel serves as a framework of overarching strategic goals in which specific KPIs are positioned. Below, we provide an overview of NRS metrics discussed in recent studies, structured according to that funnel.

Acquisition: Reach and Visibility. The acquisition stage refers to attracting a wide audience, which is crucial for both editorial (informing a wide audience [22]) and commercial purposes, as it establishes the pool of potential users who may later engage, subscribe, and remain loyal users [27, 58]. The goal of increasing reach does not only indicate audience size but also increasing ad visibility, which serves as a critical driver for advertising revenue [2]. As news organizations are increasingly dependent on traffic coming from social media or search engines [56, 76], integrating NRS algorithms into push notifications and e-mail newsletters promises to be valuable for driving audiences to the news platforms and thus, increase reach [29, 53].

Engagement: Attention to and Involvement with News. Engagement represents the next stage in the funnel. This strategic goal is a multidimensional concept [73], that encompasses the quality and depth of interactions users have with news content, blending attention and involvement as its core components [4, 14, 93]. Drawing upon existing work, we conceptualize engagement as a twofold construct involving, on the one hand, attention, as a standalone KPI, and on the other hand, involvement, which includes emotional, cognitive, and behavioral dimensions [4]. Attention, or how long users spend on news content, is measured by metrics such as *dwell time, viewport time, attention time*, or *time spent* [21, 26, 44]. These metrics capture how effectively NRSs can sustain audience focus on specific articles or topics. Involvement, meanwhile, is categorized into emotional, cognitive, and behavioral components [4]. Emotional involvement, although challenging to measure, can be assessed using sentiment analysis to evaluate readers' emotional responses [3, 30]. Cognitive involvement is more tangible, reflected in metrics like *scroll depth, CTR*, and *page views*, which indicate the user's interaction with content [10, 26, 41]. Behavioral involvement is tracked through actions such as likes, comments, and shares, offering insight into how users actively participate in the dissemination of news [4, 43]. Engagement is not only a goal in itself but also an important predictor of further action. Non-transactional outcomes, such as prolonged attention and sharing behavior, often correlate with transactional outcomes like subscriptions. Research by Davoudi et al. [15] has demonstrated a strong link between engagement metrics and subscription likelihood, highlighting the role of positive user experiences in driving conversions.

Conversion: Log-ins and Subscriptions. The third stage of the funnel is conversion, which refers to the transition from anonymous news consumption to creating an account or buying a subscription. This study focuses on conversion related to users' access to premium content as the news organizations in our sample prioritize the monetization of their audiences through subscriptions over ad revenues. To this end, there are different paywall strategies. A strict or hard paywall does not allow any news content to be consumed by non-subscribers. Alternatively, metered paywalls allow users to read a limited number of free articles before requiring a subscription; a data wall requires users to create an account in exchange for continued access, enabling the collection of personal data for personalized recommendations and targeted subscription offers [19]. Recently, a "Pay-or-OK" model has emerged, giving users the choice between paying for a subscription or consenting to data collection and personalized advertising.

Throughout this work and in the interviews, the term "conversion" is used primarily to refer to both log-in or subscribe. However, the interpretation of "conversion" might differ slightly

depending on the specific revenue model of the newspapers in our sample. Even within the same media organization, not every newspaper has the same revenue model and/or paywall strategy. To contribute to the strategic goal of conversion, NRSs are good at surfacing high-value content that encourages readers to access premium content and consider subscribing. However, a crucial factor affecting conversion is users' willingness to pay for online news. In an era of abundant free content, convincing users to pay remains challenging for many organizations [23]. Previous work has shown that willingness to pay is dependent on quality perception and reading behavior [9]. Therefore, if the NRSs are optimized to promote premium content recommendations based on user preferences, the likelihood of conversion may increase. Similarly, metrics related to the probability of conversion often correlate with engagement metrics such as *page views* and *dwell time*, because spending a lot of time on the platform can indicate a user is willing to commit to the news brand [21]. However, the most informative metric—closely tied to the sales KPI—is *paywall hits*, as it reflects the number of times a user "hits" the paywall, which serves as a proxy or indicator for their interest in premium content [2, 83].

Retention: Building Loyalty. The final stage of the funnel is retention, which focuses on maintaining a loyal base of paying, highly engaged subscribers. Retention is a long-term goal that reflects the organization's ability to provide consistent value to its users. The primary metric for retention used in previous studies is the *percentage of churn*, which measures the number of subscription terminations [62]. This metric can then be placed alongside the *number of subscription renewals*. High retention rates are crucial for maintaining revenues and ensuring a stable audience [93].

NRSs can contribute to retention by encouraging a tailored news consumption journey that keeps users engaged [82]. Recent studies have investigated clickstream data in relation to engagement and retention [39, 93]. However, there remains uncertainty on what (subscription or user behavioral) data and metrics should be analyzed for NRSs to deliver content that reinforces subscription renewals. The existing body of recommender systems literature predominantly focuses on short-term engagement metrics as CTR, number of visits, or views on a page, providing immediate feedback on user interaction [34, 35, 62, 83]. Nevertheless, the short-term performance indicators do not necessarily align with long-term organizational objectives, such as fostering loyal readership and increasing subscription retention.

2.2.3 Bridging the Computational and Media Perspective on NRS Evaluation. The overview of evaluation metrics linked to the different funnel stages reveals a lack of alignment between NRS metrics used in research and the actual (long-term) objectives of news organizations [35]. Recently, research on evaluation practices in (N)RS has been criticized for their reliance on simplistic and inconsistent metrics that fail to reflect the complexity of the domain in which they operate [6, 34, 38, 70].

There is a lack of standardized definitions for performance metrics. Tamm et al. [75] highlight inconsistencies in the calculation and interpretation of widely used metrics in academic research, emphasizing the need for clearer definitions to ensure comparability and reliability. While Fayyaz et al. [20] provide a comprehensive overview of recommender metrics, their focus on general applications neglects the unique challenges of NRSs.

Current NRS research often fails to employ nuanced evaluation measures tailored to the news context, such as metrics assessing diversity and serendipity, which are crucial for news recommendations but remain underexplored [37]. The work of Bauer et al. [6] also demonstrates this, adding the importance of aligning journalistic values with evaluation frameworks through interview-based research. The tension between optimizing for immediate, short-term engagement outcomes versus sustaining long-term value underscores the need to explore a wider range of metrics that

account for the diverse priorities of stakeholders and understanding the rationale behind these metrics, grounded in different institutional logics.

The second research question that this article aims at answering is: **RQ2: What metrics (whether currently implemented or not) are being considered in the industry to evaluate NRSs?**

2.3 Evaluating NRSs in the Context of Organizational Transformation to a Digital-first Strategy

Our analysis of the recent academic works suggests that NRSs are currently often evaluated as a stand-alone system in research. However, in real-world contexts, NRSs are not designed in isolation. Exploring the possibilities of personalization through NRS is only one aspect of the broader shift from traditional newspaper publishing to online news distribution. In order to cope with commercial pressures, legacy news organizations are adopting new technologies and rethinking the role of innovative tools within newsrooms [8, 92]. This, in turn, prompts questions about how these tools integrate into existing organizational dynamics [50].

Therefore, we argue that studying NRS implementation should not only focus on the technological aspects, but also consider the context of the entire organization being in a state of flux. The multi-stakeholder approach in our study allows to understand the stakeholder dynamics and, consequently, to analyze the differing perspectives on NRS [85]. We adopt the understanding that the decision-making process surrounding NRSs in industry typically involves three stakeholder groups: editors, commercial teams and technical developers, as established by earlier empirical studies [7, 8, 51, 82]. Smets et al. [71] highlight the role of a **product owner** (**PO**) who mediates the negotiation and acts as a bridging actor between the different stakeholder interests. Prior research has also demonstrated that such intermediaries are crucial for resolving tensions that may arise between editorial priorities and commercial objectives [7, 51, 82]. However, these stakeholder groups are not necessarily separate or in conflict; in fact, there is evidence that they are moving closer together, and increased collaboration fosters a shared "doxa" for technology adoption [8]. After all, stakeholders' attitude toward technologies, and more specifically their perceptions and expectations of NRSs, are dynamic. The three stakeholder logics described above in 2.1 are not fixed; rather, they continuously influence each other and are subject to changes, for example, in response to technological innovations. Recent works have highlighted how digital transformation of the news ecosystem reshapes the revenue models of legacy media, which calls for revisiting the commercial goals [2, 8]. Similarly, news production and distribution practices have shifted toward formats suited to online platforms, which in turn alters the professional standards and objectives of the newsroom [5, 45]. In other words, digitization is paired with an ongoing need to defend and re-define the professional and normative boundaries of their field [12, 46]. In the specific context of NRSs, Lischka [45] suggests that journalistic and commercial logic are blending together as both parties increasingly focus on audience metrics. Meanwhile, Svensson [74] argues for the rise of a new, hybrid "personalization logic" that combines commercial and technical rationalities. However, the implications of such logic convergence [45] and the potential emergence of new logics [74] remain underexplored to date. Our study does not seek to detail the current stakeholder logics or uncover new logics, but rather examines how the three stakeholder logics identified above influence the development and evaluation of NRSs.

Our interviews probe the envisioned NRS purposes and related metrics valued by each stakeholder logic. However, as highlighted in this section, NRS implementation is not merely a technical decision; rather, this is embedded within the broader strategic decision-making. To fully grasp how NRS development unfolds in real-world news organizations, our work studies NRS purposes and metrics within the context of organizational change linked to the ongoing shift toward a

Table 1. Conceptual Framework

Stakeholder group	Logic	Overarching, strategic goals	Objectives and KPIs	NRS purposes	NRS metrics
Newsroom	Journalistic Commercial				
Commercial departments	Commercial				
Technological team	Tech				

digital-first strategy [50, 85]. Therefore, our third research question explores the organizational struggles associated with this transition to a digital (distribution) strategy of which the NRS is a part. **RQ3: What difficulties arise for NRS metrics within the broader digital strategy in a multi-stakeholder environment?**

2.4 Conceptual Framework

The previous Sections have illustrated that NRS design is shaped by three organizational stakeholder groups—newsroom, commercial departments, and technological team—operating under distinct institutional logics. Accordingly, each stakeholder group articulates different expectations for how the NRSs should support broader organizational goals, leading to divergent interpretations of the system's role within the overall strategy. These differing visions regarding specific NRS purposes are, in turn, operationalized through prioritizing distinct metrics, which stakeholders perceive as meaningful indicators for assessing whether KPIs are being achieved with the support of the NRSs. This rationale presents the conceptual framework of this work. In Table 1, we visualize this conceptual framework, which informed our research design. The framework captures what the envisioned goals for the NRSs are from each stakeholder logic (RQ1: green), what metrics are being considered (RQ2: orange), and what difficulties arise for NRS metrics within the broader digital strategy (RQ3: purple). By interviewing organizational stakeholders, we address the three research questions and demonstrate how the conceptual framework advances our understanding real-world NRS practices.

3 Methodology

3.1 Data Collection

We employ semi-structured expert interviews, a method that ensures consistency across interviews while allowing for the in-depth exploration of emerging themes. Expert interviews are particularly well-suited for studying complex environments with multiple stakeholders as they elicit knowledge about the underlying dynamics shaping the development of media innovation [80]. As the aim of our study is to investigate how different organizational stakeholders understand and evaluate the role of NRSs, we interview key actors who possess first-hand experience with the NRS development, as they are (to some extent) involved in the decision-making.

3.1.1 News Organizations in Our Sample. Our sample covers four profit-oriented large-scale news organizations, each of which distributes multiple news brands across multiple regions. Two of the commercial companies operate in Flanders and the Netherlands, while the other two are active in Scandinavia (Sweden, Denmark, and Norway). These regions are comparable since both are characterized by a high concentration of media ownership with few players dominating the news landscape [11, 16, 31, 63]. All four cases were selected for their prominent market position

as dominant private players with a large market share. Nevertheless, they belong to two distinct media systems: the Western media system (Belgium and the Netherlands in our sample) and the Nordic media system (Scandinavia in our sample) and differ slightly in terms of journalistic professionalism and political parallelism; for more details, we refer to Brüggemann et al. [11], Humprecht et al. [31], Picard and Russi [63]. These different market factors do not invalidate our findings regarding the organizational stakeholder perspectives on NRSs. Furthermore, all four organizations share similar NRS adoption trajectories. Each company embarked on NRS integration between 2018 and 2019. While some initially outsourced development to third-party companies, by 2022, all had established in-house NRS design. Together, the four organizations thus provide a valuable representation of the current state of NRS development in the Northwestern European commercial news industry.

3.1.2 Sampling Approach. We relied on a combination of purposive sampling [60], based on the participants' belonging to a specific logic, and snowball sampling [61]. Concretely, we started with the PO or product manager of the NRS in every organization. They would then refer the interviewer to other internal stakeholders with whom they collaborate in the decision-making process surrounding the NRS, but who are considered to be part of another team or represent another professional logic. We continued conducting interviews at the same news media organization until each interviewee indicated that there were no other relevant colleagues to talk to regarding the purpose of our research, and when no new perspectives emerged in the interviews, indicating saturation.

3.1.5 Respondents. In total, we conducted 22 interviews across all four organizations, and in each organization, we questioned every logic. We interviewed 6 professionals representing the journalistic logic, 5 for the commercial logic, and 11 respondents aligned with the tech logic. A considerable proportion of our sample (11 out of 22) comprises respondents from the tech logic, yet these respondents have different roles within the organization, thus providing valuable insights. When contacting professionals from the journalistic or commercial logic for an interview about NRSs, they usually refer us to the PO NRS or another tech profile within the organization that deals with NRS development or data analysis. Nevertheless, the direct referrals from the POs have ensured that we still spoke to the relevant stakeholders to study their perspective.

Our sample largely consists of profiles with leadership over their team, such as the **product owner of the news recommender system** (**PO NRS**), so-called "lead" roles in the technical and commercial teams, and editors-in-chief. These people are considered qualified experts within the organization to discuss how NRS development proceeds within the broader digital strategy.

A detailed overview of all participants, their professional logic, an anonymized description of their role within their organization, and the region in which each news organization operates can be found in Table 2.

Compared to our short article [83], this work includes 11 additional interviews, which are marked with an asterisk (*) in Table 2. The interviews from the short article were conducted between November 2023 and March 2024, and all took place in the two news organizations active in **Belgium and the Netherlands** (**BE/NL** in Table). The new series of interviews was conducted between December 2024 and March 2025. These consist of two new interviews with a chief editor and a deputy chief editor, one from each news organization in BE/NL, aiming at gaining deeper insight into how NRS evaluation relates to editorial considerations and operations. The remaining 9 new interviews were conducted at the two Scandinavian commercial media organizations (SCAN in Table 2).

Throughout the findings (Section 4), respondent quotes are labeled with an ID, where the first letter indicates their logic (T = tech, C = commercial, J = journalistic).

Table 2. Overview of Respondents, Function, and Logic

Organization	Function	Logic	ID
BE/NL 1	Product owner NRS	Tech	T1
BE/NL 2	Product owner NRS	Tech	T2
BE/NL 1	Product owner UI&UX	Tech	T3
BE/NL 2	Lead data analytics	Tech	T4
BE/NL 2	Lead digital data	Tech	T5
BE/NL 2	Lead digital traffic	Tech	T6
BE/NL 2	Product director	Commercial	C1
BE/NL 1	Lead market (B2C)	Commercial	C2
BE/NL 2	Lead acquisition and conversion	Commercial	C3
BE/NL 2	Lead content performance	Journalistic	J1
BE/NL 1	Product owner news brand	Journalistic	J2
BE/NL 1	Deputy chief editor	Journalistic	J3*
BE/NL 2	Head chief editor	Journalistic	J4*
SCAN 1	Head of AI	Tech	T7*
SCAN 1	Lead NRS developer	Tech	T8*
SCAN 2	Product owner NRS	Tech	T9*
SCAN 2	Head of AI	Tech	T10*
SCAN 2	Lead NRS developer	Tech	T11*
SCAN 1	Data analyst market (B2C)	Commercial	C4*
SCAN 2	Data analyst market (B2C)	Commercial	C5*
SCAN 1	Deputy chief editor	Journalistic	J5*
SCAN 2	Deputy chief editor	Journalistic	J6*

3.2 Interview Protocol

3.2.1 Topic Guide. When conducting expert interviews, it is crucial that the interviewer has sufficient knowledge of the field in order to be perceived as a competent interlocutor by the respondents and obtain detailed, elaborate answers [80]. Given that our respondents perform such different roles, have unique perspectives on NRSs coming from a different logic, our topic guide should also be tailored to each of the respondents. Therefore, we performed a comprehensive literature review that includes the technical background of RS, journalistic values, and commercial strategic interests. The conceptual framework, as discussed in Section 2.4, provides an outline for the topic guide that should be seen as the semi-structured questionnaire for our interviews [24].

The full topic guide consists of two main sections. The first, more general section, covers questions about the perceived role and impact of NRSs within the organization, aimed at uncovering the complexities of NRS development within the organizational, multi-stakeholder context, starting with questions that prompted respondents to provide an overview or timeline of the NRS development process so far. This allows us to understand the process and different NRS objectives mentioned by the interviewees, as well as identify the different stakeholders who are involved in this process, using this as a foundation for deeper exploration during the interview. It is inherent to expert interviews to work with open-ended questions to leave room for the respondents to formulate comprehensive answers and thus also allow the interview to be differently structured than systematic interviews. Depending on the interviewee's answers, questions can be rearranged, additional questions are added, or certain questions should not be asked (if respondents have already answered them, or seem irrelevant) [80]. The second part of our topic guide is "logic-specific" and gets adapted depending on the stakeholder group the interviewee belongs to. The questions

in this second part dive deeper into the distinct priorities of each stakeholder, examining how their logics, professional identities, and backgrounds influence their perspectives on NRSs. This section includes both open-ended and theoretically-driven questions to fully cover all aspects of the phenomenon under study [24]. The complete topic guide, structured following our conceptual framework, is provided in Appendix A.

3.2.2 Interview Process. The interviews were conducted either in person or online, and all participants were informed about the study's purpose and data-handling practices. For each interview, we obtained informed consent to process responses and report their answers using quotes. However, we do not have permission to publicly share the full transcripts as they contain details about specific strategic decisions and forthcoming projects related to NRSs and personalization. The average duration of the interviews was 1 hour and 03 minutes. The interviews were conducted in Dutch or English, and the Dutch quotes used in this article have been translated into English by the researchers.

3.3 Data Analysis

The interview recordings were transcribed using Scribewave and analyzed using MAXQDA24. The analysis followed an "abductive coding" approach as described by Tavory and Timmermans [77]. Abductive coding combines both inductive and deductive coding in an iterative process. Inductive coding allows themes to emerge from the data without predefined categories, while deductive coding applies existing theoretical frameworks. Our analysis begins with a deductive phase in which predetermined codes based on our conceptual framework are retrieved from the transcripts to verify which aspect of the theory surfaces in industry. This step ensures that key theoretical concepts are systematically examined in the data, allowing us to assess how well existing frameworks align with real-world NRS practices and where gaps may exist. The topic guide in Appendix A showcases the theoretical concepts used for deductive coding of the transcripts. Next, when going through the transcripts for subsequent times, inductive coding is used to "find" new codes that do not match the predetermined codes extracted from the literature. Inductive coding allows for further exploration of these gaps and "loose ends" identified in the deductive phase. This process begins with open coding, where we assign labels (or codes) to text fragments based on the specific goals, KPIs, or metrics mentioned by the respondents. In contrast to deductive coding, this step allows us to capture nuances or even specific phrasings that allow to identify the diverse stakeholder perspectives under study. Next, axial coding is applied to systematically compare and refine the initial codes, allowing us to identify patterns and relationships across the data. Through an iterative process of reanalyzing transcripts, merging related codes, and structuring them into broader categories, we develop a final overarching coding diagram that includes both pre-made theoretical codes (deductive) and newly emerged codes (inductive) [13]. The full coding diagram is included in Appendix B.

This abductive approach enables us to bridge theory and practice by examining how established frameworks apply in a real-world industry setting while also uncovering new insights from the interview data. Hereby, we respond to the growing call in research to narrow the gap between academic studies and industry practices [6, 52], ensuring that our findings are both theoretically grounded and practically relevant.

4 Findings

This section presents the core findings of our study, structured around our three research questions RQs. Our first RQ examines the envisioned purposes for the NRS from the different stakeholders across editorial, commercial, and technical teams. In 4.1, we report the expectations and

envisioned purposes they articulate and analyze how that reflects their stakeholder logic. Our second RQ investigates which metrics (currently used or not) are considered meaningful to assess the performance of the NRS. In 4.2, we report the metrics structured according to the funnel stages: reach, engagement, conversion, and retention. Finally, our third RQ looks at the NRS within news organizations' broader digital strategy. Section 4.3 highlights the main difficulties stakeholders experience when aligning high-level strategic goals with operational metrics. We report interview findings using excerpts of quotes to ensure compact reporting that includes as many of our respondents' voices as possible. The full quotes are included as auxiliary material to convey more of the full transcripts as additional context to the results we report below.

4.1 Stakeholder Perspectives on NRS Purposes

4.1.1 Journalistic Logic. The stakeholders representing the journalistic logic in our sample are (deputy) chief editors, a news brand PO and content performance lead, who describe the role of NRSs as to *"engage readers more"* (J3) *"while maintaining editorial values such as journalistic integrity"* (J3&J4), *"independence"* (J4), and *"diversity"* (J1). Their primary envisioned purpose of the NRS is to *"excite readers with interesting recommendations that reflect the mix of content"* (J4) that each news brand brings. The editors assume that by helping readers explore both the breadth and depth of the reporting, NRSs should aim to *"encourage reading more during a session"* (J1) while ensuring the integrity of journalistic curation:

> *"There is news that no one should miss, that should be placed on the front page by the editor, but for other things like sports and entertainment, personalization is very good"* (J3)

This indicates that human curation and editorial oversight remain in place for so-called major news to ensure high journalistic standards. Algorithmic curation is used for what is considered less important news. The interviews show that NRSs should primarily help users discover articles that match their interests and thus encourage engagement and interaction with the articles. This points to similar dynamics as identified by Smets et al. [71], where editors not only follow a journalistic logic, but also consider NRSs from a strategic business perspective. They consider algorithmic curation and recommendations as a part of "personalization", which serves as a tool to excite audiences about the content on the news website or app. In this light, editors refer to NRSs to help push readers deeper into the funnel *"to make sure that as many people as possible are logged in"* (J3). This reflects the strategic side of the dual role of editors: in addition to journalistic considerations, they also have commercial objectives in mind (as discussed in [71]). There is a growing awareness among editors that data can provide insights for improving content performance, and *"users need to be logged in to get that data"* (J1). Our discussion (section 5) delves deeper into the value of logged-in users in relation to insights from data analytics for editorial decision-making.

However, some respondents warn about the risk of *"blindly relying on data insights"* (J4) without thinking about the journalistic value of the articles and therefore, for example, create click-bait articles, or *"the newsroom would shift toward what the numbers say works well: gloom and doom"* (T6). Therefore, it is important to *"extract trends from the data, but benchmark them with journalistic picks"* (J1). One editor in our sample claims to be *"data-informed, not data-driven, because there is so much data and we need enough smart people, with journalistic understanding, to interpret it"* (J4).

4.1.2 Commercial Logic. From the commercial logic, the role of a NRS goes beyond engaging users with diverse news content and is rather aimed at monetizing user actions, as one of the NRS product owners explains:

> *"Generating more clicks with recommendations is nice, but the company will be happier if those can be translated into money, whether that's ad revenues or subscription sales. At the end of the day, we need return on investment" (T1)*

At each news organization, two separate, independent departments operate under the commercial logic: B2B (business-to-business), which relates to advertisements, and **business-to-consumers** (**B2C**), which deals with subscription sales and paywall strategy. For B2B, the NRS is seen as a tool to drive ad visibility, building on the assumption that *"people who read more articles in a session saw more ads"* (T5) and *"more ad visibility is more revenue"* (C2). All respondents from the commercial logic refer to NRSs as a tool to increase incomes, and point to engagement as the key to generating more revenue. The importance of engagement is evident not only for advertising, but also for B2C, which starts from the assumption that *"the more people interact, the more likely they are to run into a paywall and eventually take action"* (T5). This user action can consist of logging in by creating a (free) account or taking a paid subscription.

What emerges among all respondents of commercial logic is that logged-in users are perceived to be particularly valuable, as they allow for *"better audience segmentation"* (C1), *"targeted advertising"* (C2), and ultimately, higher revenue per user. Thus, the NRS is deemed commercially advantageous when designed to *"recommend premium articles because to access premium content, you have to log in, and a logged-in user is worth more"* (C2), both for ad incomes as well as for turning a logged-in user into a subscriber.

Additionally, NRSs are viewed as a tool for improving user retention, as it is believed that effective NRS design and an attractive, personalized news experience help to retain paying users. Reducing churn (= ending subscription) links to maintaining a loyal, paying subscriber base, which is crucial for long-term financial sustainability. Our respondents indicate that in contemporary news markets, retention of paying users is a key concern as it *"reduces the reliance on fluctuating advertising income"* (T10). Not only is advertising highly sensitive to economic cycles, but throughout interviews with commercial logic representatives, it came up that a significant share of digital ad revenue is now captured by *"the big players"* (C2) such as Meta and Google, making it an increasingly smaller and unstable revenue source for news organizations [86, 92].

We note that both respondents from commercial logic and editors refer to the value of logged-in users and increasing engagement with the ultimate goal of user retention. Later in section 5, these shared envisioned goals for NRSs will be discussed in detail, linking them to editorial considerations and strategic business objectives.

4.1.3 Tech Logic. The largest group of respondents in our sample represents the tech logic. When asked about their specific envisioned purposes for NRSs, they mainly reiterate the journalistic and commercial goals discussed in the previous paragraphs. The tech logic thus entails a supportive rather than directive role: they position themselves as actants supporting editorial and commercial objectives with well-designed and well-integrated technologies. Our tech respondents do not perceive their position within the organization as setting strategic goals, but more as ensuring that the technical infrastructure can reliably serve the needs and goals of other stakeholder groups.

Although *"they have grown closer together"* (T6), the journalistic and commercial logic are *"different worlds"* (T1) and *"in between, we have to create a product that can respond to those different objectives"* (T2) and *"sometimes it goes very smoothly, sometimes it stumbles a little because there are other interests"* (C2).

> *"We must take editorial values, user interests, the commercial sales and advertising interests, and the organizational goals"* (T2) and turn them into a functional recommender system, which is *"very tricky"* (T3)

As previous empirical studies [7, 51, 53, 71] show, mediators are needed to bridge the gap between newsroom and commercial teams. This is echoed by our respondents:

> *"Our goal is to preserve the equilibrium that currently exists, the established balance of internal commercial objectives, newsroom concerns and other considerations"* (T7)

However, not only do the POs see themselves as responsible for *"bringing them together"* (T2), but all respondents from the tech logic see themselves as supporting the editors and commercial teams by *"building the recommender for the editorial decision-making and for specific business cases"* (T5). The tech logic is tasked with creating a NRS to support a *"good news flow"* (T7) and contribute to a favorable user experience. Since news consumption is highly individual, personalization is considered a tool to *"create a logical path in which the user recognizes his needs"* (T1). Nevertheless, they are aware that NRSs play *"a relatively small role in the bigger picture"* (T8) that constitutes a personalized user experience.

4.2 Metrics for NRS Evaluation

Based on the interviews, the previous section provides a well-founded understanding of the envisioned purposes of the NRS. When describing those specific purposes and the role that NRSs can play in achieving more high-level organizational goals, we note once again that respondents of the journalistic and commercial logic framed their reasoning in terms of the funnel rationale.

However, the goals discussed above are more *"strategic KPIs, but they are not that easy to capture in evaluation dashboards, so we have to break those down into some day-to-day metrics"* (T10). In this section, we describe all the metrics (whether currently used or not) cited by our respondents, grouped per funnel stage. To aid clarity, each metric is indicated in bold throughout our findings. Table 3 represents an overview of all metrics per funnel stage.

4.2.1 Reach. Reach refers to the proportion of people in the targeted audience who are exposed to the news offering that is distributed across different channels [25]. Although reach is not really discussed as an envisioned NRS purpose by any stakeholder group (see 4.1), aiming at maximizing reach is acknowledged as an important goal. In the context of NRS, *"we don't talk about reach a lot, but still, it's a very important goal"* (T10), since reach remains a fundamental first step, "*as you need to fill your fishpond"* (J4) with potential users.

> *"Everyone [every stakeholder] wants the largest possible reach as it helps to achieve their own goals. So, we looked if using a recommender algorithm could get more reach"* (T3)

The interviews reveal that reach is typically measured by counting the **daily active users** (**DAU**) and **monthly active users** (**MAU**). Additionally, multiple respondents mention **impressions** as a metric currently used to assess reach. Impressions count every time a page is exposed to a user, regardless of how long the page is being opened. On the one hand, there is *"reach measured as unique impressions for the journalistic mission"* (J3). Unique impressions refer to impressions of a page by unique users. When the same user (with the same user-ID) is exposed to a page multiple times, in other words, when multiple impressions occur on the same page, this is still only counted once in the analysis according to journalistic logic. This interpretation of reach is related to measuring how large an audience is reached with the published content. This reflects the values and professional responsibilities upheld by the editorial team (inherent to their logic). On the other hand, there are *"impressions linked to the number of page views for ad revenues"* (C2) as advertising revenue is based on **cost per mille** (**CPM**), the cost an advertiser pays for one thousand impressions of an advertisement. The commercial teams will therefore count all impressions

Table 3. Overview of Strategic Goals, NRS Purposes, and Related Metrics by Stakeholder Group and Logic

Stakeholder Group	Logic	Strategic Goal	Objectives/KPIs	NRS Purposes	NRS Metrics
Newsroom	Journalistic + Commercial	Reach	Inform many people as possible	Show news to a broad audience	Unique impressions DAU/MAU
		Engagement	Maximize user interactions with news offering	Make users read more Excite users with offer Discover mix of content	Article CTR Quality clicks vs. empty clicks Clicks per user (CPU) Attention time: • Non-regrettable attention time • Quality time spent Scroll-depth Page views per session Intra-user diversity: • Unique teasers shown per user • Page views to different categories • Attention time to different categories
		Conversion	Increase log-ins	Push readers deeper into the funnel	Share of logged-in users
		Retention	Retain loyal user base	Stimulate returning use	Subscriber attention time
Commercial departments	Commercial	Reach	Increase pool of potential subscribers	Maximize number of users	Total impressions
		Engagement	Monetize engagement	Drive ad visibility Increase willingness-to pay	Page CTR Page views
		Conversion	Maximize log-ins Sell subscriptions	Promote premium article	Paywall hits Log-in rate Subscription rate Subscriptions sold
		Retention	Long-term financial stability	Reduce churn	Subscription renewals Churn rate Visit frequency
Technological teams	Tech	Reach	Increase traffic to news platform	—	—
		Engagement	Enhance UX/news flow	Create logical reading path	Block CTR
		Conversion	Predict propensity to convert	Support paywall strategy	Paywall hits after recommended
		Retention	Predict propensity to churn	Increase user visits	Churn rate

This table lists the metrics as they emerged from our interviews. It should not be read as an exhaustive list.

without paying attention to the number of exposures per user ID, as it does not matter who has been exposed to the pages for advertising revenue. Page views differ from impressions in that they have a time aspect: a page must be open for a minimum of seconds (varies per organization) in order for the exposure (= impression) to be recognized as viewed (= page view). Impressions only measure exposure to a page; this does not mean that the page has actually been viewed or engaged with.

Measuring reach appears to have been more important in the past than it is now, where our interviewees mainly focus on measuring whether their NRS contribute to other objectives and KPIs (related to other funnel stages). This reflects a shift in strategic priorities, which shifts goals and, consequently, requires evaluation practices to be adjusted:

> *"Reach has declined in importance because at one time it was the only thing that mattered as we didn't have a paywall"* (J4)

In other words, other strategic goals have increased in importance with the advancement of a digital business model. This observation reflects the need to understand the adoption and evaluation of the NRS in the context of digital strategy, which is the focus of our third research question.

At the same time, our respondents recognize a rather limited potential of NRSs to directly increase reach, as recommendations tend to be better for *"already known, engaged users"* (J3) for whom data is available in contrast to the classic cold-start problem for new, anonymous visitors. Most (academic) discourse—as well as this article —concentrate on NRS use within news websites and apps, however, our respondents point to the potential of implementing recommender algorithms in *"personalized push notifications"* (C2, T3, and J3) and *"personalized email newsletters"* (C2, T6, C1, and J4) aiming at driving traffic to the news platforms. These alternative *"personalized products"* (T6) enable news brands to reach audiences through their *"own channels"* (C2) rather than relying on social media or search engines. A major limitation of both push notifications and personalized newsletters is that recipients must sign up via an already installed app or register for an email list. As a result, these approaches rely on engaging registered or logged-in users, which re-iterates the understanding that the NRS perform best for known users, rather than reaching entirely new audiences and increasing total reach.

Therefore, in what follows, we dive deeper into the use of NRS as tailored recommendations on the news platform with the aim to *"push them deeper in that funnel"* (C5) and eventually transform reach into *"more log-ins or more subscriptions"* (T3).

4.2.2 Engagement. The second funnel phase is user engagement, which, as explained in Section 2.2, includes attention to and (actionable) involvement with the content, but also interactivity on the news platform. All respondents, regardless of logic, describe engagement as an important strategic goal contributing to *"the ultimate goal: creating a good user experience"* (T8).

Accuracy Metrics. Our interviews show that the *"typical, simple evaluation metrics"* (T8) are currently used to evaluate the performance of NRSs in practice.

> *"For personalization, we agreed the key metric is **click-through rate**"* (J1).

Our interviews uncover that there are different types of CTR reported in the dashboards. First, there is **block CTR**, the click-through rate on the specific block or list of articles recommended by the NRS. This CTR *"computes just the CTR for the recommendations. So, a person sees five recommendations, did they click on any of them?"* (T8). The block CTR does not account for the content, theme, title, or other article-specific characteristics. To gain insight into the number of clicks on a specific article, a second CTR—namely the **article CTR**—is used to indicate how many users who were presented with a certain article (whether or not suggested by the NRS) clicked through. This

is considered an insightful metric by respondents of the journalistic logic to assess how individual articles are performing as *"higher clicks is promising, it conveys something about how they [users] perceive the articles"* (J6). A third CTR is the **page CTR**, which calculates the percentage of users who click through from a certain page to another page. For example, when 23 per cent of all users click through from the home or front page to another page, the homepage CTR is 23%. Similarly, page CTRs can be calculated for each *"article detail page"* (T6). This CTR is referred to as *"the overall CTR"* (T2) that counts all the clicks on a given page, without considering the content or specific sections on that page. Using such page CTRs to assess NRS performance is mainly interesting for the B2B department, as increasing page CTR serves as an indicator of how the NRS contributes to increasing page views and, therefore, increasing ad revenues. This page CTR takes into account every click, regardless of whether it is on a prominent article, a NRS-recommended article, or another link, such as an external affiliated page or ad. Hence, *"if the recommender is very good and you roll it out more, you might see the overall CTR increase"* (T8).

However, there is growing criticism of CTR as it fails to fully capture the complexity of meaningful engagement. Our respondents recognize that users may click on articles because of sensational headlines (clickbait) or out of curiosity, only to quickly abandon the articles. *"We assume that we can measure success with click-through, but who says so? What is the value of that to our reader?"* (T6).

> *"Clicks are easy to translate, like, if you generate more clicks, you generate more page views, more ads, and more ad revenue. So, companies are very interested in CTR rates… whether that's a good KPI for long-term satisfaction? I'm not sure"* (T8)

This is why respondents have started to look at ways to reduce ***"empty clicks"*** (J4), also called *"quick exits"* (T10) or ***"bounce rate"*** (T4), which refers to clicks that are dismissed again within a few seconds. Another click-related metric considered valuable is ***"click-per-user (CPU)"*** (C5), which tracks the total number of articles clicked within a session for each individual user. When the NRS is performing well, *"CPU can actually decrease, 'cause they're finding what they're looking for quicker, but lower CPU is not always a good sign, similar with CTR, it's challenging because you don't know what the intention is behind clicks."* (C5).

Despite this critique on CTRs, respondents indicate that it remains an informative metric: *"if you're modeling, click rates work as they have a very high correlation to what we deem important"* (T10). This confirms the paradoxical attitude toward CTR found by van Es and Nguyen [82]. Therefore, in addition to clicks, other evaluation metrics should be included "to get a holistic picture of engagement" (J2).

An often-used metric, recalled by all respondents, is **attention time** or time spent, which tracks how long users spend on a specific page.

> *"We put attention time as a main KPI as that one is understandable for all stakeholders: a lot of time spent means people find the content relevant and enjoyable, and that means journalists create good pieces, people are more likely to buy a subscription and it somewhat translates to more eyes on the ad banners as well"* (T5)

Our respondents believe that measuring the time spent consuming (recommended) articles provides insights into the users' evaluation of the content. In the case of a recommended item, high attention time is seen as an indication of whether the recommendations match user preferences and interests. However, similarly to CTRs being criticized, our respondents refer to attention time as not completely sound. For instance, users may leave the page open without actually reading it. Given that attention time cannot tell whether users found the article informative or valuable,

there is a call among some of our respondents to develop metrics such as ***"quality time spent"*** (T7) or ***"non-regrettable attention time"*** (T4).

Still, to date, the tech logic respondents interviewed had little success developing metrics to measure such qualitative, complex constructs. Therefore, the current dashboards combine multiple metrics to assess NRS performance and how they contribute to KPIs or strategic goals. Weighing attention time with **scroll-depth** might provide more informative insights. Scroll-depth measures how far down users scroll, assessing whether readers encounter an article's content beyond the headline and introduction. However, while deeper scrolling suggests reading more, it does not necessarily indicate satisfaction or comprehension of the content.

Another engagement metric frequently mentioned as currently used is **page views**, which measures the amount of views on a specific page. The total number of page views is specifically insightful for the commercial logic, since it relates to ad revenue. In addition, the number of page views generated by recommendations can be measured. While useful for assessing additional traffic thanks to recommendations, respondents stress there is no point in optimizing the NRS for page views because this metric can be artificially inflated by repetitive or irrelevant recommendations that lead to empty clicks rather than genuine interest. It might be more interesting to look at ***"page views per session per user****: does the NRS lead to more page views within the same session?"* (J1). Additionally, page views are insightful from a commercial logic for ad revenue.

The number of page views is not the only metric that is measured per session. Respondents also bring up *"attention time in a session, or number of articles per session"* (T2) as meaningful indicators to provide session-related insights. Linked to this, the **number of sessions** a user undertakes within a given period points to user visit frequency, which is believed to relate to user retention. This will be discussed in Section 4.2.4.

Beyond-accuracy Metrics. All the engagement metrics mentioned above can be categorized as "accuracy" metrics, following current literature. Our tech respondents describe these metrics as *"technically easy"* (T5) to implement and providing some *"short-term insights for each stakeholder"* (T2):

> *"Right now, the metrics the system is optimized for are clicks and time spent, while these often lean toward commercial objectives, they are not exclusively so. I agree they work well for commercial optimization, although engagement can also serve as an editorial value"* (T7)

These traditional evaluation metrics such as CTR, attention time, scroll-depth, page views and session-related metrics, provide a quantitative measure of engagement. Respondents stress that all these metrics should be viewed together when measuring the impact of NRSs to assess whether there is *"a correlation between CTRs, attention time and scrolling; to see if they stay longer on the article and are reading more articles?"* (T10). As illustrated when discussing clicks and time spent, our respondents point to a need for more qualitative, informative metrics that allow for more complex analysis of engagement at an aggregated level.

In the same vein, respondents advocate for *"beyond-accuracy metrics to get as much of a holistic view of the recommenders as possible"* (T8). Using only accuracy metrics to measure the impact of NRS on engagement does not take into account the editorial envisioned NRS purpose of helping readers *"explore the breadth and depth of the news offer"* (J4). To this end, respondents indicated that diversity, serendipity and coverage serve as indicators of meaningful engagement with news content.

The first beyond-accuracy metric which is frequently mentioned is **diversity**, which links to the perceived risk in the journalistic logic that NRSs could create filter bubbles. Respondents from

other stakeholder groups acknowledge *"diversity is an important one for editors"* (T1), and NRS evaluation should include an assessment of the diversity of articles recommended. However, *"values like diversity cannot simply be translated into ones and zeros"* (C1). Likewise, respondents point to the limitations of existing metrics:

> *Intra-list diversity serves as a poor metric for news, right? It's good for a search result page, but it doesn't tell you much about anything relevant for news"* (T7)

Instead, other ways of measuring diversity should be explored. Possible avenues mentioned by our interviewees are ***"intra-user diversity over time"*** (T7) or *"looking at the* ***different unique article teasers shown to a user****"* (C5). Yet, to date, no diversity metric has been implemented that is considered to be a good *"domain-specific metric"* (T7), nor does that seem to be a priority:

> *"A lot of it is still editorial curation. So, when they're talking about diversity, it's more as an extra nice thing to have in your recommender to push some more long tail"* (T8)

Noteworthy, our respondents connect diversity to unlocking long-tail items using NRS and *"link it to coverage: how good is the recommender in finding all the "edge cases" for the different users?"* (T8). In this light, evaluating coverage includes investigating whether *"specific or niche categories are coming forward more on the pages"* (C5), which links to measuring diversity as *"how many views are different article categories getting"* (C5). Another respondent mentions *"attention time to different topics to see diversity and coverage"* (T2). In addition to recommending a diverse set of articles, the NRS should reflect the broad topic coverage and help surface long-tail items. Our respondents indicate that recommendations should *"offer users a good mix, something interesting that they didn't know before"* (T6). This aspect of recommending novel, unexpected and positively surprising items is called serendipity by a few respondents. Although this concept is more common in existing literature [6, 70], there is still much ground to cover in conceptualizing and operationalizing serendipity, both for academic research and industry, as *"serendipity is a crucial metric, but it's challenging to interpret, making it difficult to get journalists to understand"* (T7).

This section shows that CTR, attention time, page views and session-based metrics remain the most widely used measures of engagement and evaluating the impact of NRS on engagement, as these metrics can be understood by all stakeholder logics. Nonetheless, respondents increasingly recognize the limitations of these simple, accuracy-based metrics and advocate the development of more sophisticated, qualitative metrics that capture meaningful engagement and user satisfaction, such as quality clicks and quality or non-regrettable time spent. Along that pursuit of nuanced metrics, respondents touch upon the challenge of translating beyond-accuracy metrics into computational measures because it requires more complex analysis at the aggregate level and a thorough reflection on *"how engagement is understood and how success [of NRS to increase engagement] is measured"* (T7).

4.2.3 Conversion. Section 4.1 has uncovered increasing conversion as an envisioned NRS goal, mainly for the commercial logic. Conversion refers to the process of converting users into logged-in users and then paying subscribers, making more logins and more sales a KPI for the news organizations in our sample. However, measuring the impact of the NRS on conversion is difficult as *"that's the outcome of a process or journey"* (C2), wherein *"a lot of different factors add to conversion, so it's difficult to measure"* (T10). Nevertheless, there are some metrics adopted, aiming to grasp the impact of NRSs on conversion.

First, there is **paywall hits**, which indicates the number of times a user has encountered a subscription barrier when reading an article. As mentioned above, *"for sales, it's a long process from the actual clicks on personalized recommendations to an uplift in actual sales"* (C5). Therefore, *"intent*

to buy [paywall hit] is easier to measure" (C5). Paywall hits is perceived as a valuable indicator or proxy for the intent to log-in or subscribe:

> *"Ultimately, a certain percentage of paywall hits always leads to a conversion, so the more paywall hits you have, the greater the chance that users will convert"* (T6)

Additionally, *"paywall hits somewhat demonstrate the relevance of the articles that the recommender spits out"* (T2). However, paywall hits do not necessarily translate into actual sales, as users may hit a paywall but abandon the platform without subscribing. In other words, this metric alone cannot fully capture the effectiveness of NRS in driving conversions, although it remains useful to assess how well the NRS steers users toward premium content.

A second conversion metric currently used is the **log-in and subscription rates**, which calculates the proportion of all visitors who have logged in or paid for a subscription. Some respondents mention that a more definitive way to measure the impact of NRS on conversion is tracking actual subscription purchases resulting from recommendations. If a user clicks on a recommended article and subsequently subscribes, this could indicate a successful NRS-driven conversion. However, measuring this direct link is challenging, as discussed above, the decision to create an account or pay for a subscription is the result of multiple interactions over time.

> *"Increasing sales is good for talking with other stakeholders: If you see that the recommender generates sales, it's all good, but I'm actually not a hundred percent certain how to measure it… it's seen as you clicked on an article that was recommended and the next thing you did was to pay for it, but I'm a little skeptical if that is a sale from the recommender"* (T8)

This is closely tied to a third metric frequently mentioned, namely, the absolute number of subscriptions sold. Although this metric does not specifically examine the role of the NRS, the total number of subscriptions sold over a period of time serves as an important KPI. It stands to reason that this is key for the B2C team (commercial logic), but remarkably, the number of subscription sales for a specific article is also mentioned by the editors in our sample:

> *"When people are pulling out a wallet, that's a good sign. Then we've either informed them enough and/or amused them enough too… and you can't derive everything from it, but it's still a valid metric"* (J4)

Our findings confirm the dual role of editors, balancing journalistic values with strategic business interests. During the interviews, this bridging between journalistic and commercial logic was evident when respondents showed newsroom dashboards that track both editorial and business metrics. Alongside anticipated metrics such as most-read articles based on clicks and attention time, these dashboards include how many subscriptions have been sold on each article. This observation confirms that increasing the number of log-ins and/or subscribers is not only a commercial, but also an editorial goal, as conversion is assumed to be a result of high-quality, relevant content that informs and engages readers.

4.2.4 Retention. The final funnel stage and the last envisioned NRS purposes mentioned by our respondents is building and maintaining long-term relationships with subscribers. Various stakeholders name retention as a key KPI: For the journalistic logic, if *"they pay and they stay"*, that indicates users *"see value in paying for our reporting and our content"* (J4). For commercial logic, subscription retention and renewals are a certainty of stable revenue. As such, retention is an important objective. However, it proves difficult to assess whether NRS effectively contributes to subscriber retention because of the *"long feedback loop"* (T10). To know which measures to look

at to understand retention, *"you need to derive some hypothesis about important metrics that might affect users' visits in the long run"* (T10).

The metric that was most frequently raised by several respondents is **churn rate**, which measures the percentage of subscribers who cancel their subscriptions. Each news organization in our sample has developed some (rudimentary) propensity models to predict who is at risk of churning and when.

> *"There already are some kind of churn models—when are people at risk of leaving, when do they stop using the product… but it's a first step and it's something worth investing in"* (T8)

Our respondents point to **frequency** playing a key role in retention, as inactive users are more likely to cancel their subscriptions.

> *"The chance of you actually ending your subscription increases quite a lot if you don't use the product"* (C5)

This suggests the assumption that there is a correlation between frequent use of the news platform and churn reduction. In other words, our respondents believe that if the NRS successfully encourages regular content consumption, personalized recommendations could have a positive impact on subscriber retention. The NRS is optimized *"to chase engagement because it's our belief is that highly engaged users are less likely to churn"* (T11). However, to date, there is no data to support such causality.

Another recently adopted metric cited by our respondents is **subscriber attention time**, which tracks how long paying users engage with (recommended) news content. However, defining a suitable measurement approach remains complex:

> *"How are we going to measure it? Per subscriber session, or per article? And can we weigh it against length and content of article, and what about video?"* (J4)

However, evaluating whether the NRS contributes to retention remains an ongoing challenge, as one respondent explains:

> *"We have agreed that all KPIs related to conversion, churn, and retention… that's long-term for now, we don't know enough yet about how exactly to integrate it into the [propensity] model"*(T2)

While propensity models offer a promising approach, *"it's hard to translate the recommender as part of a digital strategy into retention, because that's a long-term thing"* (T4). Performing long-term A/B tests could provide long-term insights into how to decrease churn or increase user frequency and subscriber attention time. However, *"A/B testing for long-term effects is all about trying to find that tradeoff of ńWhat's long enough?»"* (T8). We note that multiple respondents point to a tradeoff between short-term and long-term measurements. One respondent explicitly attributes this gap to a different focus for stakeholder logics:

> *"Journalists and editors are used to everything going fast fast fast… If you write an article, a day later it is no longer relevant. That is strongly opposed to a strategy, then you have to look at the longer term. But they are not used to it"* (T5)

This results in a specific practical challenge, because *"it's difficult to run an experiment for some time because while that test is running, some numbers might drop. Then editors start to complain. So*

I'm still figuring out how to set this up and convince them that retention is in everyone's interest, and looking at the long term will be beneficial" (T4).

This highlights the delicate balance between experimental rigor and building on well-informed *"hypotheses"* (T10). In order to develop valid assumptions, respondents call for weighing "quantitative" metrics such as visit frequency, attention time, and subscription cancellation/renewal against more qualitative metrics about user experience. Accordingly, *"Net Promoter Scores (NPS), reader surveys and polls done by the customer service and retention team"* (J4) to gain deeper insights into factors affecting retention. This demonstrates that news organizations also consider end users to be important stakeholders. By explicitly asking them about their experiences, our respondents claim to gain insights into the factors that influence user satisfaction and learn lessons about subscriber retention. Nevertheless, NPS scores are criticized in the literature for their methodological simplicity and weak predictive power. It lies beyond the scope of this work to examine **customer relationship management** (**CRM**) in news organizations.

Our findings illustrate the variety of KPIs and different metrics cited by different stakeholders within each funnel stage to evaluate NRS performance. In section 5, we further elaborate on the complexity stemming not only from NRS-specific metrics, but also from the multitude of content performance and user tracking data used by editorial, commercial and technical teams. In what follows, we first synthesize the findings from Sections 4.1 and 4.2 in Table 3. After that, we report the findings related to the third research question, which delves deeper into the challenges associated with NRS adoption.

4.2.5 Overview of NRS Purposes and Related Metrics. Building on our conceptual framework in Section 2.4, Table 3 illustrates how each of the three stakeholder groups translates strategic goals into objectives and KPIs, linked to envisioned NRS purposes and operationalized through specific metrics.

The newsroom is primarily represented by (deputy) chief editors in our sample, who operate under both the journalistic and commercial logic [71]. This stakeholder group considers "reach" an essential strategic objective in light of their journalistic mission to inform a broad audience [22], and is measured in unique impressions and the number of daily or monthly active users. However, as our interviews focus on on-platform NRS implementation, respondents state the limited direct impact of RSs on attracting new users. The newsroom's primary envisioned NRS purpose lies in increasing engagement, defined not only in terms of article consumption (e.g., article CTR, clicks per user, attention time and scroll depth), but also through facilitating exposure to a diverse mix of content (reflecting beyond-accuracy values). While these metrics help monitor how users interact with the platform, editors stress that higher engagement does not equate to better journalism. They express concern about solely and overly relying on traditional metrics like CTR and time spent, advocating instead for more qualitative indicators such as "quality clicks" or "non-regrettable attention time" to better assess the role of NRSs in achieving editorial objectives. Engagement is also viewed as a catalyst for downstream outcomes in terms of conversion, whether that is log-in or a subscription. Ultimately, according to editors, increased engagement indicates positive user experiences, which contribute to the long-term goal of retaining loyal readers

The commercial departments frame their strategic goals in terms of revenue generation through both advertising (B2B) and subscriptions (B2C). Reach is measured via total impressions, which serve as the basis for advertising incomes linked to CPM (cost-per-mille) models. Here, engagement is tied to monetization, with page views and page CTR acting as key metrics that directly impact ad visibility. The commercial logic also includes B2C professionals who see engagement as instrumental for driving conversion, particularly by optimizing NRSs for paywall hits to increase the likelihood of logins and subscriptions. These conversion goals are tracked through metrics such as login rates, number of new subscriptions, and subscriptions sold However, even within

the commercial logic, tensions arise around the role of the NRS and which purpose the recommender should fulfill. As one respondent explains:

> *"If you want more subscribers, you have to make more premium content, so put more articles behind the paywall, because then you increase your number of paywall hits. Then, you'll see that some of the readers who consume free content will go elsewhere to read. So, you'll see a decrease in reach, and the moment that happens, the newsroom gets a visit from the ad department, who see reach and eyeballs decreasing"* (T4)

Commercial departments are also concerned with retention, which links to financial stability, with metrics such as subscription renewals and churn rate serving as key indicators. Similar to editors, it is assumed that more engagement fosters long-term relationships with users, which might explain the use of engagement metrics to assess the NRS' impact on retention.

Finally, the technological teams may appear to discuss fewer metrics compared to the newsroom and commercial departments based on Table 3. However, this should be understood in context. Throughout these interviews, there was comparatively little discussion of system-level evaluation of the algorithm. Importantly, this was not a result of the interview design, but of how respondents from the tech teams consistently framed their role: they described their task as building a NRS that enables the newsroom and commercial departments to achieve their objectives. Consequently, technological stakeholders align their evaluation practices with the objectives defined by the other logics, rather than developing a separate set of organizational goals for the NRS.

4.3 Evaluating NRSs within Broader Digital Strategy

The third research question of this study aims at looking beyond the system-level metrics and goals. Section 2.3 shows that NRS integration cannot be viewed separately from the broader digital strategy of news organizations. Our interviews echo this approach to *"the recommender as part of a digital strategy"* (T4), which entails challenges related to computational difficulties in system integration, as well as organizational barriers arising from divergent stakeholder goals and interests.

4.3.1 Computational Challenges. Our interviews uncover five computational challenges that complicate the implementation of NRSs. First, NRSs face technical challenges related to **scalability** and real-time performance:

> *"Platform performance in high-traffic times is crucial: When there is breaking news that is so big that millions of users visit our platform at the same time, the NRS needs to be fast enough… now we had to take it offline"* (J2)

Tech respondents link this to the speed of publishing news items and the so-called cold-start problem: *"The downside of NRS is that you're often a little behind"* (T5) because *"very recent news needs to be warmed up if nobody has seen or clicked it yet"* (T6). Although not every organization in our sample struggles with *"handling the load of more than a hundred requests per second"* (T2) to the same extent, the tech respondents who report limited issues with scalability signal other difficulties. A second computational challenge is the issue of **data sparsity**, which limits the ability of NRSs to make accurate, personalized recommendations. This refers not only to a lack of sufficient data for each user, but also unclear data, as many users *"block cookies"* (T5), *"browse in incognito mode"* (T4, T5, and T6), or *"share accounts"* (T4), making it difficult to collect reliable behavioral data.

> *"The median user had five clicks, so determining preferences on five noisy clicks? That's really sparse"* (T9)

Even when data is available, there is a **lack of quality data insights** adequate for extracting an informative, nuanced understanding of user behavior. Whereas metrics provide quantitative reports, *"what we miss most of all is the qualitative feedback"* (T1). Multiple respondents warn for reliance on quantitative data as it might lead to an insufficiently nuanced analysis: *"We have so much data and not enough smart people, with journalistic understanding, to interpret it. Too often I look at data reporting saying this graph goes up and this graph goes down… Yes, but why?"* (J4).

> *"There's way too much tracking of data, and it's hard to make that into actual numbers that are stable and reliable because there are typically a lot of technical things you need to adjust for. So right now, it's not the data gathering that's the problem, the problem is the data processing and metric design"* (T10)

This respondent touches upon the next complicating factor in NRS implementation, namely the **simultaneous execution of multiple A/B tests**. As different teams within the same organization are each running different A/B tests to optimize their KPIs, it can distort results and introduce confounding variables: *"If there are too many things running at the same time, it's definitely something that could skew the results a bit as well"* (C5). In each organization in our study, there seems to be no overarching overview of the A/B tests planned or being performed across the platforms *"when we run the A/B test, hopefully, no one else will do any radical change, like changing the whole color scheme or something that can really affect our outcomes"* (T8).

The last computational challenge that emerged from our interviews is the lack of a prominent place for NRS output, which some respondents refer to as *"the story of the chicken or the egg"* (J1 and T1). This refers to the issue of NRS-generated recommendations often being relegated to *"a hidden corner of the website or app, down on the page, where there is, of course, little interaction with it, which influences its performance"* (T4). Here, respondents refer to the importance of the interface, an issue often linked to Francisco Martin's keynote at RecSys 2009, stating that up to 50% of the value of recommenders comes from a well-designed interface [48]. As one respondent notes, the NRS algorithm could be state-of-the-art, as long as only a small fraction of the users encounter the recommendations *"peanuts remain peanuts"* (T2). This is particularly challenging as many of the tech respondents state having *"to prove the value of NRS by numbers"* (T1) before it can be rolled out further or gain a more prominent place. However, such performance numbers cannot be obtained if the NRS is not positioned *"above the scroll"* (C1). This creates a paradox where the NRS must demonstrate its value through engagement metrics, yet it cannot reach its full potential without first being given sufficient visibility.

4.3.2 Organizational Challenges. The qualitative nature of this research allows us to look beyond the system-level computational challenges and explore how NRS adoption is embedded within the broader digital strategy of the news organization. In their shift to a digital-first distribution strategy, respondents point to exploring various innovative technologies, of which NRSs (and personalization) are only one example. To understand the context within which NRSs are developed, our interviews inquire about organizational challenges. A first major challenge that respondents refer to is measuring the long-term impact of decisions made. When discussing retention metrics in Section 4.2.4, some tensions regarding long-term goals already become apparent. Evaluating the long-term impact of such strategic choices is crucial for both the organization's commercial viability and journalism's democratic role in informing a wide audience [22, 28]. Innovation is part of the search for *"how do we ensure that what we do is distinctive?"* (J4) within today's complex digital (news) ecosystem.

> *"In the days when we made front pages on paper to sell physically, news and facts were the currency. But today, news is no longer a currency because it is instantly copied,*

> *and it is everywhere online for free. You cannot distinguish yourself with news by itself. So, we are looking at what do we add? Where are we unique? What can't just be copied online? And with that comes now: What can't just be written by AI?"* (J4)

Editors and commercial teams alike are forced to consider what aspects of their journalistic offerings remain uniquely valuable. Here lies an opportunity for NRSs and personalization to elevate the user experience and contribute to that distinctiveness, serving as a value proposition to consumers. However, while there are opportunities, it is equally important to consider the long-term negative impacts. As one respondent points out: *"We started worrying about news avoiders about 3 years ago, today I worry just as much about the doomscrollers who keep reading, they don't go away, but it's not sustainable… in the long run, it's not good"* (J4). To this end, editorial oversight on NRSs and personalization from the journalistic deontological perspective is needed to avoid potential negative side effects.

Another potential long-term negative effect of NRSs that emerged during the interviews is the *"fear of a quality drop"* (T10 and J2) in the curation of the homepage. This concern comes from the perception that NRSs would recommend *"too much of the same"* (J1) and create *"bubbles"* (J2 and J3). However, across all four news organizations in our study, these worries have declined considerably compared to the early days of NRS integration. We observe that this mainly has to do with a kind of "organizational learning" about what NRSs actually can (and cannot) do. For example, one editor justifies the use of NRSs by explaining that the recommender is *"only going to adjust the order in which articles, published by the journalists, are shown"* (J4).

> *"So we may find an article about some celebrity not super important, and it might be more prominent among those personalized items, well we did choose to publish it, so we stand behind it, right?"* (J3)

Thus, we observe that the editors we interviewed view NRSs as a tool that assists them, rather than a replacement for human curation. This minimizes the risk of reduced quality or shaping of filter bubbles. Moreover, technologists also explicitly acknowledge these concerns; however, this is *"not to make them anxious, but to make it clear that we are working on it. We have already been able to show that we contribute to higher diversity, so those statistics do help"* (T1). Another strategy is "framing" personalization *"as FOMO-tization, which is kind of the opposite of a filter bubble"* (T10). Here, the belief is that NRS can *"create a sense of FOMO"* (T3). According to our respondents, the so-called fear of missing out can be triggered among readers *"by recommending things that other people have read but you haven't yet"* (T6). However, each of these initiatives requires a lot of effort and involvement from several stakeholders to get this knowledge across. Nonetheless, our findings may indicate that filter bubble concerns in the past were mainly informed by an overhyped impact of NRSs, resembling the technology hype cycle, where initial fears often give way to a more nuanced understanding of implementation realities.

A second issue where hesitancy, resistance, or concern arose at the start of NRS implementation at all four news organizations is the relinquishment of editorial control.

> *"There was some initial backlash because editors are used to being in control"* (T9)
> *"At first, there is some feeling of losing control and not understanding what's going on, combined with fear that they are gonna be replaced"* (T10)

This links in part to the previously mentioned fear of *"low-quality recommendations"* (J2). A common strategy to deal with newsroom reluctance is to install *"some safeguards"* (J3). Firstly, multiple respondents from all four organizations refer to ensuring human oversight with a *"human-in-the-loop"* (J1, J4, T2, T6, T9, T10, C1, and C2). A second safeguard corresponds to what Smets et al. [71] refer to as "preconditions":

"Editorial requirements should prevent strange quirks of the NRS" (T1)

Such so-called quirks could result from *"blindly following the data"* (J4). The key to *"building trust with editors"* (T10) is *"guarding the balance between what the numbers say and what we think is important from a journalistic point of view"* (J1).

These difficulties are *"overcome slowly but surely"* (T2), which allows *"flexibility to explore the opportunities of recommendations"* (J2). A full examination of the management strategies involved in intermediating perspectives is beyond the scope of this article. For insight into stakeholder alignment strategies, consider the work of van Es and Nguyen [82], Ranaivoson et al. [65], and Vandenbroucke and Smets [84].

5 Discussion

Our findings underscore the complex nature of developing and evaluating NRSs in a multi-stakeholder environment. While different actors share an interest in making the NRS successful, they diverge in the purposes, objectives, and metrics they associate with success. In this discussion, we unpack several tensions related to NRS evaluation. First, engagement emerges as a shared but contested goal. Second, while simple metrics dominate, there is a call for more informative, as well as beyond-accuracy metrics. Next, more metrics do not result in better evaluation, and finally, NRSs should not be understood in isolation, but rather as part of broader personalization efforts.

5.1 Engagement as a Shared Goal Between All Logics

Our analysis reveals that currently, all stakeholders agree that increasing user engagement is the overarching strategic goal to which NRSs should contribute. Despite the common belief that NRSs should support users in reading more (e.g., CTR, page views, scroll depth), reading more frequently (e.g., session frequency), and reading longer (e.g., attention time), these indicators only partially reflect what different stakeholders interpret as meaningful engagement.

Our analysis shows that the ways engagement is conceptualized and subsequently operationalized into NRS purposes and metrics vary across stakeholder groups.

Under the journalistic logic, engagement is linked to editorial considerations such as fostering a well-informed readership [18, 28]. Editors emphasize that NRSs should encourage users to discover both the breadth and depth of the reporting while safeguarding journalistic values. Here, engagement refers not only to the quantity of interaction with articles, but also reflects cognitive and emotional involvement with the content [4, 14, 73]. This reflects the journalistic conception of engagement, which indicate a rather instrumental view of successful user interaction: editors assume that more time spent on the site, more articles read, or more scrolling automatically indicate valuable engagement. Although this assumption is consistent with the broader organizational rationale surrounding a focus on increasing engagement, nuance is needed. A significant portion of the audience values convenient, concise, and easy-to-navigate news products. For some users, the ability to quickly stay informed is a core value proposition; long-standing formats such as the Wall Street Journal's "What's News" short summaries are an example of this preference for efficient updates. From this perspective, a longer session does not necessarily indicate higher satisfaction or meaningful engagement. In addition, longer sessions may also indicate difficulties in finding the desired or relevant information. This suggests that engagement metrics in the context of news recommendations should be interpreted in relation to users' goals and situations, and that organizational interpretations of "more is better" should be critically weighed against audience-focused studies.

In contrast, the commercial logic adopts a more transactional view on engagement with the main KPI being monetization of user interaction and maximizing the transactional outcomes that result from it [4, 15]. Here, attention is treated as a currency, with page views and CTR serving as key proxies. Still, commercial stakeholders echo editors' belief that engagement indicates a positive user experience. Prior research shows that willingness to pay is strongly influenced by quality perceptions and reading behavior [9], which explains why conversion-related outcomes often correlate with engagement metrics [21].

Finally, the technological logic frames engagement as a reflection of a favorable user experience on the platform. However, our tech respondents emphasize the difficulties involved in deducing user satisfaction based on data about how users interact with the news platform, the recommendation system and the news content. Additionally, the technological team is tasked with measuring user engagement in line with the different interpretations of journalistic and commercial logic.

The theory of institutional logics [7, 46, 51] suggests that stakeholder logics are dynamic and can be integrated to combine practices and bridge perspectives. In this line of work, Svensson [74] points to the possible emergence of a new hybrid "personalization logic". Although our findings do not indicate the emergence of a new, shared logic, they show that engagement plays a key role in bridging the different logics within news organizations. In this sense, engagement acts as a boundary object [12, 46]: a shared reference point that enables communication and collaboration across stakeholder groups despite divergent goals and perspectives.

5.2 Struggles to Move from Simple to Informative Metrics

Our analysis shows that metrics based on easily observable behavior, such as user–article interactions, continue to dominate evaluation practices in industry. Our respondents refer to these as "simple" metrics because they are straightforward for technologists to implement, measure, and report. We believe this ease of use also helps explain why such metrics are so well established in the RS literature: they provide a quantifiable proxy for more complex goals such as engagement, conversion, or retention. As a result, they have become the default choice for evaluating NRS performance, both in industry and in academic research [6, 70].

This is most visible in the engagement metrics reported by our participants (see Table 3). The majority of metrics mentioned, e.g., page CTR, page views, block CTR, reduce engagement to single interactions. Even later funnel stages are measured in similar ways: conversion is captured as "paywall hits after recommended", and retention as "subscriber attention time" or "subscriber visit frequency". The same dominance of simple interaction-based measures has also been noted in the literature, where CTR [41, 57], page views [93], or time spent [44] are consistently the most studied indicators of engagement [6, 40]. While convenient, these metrics fail to capture the multidimensional nature of engagement and the broader strategic goals of news organizations.

At the same time, our interviewees emphasize that traditional engagement measures are too coarse to reflect the complexity of user experiences and long-term organizational outcomes. Stakeholders of different logics argue that more qualitative and informative metrics are needed to better capture the contribution of NRSs to strategic goals such as meaningful engagement (whatever that implies, according to them), conversion, and retention. For instance, respondents proposed metrics such as CPU, quality clicks, or non-regrettable attention time. These indicators aim at going beyond superficial interaction and assess engagement at a more aggregate level. However, they also acknowledged uncertainty about how such metrics should be defined, operationalized, and implemented in practice. This points to an ongoing tension: while simple metrics remain dominant due to their ease of use, there is a clear demand for more nuanced approaches that align evaluation practices with the broader strategic, long-term goals of the news industry.

5.3 Translating Beyond-accuracy Values into Metrics Remains Aspirational

While academic research increasingly focuses on translating beyond-accuracy values such as diversity and serendipity into metrics [6, 37, 89], our interviews suggest there is only limited effort to operationalize these in practice. Our respondents acknowledge beyond-accuracy values, such as promoting diversity, unlocking long-tail items and avoiding the pitfalls of personalization (i.e., filter bubbles [49], over-reliance on data resulting in poor quality news or click-bait [81]) as important. However, in day-to-day practice these are described as abstract aspirations rather than concrete measurable goals.

This has much to do with the context- and content-specific nature of such values in news. For example, what should be diversified (e.g., topics, sources, political viewpoints, geography, formats) and over what time horizon depends on editorial priorities and on the current news agenda. According to our respondents, existing metrics in the literature (e.g. **intra-list diversity** (**ILD**) [49, 88], and **intra-list similarity** (**ILS**) [6, 70]) offer useful starting points but are not yet fully suited to the news domain because they do not capture aspects such as recency or editorial relevance. While respondents point to the need to explore other diversity metrics, with examples being intra-user diversity over time or unique article teasers shown to a user, developing new beyond-accuracy metrics themselves is currently not a priority for practitioners. Since the risk of filter bubbles is not perceived as a threat (as described in Section 4.3.2), informative metrics to assess NRS impact on diversity, coverage, or serendipity are only desirable by-products. Instead, an estimate for those three beyond-accuracy values is calculated by aggregating the known engagement metrics, such as the number of unique views and attention time, balanced against different content categories. Although it is often brought up during interviews that NRSs can play a valuable role in surfacing long-tail content and surprising users with unknown but relevant articles, there seems to be little momentum to translate these values into metrics that reflect the specific characteristics of the news domain.

Hence, while academic research has made important progress in formalizing beyond-accuracy metrics such as diversity, serendipity, and coverage, our interviews show that these are only weakly adopted in industry. This suggests that there is a need for more ready-to-be-implemented metrics for these beyond-accuracy values. This is precisely where another obstacle appears: many of these values are highly content- and context-dependent, and therefore require normative choices about what should be diversified or surfaced for which audience and under which editorial priorities. Furthermore, to date, the limited news datasets from real-world news platforms that are publicly available for academic research lack the complexity, real-time nature, and variable diversity that legacy news companies encounter in practice [87, 91]. This complicates any research into context- and content-specific beyond-accuracy metrics suited for news companies.

5.4 More Metrics Do Not Lead to Better Understanding of NRS Performance

Despite the call for more informative metrics, merely adding more metrics does not facilitate the evaluation of NRS performance and its impact on KPIs. Instead, a "jungle of metrics" arises in which the abundance of indicators and measures complicates, rather than helps to extract insights and subsequently aid decision-making. Table 3 showcases a multitude of metrics to reflect various envisioned NRS purposes and different KPIs cited by the different stakeholders within each funnel stage. As a result, there is a struggle to determine which metrics to prioritize, what they signify, and how to optimize the NRS accordingly. In addition to tracking NRS performance, a multitude of content performance and user tracking metrics are used across editorial, commercial, and technical teams. The rise of data tracking has led to an overload of different metrics in dashboards, which leads to stakeholders experiencing difficulties interpreting the many figures and statistics in relation to their day-to-day practice. Our interviews reveal an increase in complaints about too

many numbers and too many goals among editors, and requests for simpler goals and a defined set of key metrics. However, the challenge to streamline KPIs and metrics is not only due to the amount of data and metrics, but also exists in a differing focus as the editorial and commercial considerations are not aligned:

> *"The editorial team is very concerned about whether a user is engaging with content, and we've noticed that we're two teams working on two different sets of goals and different kinds of metrics"* (C5)
> *"The newsroom is used to talk about absolute numbers: total number of clicks per day on an article, total number of page views… and that is what their goals are set on, but the problem is that personalization doesn't work like that. The recommender is optimized for read rate, click rates or sales rates"* (T10)

This mismatch between looking at absolute numbers versus ratios reiterates the need to revisit the current KPIs and metrics. This requires alignment of both high-level strategic goals as well as the specific NRS purposes across all teams and defines a clear, limited set of metrics which are agreed upon by all recommender stakeholders.

On the contrary, many respondents also highlight the advantages of data insights, both for the journalistic and commercial logic. A particularly noteworthy finding is the emphasis across respondents on the higher value of logged-in users compared to anonymous visitors. For the commercial logic, this makes sense as logged-in users generate more advertising revenue because more granular data allows for better targeting. Additionally, the probability for logged-in users to convert into paying subscribers is higher. Interestingly, this also emerged during interviews with editors, who see login behavior as an indicator of user appreciation for journalistic reporting. They perceive users logging in to access premium articles as a result of good journalism in line with their journalistic mission [22]. Beyond the journalistic interests, there are also strategic editorial considerations at play: logged-in users provide more affluent data that enable better analysis of audience trends and content performance, shaping editorial decision-making in line with business goals. Finally, under the tech logic, logged-in users mitigate data sparsity issues, enabling better recommendations.

5.5 NRS and the Personalization Project as Part of a Broader Strategic Shift

Our analysis reveals that stakeholders do not see NRSs in isolation from the broader personalization "project" that encompasses various data-driven strategies aimed at creating a personalized user experience on the news platform. Beyond on-platform content recommendations (which is the focus of this study), this personalization project includes NRS applications such as tailored push notifications, e-mail newsletters, and individualized CRM. Throughout the interviews, respondents refer to the use of predictive models to determine when a logged-in reader is likely to subscribe or calculate the propensity to cancel a subscription. Such user-level monitoring is called individualized CRM, where individual user journeys are tracked to make a personalized offer at the right moment. That is a promising avenue for future work that links the evaluation of NRS performance to user preferences. Our respondents acknowledge that technology-wise, there are many possibilities to adopt innovative technologies, but from a strategic perspective, the question is: What business use cases are interesting?

Instead of focusing solely on improving their NRS design, the industry needs to evaluate how it fits into the wider shift toward a digital-first strategy. This includes tradeoffs in allocating resources to alternative innovations, such as recommending different content formats and leveraging generative AI tools to summarize articles or improve personalized email newsletters. By shifting the

focus from solely NRSs to a more holistic approach, recognizing personalization as a fully tailored news experience, our respondents stress the importance of return on investment.

This perspective highlights that NRS evaluation involves more than measuring algorithmic performance; it requires considering its placement within the platform, the multiple purposes it serves, and the value it adds relative to other personalization initiatives. This approach contrasts with much of the academic literature, which continues to emphasize algorithmic accuracy at the system level [34]. Our findings echo the call to position NRS research within the broader organizational context it operates, recognizing that NRS adoption is not a stand-alone effort, but one component of a larger personalization project that is part of a strategic shift toward digital distribution in legacy news organizations.

5.6 Strategic Tensions Between Stakeholders

5.6.1 The Long-term Implications of Innovations. A key challenge that cuts across both the computational and organizational levels is evaluating the long-term impact of strategic decisions, such as adopting innovative technologies. In the case of NRSs, our analysis reveals that currently, short-term engagement metrics such as subscriber attention time or visit frequency are mostly used to evaluate the impact on retention. However, this approach cannot capture long feedback loops or causal attributions. Setting up long-term A/B tests seems a solution to this, but in practice, it does not prove to be easy. Similar to the discrepancy between examining relative and absolute metrics described above, our interviews point to a discrepancy in temporal orientations of the different stakeholder groups. The newsroom is accustomed to the daily rhythm of news cycles, which is measured with immediate content performance data. On the contrary, insight into the journey leading up to conversion, as well as analysis of factors contributing to subscription renewal/cancellation, requires long-term measurements. Technologists find themselves caught between these two logics, tasked with building evaluation practices that include metrics that satisfy both perspectives. This dilemma extends beyond NRSs and includes all aspects of the personalization project and other strategic decisions related to innovations such as generative AI. As highlighted in our Background section 2.3, news organizations are in a state of flux, navigating the shift to a digital-first strategy. This mismatch between short-term metrics and long-term goals is one example of tension between the stakeholder logics.

5.6.2 Stakeholder Alignment. Our analysis suggests there are some potential areas of tension surrounding the integration of NRSs within existing workflows and business models. It is beyond the scope of this study to map out all the underlying tensions between competing institutional logics, but our overview of envisioned NRS purposes and metrics reveals differences in focus, varying conceptualization of KPIs and, as a result, the use of conflicting metrics. During the interviews, respondents pointed out the possibility that the further development of NRSs could be delayed by these differing perspectives.

Referring back to the theoretical framework of diverse institutional logics [7, 51, 78], we point here to the tendency of stakeholders to protect and re-define the boundaries of their professional identity with the emergence of data-driven innovations. The need to align new practices with the norms and values of the stakeholder group [12, 46] might cause tensions. Our analysis reveals that all four organizations implemented their first NRS between 2018 and 2019, and the early experimentation was *"met with backlash"* resulting in *"having to dial it back"* (T9). This hesitance and resistance in newsrooms has also been identified in previous studies, reflecting fear about the negative effects of personalization, such as filter bubbles [49]. On the other hand, newsrooms are reluctant to NRSs because of a perceived loss of control over content curation [50].

Over time, however, these concerns are resolved. Our findings suggest that processes of organizational learning and stakeholder alignment have gradually built a better understanding of NRSs in newsrooms and commercial teams. Our respondents briefly hint toward knowledge management as a tool to align the different stakeholder perspectives. Although very interesting, a thorough analysis of the management strategies for stakeholder alignment mentioned during our interviews is beyond the scope of this article. Nonetheless, we would like to highlight that our analysis shows that adopting a collaborative decision-making process around NRS design by *"incorporating those who work with the system into its development process"* (T8) proved to be beneficial in the organizations studied. In other words, involving the various stakeholders from the start facilitates the introduction of NRSs. This affirms why the multi-stakeholder approach in research is so highly valuable for studying the real-world settings in which NRSs operate [1, 33, 40].

6 Limitations and Future Work

The results and contributions of this work should be viewed in light of the following limitations. Methodologically, two limitations should be noted. First, our interviews were conducted at two different times, with roughly a one-year gap. This temporal distance may have introduced slight differences in respondents' perspectives due to organizational or technological changes throughout the year. At the same time, the extended collection period proved valuable, as it allowed us to observe how stakeholders' reasoning about NRS goals and metrics remained consistent with the logic of each stakeholder group. Second, our sample is limited to four large, profit-driven news organizations in Northwestern Europe. This constrains the generalizability of our findings to other regions and excludes contexts such as public service media or smaller, local outlets. A further scope-related limitation is that each organization in our sample operates as a multi-brand group. Our study does not delve into the added complexity of negotiating personalization across brands and organizational levels. Our interviews suggest differences: some brands develop their own NRS in-house, while others depend on group-level systems. Additionally, we highlight the indirect involvement of journalists in NRS development. Echoing Smets et al. [71], our interviews confirm that journalists are rarely included directly in discussions; rather, their perspectives are represented by editors. In practice, when approached for interviews, journalists or newsroom analysts often redirected us to technical profiles. This limited our ability to capture journalists' first-hand perspectives on NRSs.

Our study has important implications for future RS research, particularly as news organizations increasingly seek comprehensive approaches to evaluate personalization efforts, of which NRS are one component. Traditional offline evaluation methods in computational research focus narrowly on algorithmic optimization through system-level metrics; however, these are poorly suited to capturing the impact of NRS on broader organizational goals. Therefore, we encourage future research to critically examine evaluation frameworks that account for the strategic role of personalization within the wider digital news ecosystem, building upon our work. Specifically, future research could also take up the development and validation of more informative metrics that were suggested by our respondents, such as *non-regrettable or quality attention time, quality clicks per user*, and explore more retention metrics beyond the combination of engagement metrics. Next, we call for future work to replicate our research questions (i) in other European media systems or non-European contexts, and (ii) in other types of news organizations. Future work could also explore how personalization is coordinated across brands within the same media group. Moreover, further research should center on journalists themselves. Ethnographic and qualitative studies [53–55, 68] provide promising first steps, but more work is needed to understand how journalists perceive and interact with NRS, and how editors mediate these perceptions in organizational decision-making about evaluation and development. Finally, both the theoretical framework in

this work and our findings highlight the dynamic and adaptive nature of stakeholder logics. The introduction of personalization efforts, among other data-driven practices, creates tensions around KPIs and evaluation practices, resulting in a "metrics jungle" that complicates alignment. We encourage future studies to examine management strategies on how to deal with innovations that reshape professional boundaries and trigger processes of re-alignment across logics.

7 Conclusion

The evaluation of NRSs in practice remains a complex challenge, shaped by the diverse objectives of multiple stakeholders within news organizations. While traditional engagement metrics such as CTR and attention time remain commonly used, they provide only a partial view of the actual impact and effectiveness of NRSs. Despite growing academic interest in NRS evaluation, existing research primarily focuses on technical optimization and user-centric metrics, often overlooking the broader organizational and strategic considerations that shape real-world implementations. This disconnect results in a lack of understanding of how NRSs function within the complex organizational dynamics in news media outlets. This study has focused on the provider-side of NRSs, examining how internal organizational stakeholders define and interpret performance metrics. While users remain a crucial stakeholder group in the broader NRS ecosystem, their perspective falls beyond the scope of this study.

In this article, we extend our earlier work [83], further exploring the various organizational stakeholder perspectives on evaluation metrics and addressing this gap between academic research and industry practices. By relying on insights from social sciences literature on institutional logics and organizational dynamics, we provide a comprehensive overview in Table 3, building on our conceptual framework in Table 1. Our analysis bridges the gap between system-level optimization and the strategic dimensions of NRS development in news organizations. Drawing upon 22 interviews with representatives from four profit-oriented news organizations, we examine how different organizational stakeholders conceptualize the purposes of NRSs (RQ1) within their respective institutional role (i.e., journalistic, commercial or technological). Furthermore, we explore which evaluation metrics are currently in use or considered valuable (RQ2) and how these align with the aforementioned organizational objectives. Finally, we analyze the practical challenges of integrating NRSs as part of a broader digital strategy (RQ3), focusing on both computational and organizational difficulties.

We contribute four key insights to the body of academic knowledge on NRSs. First, engagement serves as the primary goal across stakeholder groups. However, its conceptualization and measurement differ based on stakeholder logics. While business teams associate engagement with conversion and retention, newsrooms frame engagement in terms of content diversity and journalistic integrity. Secondly, there is a growing call for more informative metrics that capture values central to the news domain. Our study shows that while industry still relies heavily on simple, accuracy-based metrics, stakeholders increasingly recognize their limitations. There are efforts to develop new metrics; however, to date, metrics are combined to estimate NRS performance at a more aggregate level. Thirdly, we highlight that the increasing availability of data has resulted in a **jungle** of metrics. Despite the benefits of data analytics, the growing reliance on tools for tracking audiences and gaining content performance insights has not necessarily led to a better understanding of NRS performance. Instead, decision-makers struggle to extract clear insights into the systems' impact on KPIs and call for limiting the number of metrics in their dashboards. Finally, we find that **NRSs are** not viewed in isolation; rather they are one component of a personalization project that is part of a broader strategic shift to digital-first distribution. Stakeholders under the journalistic and commercial logic focus on how all personalization efforts combined align with their overarching organizational goals, such as content distribution and audience monetization.

This is evidenced by the dominance of traditional metrics such as CTR, attention time, and page views, which are primarily valued for their ability to approximate KPIs rather than for evaluating the NRS performance on its own.

In conclusion, our study contributes a more contextualized understanding of NRSs in practice, emphasizing the need for future research to bridge the gap between academic evaluation metrics and the practical metrics required for informed decision-making in the news industry. Our findings show that bridging this gap requires more than just refining evaluation techniques; it also demands a more comprehensive focus on organizational structures, cross-departmental collaboration, and decision-making processes that influence how NRSs are implemented. Without addressing these underlying institutional dynamics, such as competing stakeholder priorities, the operational constraints of technology teams, and the strategic embedding of NRSs within business models, efforts to improve evaluation frameworks risk remaining disconnected from industry realities.

Acknowledgments

We would like to express our gratitude to the participants in this study for their time and willingness to share their experiences and insights. Their openness and flexibility were crucial for conducting this research. We also appreciate constructive feedback from colleagues who helped refine our analysis and reviewers whose comments have strengthened our work. We declare no conflicts of interest.

References

[1] Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnodebski, and Luiz Pizzato. 2020. Multistakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction* 30, 1 (2020), 127–158.

[2] Sinan Aral and Paramveer S. Dhillon. 2021. Digital paywall design: Implications for content demand and subscriptions. *Management Science* 67, 4 (2021), 2381–2402.

[3] Ioannis Arapakis, Mounia Lalmas, B. Barla Cambazoglu, Mari-Carmen Marcos, and Joemon M. Jose. 2014. User engagement in online News: Under the scope of sentiment, interest, affect, and gaze. *Journal of the Association for Information Science and Technology* 65, 10 (2014), 1988–2005.

[4] Mojtaba Barari, Mitchell Ross, Sara Thaichon, and Jiraporn Surachartkumtonkun. 2021. A meta-analysis of customer engagement behaviour. *International Journal of Consumer Studies* 45, 4 (2021), 457–477.

[5] 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.

[6] 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 Transactions on Recommender Systems* 2, 3 (2024), 1–40.

[7] Sina Blassnig, Edina Strikovic, Eliza Mitova, Aleksandra Urman, Anikó Hannák, Claes de Vreese, and Frank Esser. 2023. A Balancing Act: How media professionals perceive the implementation of news recommender systems. *Digital Journalism* 13, 4 (2023), 1–23.

[8] Balázs Bodó. 2019. Selling News to Audiences–A Qualitative Inquiry into the Emerging Logics of Algorithmic News Personalization in European Quality News Media. *Digital Journalism* 7, 8 (2019), 1054–1075.

[9] Udo Bomnüter, Nele Hansen, Michael Beuthner, Ulf König, and Thorsten Hennig-Thurau. 2024. More important than ever before? Assessing readers' willingness to pay for local news as a constituent for sustainable business models. *Journal of Media Business Studies* 21, 2 (2024), 109–132.

[10] Marcel Broersma. 2019. Audience engagement. *The International Encyclopedia of Journalism Studies* 1, 6 (2019), 10–1002.

[11] Michael Brüggemann, Sven Engesser, Florin Büchel, Edda Humprecht, and Laia Castro. 2014. Hallin and Mancini revisited: Four empirical types of Western media systems. *Journal of communication* 64, 6 (2014), 1037–1065.

[12] Matt Carlson and Seth C. Lewis. 2019. Boundary work. In *Proceedings of the Handbook of Journalism Studies*. Routledge, 123–135.

[13] Juliet M. Corbin and Anselm L. Strauss. 2015. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (4th. ed.). Sage Publications, Thousand Oaks, CA.

[14] Rachel Davis Mersey, Edward C. Malthouse, and Bobby J. Calder. 2010. Engagement with online media. *Journal of Media Business Studies* 7, 2 (2010), 39–56.
[15] Heidar Davoudi, Morteza Zihayat, and Aijun An. 2017. Time-aware subscription prediction model for user acquisition in digital news media. In *Proceedings of the 2017 SIAM International Conference on Data Mining*. SIAM, 135–143.
[16] Karen Donders, Gunn Enli, Tim Raats, and Trine Syvertsen. 2018. Digitisation, internationalisation, and changing business models in local media markets: An analysis of commercial media's perceptions on challenges ahead. *Journal of Media Business Studies* 15, 2 (2018), 89–107.
[17] Mats Ekström, Amanda Ramsälv, and Oscar Westlund. 2022. Data-driven news work culture: Reconciling tensions in epistemic values and practices of news journalism. *Journalism* 23, 4 (2022), 755–772.
[18] Sarah Eskens, Natali Helberger, and Judith Moeller. 2017. Challenged by news personalisation: five perspectives on the right to receive information. *Journal of Media Law* 9, 2 (2017), 259–284.
[19] Tom Evens and Kristin Van Damme. 2016. Consumers' willingness to share personal data: Implications for newspapers' business models. *International journal on media management* 18, 1 (2016), 25–41.
[20] Zeshan Fayyaz, Mahsa Ebrahimian, Dina Nawara, Ahmed Ibrahim, and Rasha Kashef. 2020. Recommendation systems: Algorithms, challenges, metrics, and business opportunities. *applied sciences* 10, 21 (2020), 7748.
[21] Elizabeth Fernandes, Sérgio Moro, and Paulo Cortez. 2024. A data-driven approach to improve online consumer subscriptions by combining data visualization and machine learning methods. *International Journal of Consumer Studies* 48, 2 (2024), e13030.
[22] Raul Ferrer-Conill and Edson C. Tandoc Jr. 2018. The audience-oriented editor: Making sense of the audience in the newsroom. *Digital Journalism* 6, 4 (2018), 436–453.
[23] Richard Fletcher and Rasmus Kleis Nielsen. 2017. Paying for online news: A comparative analysis of six countries. *Digital Journalism* 5, 9 (2017), 1173–1191.
[24] Anne Galletta and William E. Cross. 2013. *Mastering the Semi-structured Interview and Beyond: From Research Design to Analysis and Publication*. NYU Press.
[25] Matthew H. Goldberg and Abel Gustafson. 2023. A framework for understanding the effects of strategic communication campaigns. *International Journal of Strategic Communication* 17, 1 (2023), 1–20.
[26] Tim Groot Kormelink and Irene Costera Meijer. 2020. A user perspective on time spent: Temporal experiences of everyday news use. *Journalism Studies* 21, 2 (2020), 271–286.
[27] Brian Halligan and Dharmesh Shah. 2014. *Inbound Marketing, Revised and Updated: Attract, Engage, and Delight Customers Online*. John Wiley & Sons.
[28] Natali Helberger. 2021. On the democratic role of news recommenders. In *Proceedings of the Algorithms, Automation, and News*. Routledge, 14–33.
[29] Jonathan Hendrickx, Annelien Smets, and Pieter Ballon. 2021. News recommender systems and news diversity, two of a kind? A case study from a small media market. *Journalism and Media* 2, 3 (2021), 515–528.
[30] Jennifer Hessler. 2021. From Social Content Ratings to Sentiment Analysis: The cultivation and commodification of affective television engagement. *Participations: Journal of Audience and Reception Studies* 18, 2 (2021), 269–293.
[31] Edda Humprecht, Laia Castro Herrero, Sina Blassnig, Michael Brüggemann, and Sven Engesser. 2022. Media systems in the digital age: An empirical comparison of 30 countries. *Journal of Communication* 72, 2 (2022), 145–164.
[32] Dietmar Jannach and Himan Abdollahpouri. 2023. A survey on multi-objective recommender systems. *Frontiers in Big Data* 6 (2023), 1157899.
[33] Dietmar Jannach and Gediminas Adomavicius. 2016. Recommendations with a purpose. In *Proceedings of the 10th ACM Conference on Recommender Systems*. 7–10.
[34] Dietmar Jannach and Christine Bauer. 2020. Escaping the McNamara fallacy: Towards more impactful recommender systems research. *Ai Magazine* 41, 4 (2020), 79–95.
[35] Dietmar Jannach and Michael Jugovac. 2019. Measuring the business value of recommender systems. *ACM Transactions on Management Information Systems* 10, 4 (2019), 1–23.
[36] Olivier Jeunen, Shubham Baweja, Neeti Pokharna, and Aleksei Ustimenko. 2024. Powerful A/B-Testing metrics and where to find them. In *Proceedings of the 18th ACM Conference on Recommender Systems*. 816–818.
[37] Marius Kaminskas and Derek Bridge. 2016. Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems* 7, 1 (2016), 1–42.
[38] Mozhgan Karimi, Dietmar Jannach, and Michael Jugovac. 2018. News recommender systems–Survey and roads ahead. *Information Processing and Management* 54, 6 (2018), 1203–1227.
[39] Su Jung Kim, Yayu Zhou, Edward C. Malthouse, and Yasaman Kamyab Hessary. 2024. In search for an audience-supported business model for local newspapers: findings from clickstream and subscriber data. *Digital journalism* 12, 9 (2024), 1283–1303.

[40] Bart P. Knijnenburg and Martijn C. Willemsen. 2015. Evaluating recommender systems with user experiments. In *Proceedings of the Recommender systems handbook*. Springer, 309–352.
[41] Tim Groot Kormelink and Irene Costera Meijer. 2018. What clicks actually mean: Exploring digital news user practices. *Journalism* 19, 5 (2018), 668–683.
[42] Allie Kosterich. 2020. Managing news nerds: Strategizing about institutional change in the news industry. *Journal of Media Business Studies* 17, 1 (2020), 51–68.
[43] Isabelle Krebs and Juliane A. Lischka. 2019. Is audience engagement worth the buzz? The value of audience engagement, comment reading, and content for online news brands. *Journalism* 20, 6 (2019), 714–732.
[44] Dmitry Lagun and Mounia Lalmas. 2016. Understanding user attention and engagement in online news reading. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining*. 113–122.
[45] Juliane A Lischka. 2020. Fluid institutional logics in digital journalism. *Journal of Media Business Studies* 17, 2 (2020), 113–131.
[46] Wilson Lowrey. 2018. Journalism as institution. *Handbooks of Communication Science* 19 (2018), 125–148.
[47] Cristian-Ramón Marín-Sanchiz, José María Valero-Pastor, Miguel Carvajal, and Félix Arias-Robles. 2023. Mapping the use of artificial intelligence for the optimization of paywalls in the news media industry: How Firms are taking advantage of machine learning and related technologies to increase reader revenue. In *Proceedings of the Digital Disruption and Media Transformation: How Technological Innovation Shapes the Future of Communication*. Springer, 153–162.
[48] Francisco J. Martin. 2009. Recsys'09 industrial keynote: top 10 lessons learned developing deploying and operating real-world recommender systems. In *Proceedings of the 3rd ACM Conference on Recommender Systems* (New York, New York, USA). Association for Computing Machinery, New York, NY, USA, 1–2. DOI:https://doi.org/10.1145/1639714.1639715
[49] 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.
[50] Marko Milosavljević and Igor Vobič. 2019. Human still in the loop: Editors reconsider the ideals of professional journalism through automation. *Digital Journalism* 7, 8 (2019), 1098–1116.
[51] 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.
[52] Eliza Mitova, Sina Blassnig, Edina Strikovic, Aleksandra Urman, Aniko Hannak, Claes H. de Vreese, and Frank Esser. 2023. News recommender systems: A programmatic research review. *Annals of the International Communication Association* 47, 1 (2023), 84–113.
[53] Lynge Asbjørn Møller. 2022. Between personal and public interest: How algorithmic news recommendation reconciles with journalism as an ideology. *Digital journalism* 10, 10 (2022), 1794–1812.
[54] 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.
[55] Lynge Asbjørn Møller. 2023. Designing algorithmic editors: how newspapers embed and encode journalistic values into news recommender systems. *Digital Journalism* 12, 7 (2023), 926–944.
[56] Tai Neilson and Timothy A. Gibson. 2022. Social media editors and the audience funnel: Tensions between commercial pressures and professional norms in the data-saturated newsroom. *Digital Journalism* 10, 4 (2022), 556–578.
[57] Jacob L. Nelson. 2021. The next media regime: The pursuit of 'audience engagement'in journalism. *Journalism* 22, 9 (2021), 2350–2367.
[58] Ragnhild Kristine Olsen, Aske Kammer, and Mona Kristin Solvoll. 2020. Paywalls' impact on local news websites' traffic and their civic and business implications. *Journalism Studies* 21, 2 (2020), 197–216.
[59] Özlem Özgöbek, Jon Atle Gulla, and Riza Cenk Erdur. 2014. A survey on challenges and methods in news recommendation. In *Proceedings of the International Conference on Web Information Systems and Technologies*. Retrieved from https://api.semanticscholar.org/CorpusID:19984721
[60] Lawrence A. Palinkas, Sarah M. Horwitz, Carla A Green, Jennifer P. Wisdom, Naihua Duan, and Kimberly Hoagwood. 2015. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research* 42 (2015), 533–544.
[61] Charlie Parker, Sam Scott, and Alistair Geddes. 2019. Snowball sampling. In *SAGE Research Methods Foundations* (2019), 3–12.
[62] Vincent C. Peña, Edward C. Malthouse, and Rachel Davis Mersey. 2023. Churning off the news: An analysis of newspaper subscriber churn across digital devices. *Newspaper Research Journal* 44, 2 (2023), 190–205.
[63] Robert G. Picard and Loris Russi. 2013. Comparing media markets. In *Proceedings of the Handbook of Comparative Communication Research*. Routledge, 234–248.

[64] Payam Pourashraf and Bamshad Mobasher. 2022. Using user's local context to support local news. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. 359–365.
[65] Heritiana Ranaivoson, Annelien Smets, and Pieter Ballon. 2024. 15 Challenges and opportunities for recommender systems in media markets. *De Gruyter Handbook of Media Economics* 15 (2024), 215–227.
[66] Shaina Raza and Chen Ding. 2022. News recommender system: a review of recent progress, challenges, and opportunities. *Artificial Intelligence Review* 55, 1 (2022), 749–800.
[67] Jorge Remondes, Paulo Madeira, and Carlos Alves. 2024. *Connecting with Consumers Through Effective Personalization and Programmatic Advertising*. IGI Global.
[68] Anna Schjøtt Hansen and Jannie Møller Hartley. 2023. Designing what's news: An ethnography of a personalization algorithm and the data-driven (re) assembling of the news. *Digital Journalism* 11, 6 (2023), 924–942.
[69] Dhaval Shah, Pramod Koneru, Parth Shah, and Rohit Parimi. 2016. News Recommendations at scale at Bloomberg Media: Challenges and Approaches. In *Proceedings of the 10th ACM Conference on Recommender Systems (Boston, Massachusetts, USA)*. Association for Computing Machinery, New York, NY, USA, 369. DOI:https://doi.org/10.1145/2959100.2959118
[70] Thiago Silveira, Min Zhang, Xiao Lin, Yiqun Liu, and Shaoping Ma. 2019. How good your recommender system is? A survey on evaluations in recommendation. *International Journal of Machine Learning and Cybernetics* 10, 5 (2019), 813–831.
[71] 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.
[72] Alain Starke, Sanne Vrijenhoek, Lien Michiels, Johannes Kruse, and Nava Tintarev. 2024. NORMalize 2024: The second workshop on normative design and evaluation of recommender systems. In *Proceedings of the 18th ACM Conference on Recommender Systems*. 1242–1244.
[73] Steen Steensen, Raul Ferrer-Conill, and Chris Peters. 2020. (Against a) theory of audience engagement with news. *Journalism Studies* 21, 12 (2020), 1662–1680.
[74] Jakob Svensson. 2023. Logics, tensions and negotiations in the everyday life of a news-ranking algorithm. *Journalism* 24, 7 (2023), 1518–1535.
[75] 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.
[76] Edson C. Tandoc Jr and Tim P. Vos. 2016. The journalist is marketing the news: Social media in the gatekeeping process. *Journalism Practice* 10, 8 (2016), 950–966.
[77] Iddo Tavory and Stefan Timmermans. 2014. *Abductive Analysis: Theorizing Qualitative Research*. University of Chicago Press.
[78] Patricia H. Thornton and William Ocasio. 2008. Institutional logics. *The Sage Handbook of Organizational Institutionalism* 840, 2008 (2008), 99–128.
[79] Chun-Hua Tsai and Peter Brusilovsky. 2021. The effects of controllability and explainability in a social recommender system. *User Modeling and User-Adapted Interaction* 31, 3 (2021), 591–627.
[80] Leo Van Audenhove and Karen Donders. 2019. Talking to people III: Expert interviews and elite interviews. In *The Palgrave Handbook of Methods for Media Policy Research* (2019), 179–197.
[81] Hilde Van den Bulck and Hallvard Moe. 2018. Public service media, universality and personalisation through algorithms: mapping strategies and exploring dilemmas. *Media, Culture and Society* 40, 6 (2018), 875–892.
[82] Karin van Es and Dennis Nguyen. 2024. Balancing Needs and Values: A Multi-Stakeholder Examination of Algorithmic News Recommenders in the Netherlands. *Journalism Practice* (2024), 1–19.
[83] 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*. 999–1003.
[84] Hanne Vandenbroucke and Annelien Smets. 2025. From experimentation to strategy: the ambidextrous balancing act of developing news recommender systems. *Journal of Media Business Studies* (2025), 1–28.
[85] Henk W. Volberda, Saeed Khanagha, Charles Baden-Fuller, Oli R. Mihalache, and Julian Birkinshaw. 2021. Strategizing in a digital world: Overcoming cognitive barriers, reconfiguring routines and introducing new organizational forms. *Long Range Planning* 54, 5 (2021), 102110.
[86] M. Bjørn von Rimscha and Robin Riemann. 2024. 13 Developments in advertising markets and their effects on media companies. *De Gruyter Handbook of Media Economics* 13 (2024), 185–200.
[87] Sanne Vrijenhoek. 2023. Do you MIND? Reflections on the MIND dataset for research on diversity in news recommendations. In *Proceedings of the International Workshop on Algorithmic Bias in Search and Recommendation*. Springer, 147–154.
[88] 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*. 208–219.

[89] Sanne Vrijenhoek, Savvina Daniil, Jorden Sandel, and Laura Hollink. 2024. Diversity of What? On the Different Conceptualizations of Diversity in Recommender Systems. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 573–584. arXiv: 2405.02026. Retrieved from https://arxiv.org/abs/2405.02026

[90] 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*. 173–183.

[91] Samia Nawaz Yousafzai, Hooria Shahbaz, Armughan Ali, Amreen Qamar, Inzamam Mashood Nasir, Sara Tehsin, and Robertas Damaševičius. 2024. X-News dataset for online news categorization. *International Journal of Intelligent Computing and Cybernetics* 17, 4 (2024), 737–758.

[92] Rodrigo Zamith. 2018. On metrics-driven homepages: Assessing the relationship between popularity and prominence. *Journalism Studies* 19, 8 (2018), 1116–1137.

[93] Yayu Zhou, Bobby J. Calder, Edward C. Malthouse, and Yasaman Kamyab Hessary. 2022. Not all clicks are equal: detecting engagement with digital content. *Journal of Media Business Studies* 19, 2 (2022), 90–107.

Appendices

A. Topic Guide

Table 4. General Section for all Respondents

Topic	Guiding Questions
Implementation of NRS: What?	– What specific NRS is utilized in the news organization? – Can you give an overview of the NRS that is utilized right now? How was it developed? – What goals is the current NRS designed for? – To what extent does the current NRS design align with your goals?
Multi-objective NRS design [Jannach and Bauer, 2020]	– Can you describe the different values you would wish to see reflected in the NRS design? – What are the primary objectives/purposes according to you for implementing RS and personalization? – To what extent are certain objectives/purposes prioritized when discussing the NRS design? – Do you experience any differences in the goals of the others involved? – Can you describe how you perceive those different values and perspectives central to the [two other logics] related to NRS?
Implementation of NRS: Who? [Smets et al., 2022]	– Can you identify the different teams in the organizations that are involved in the decision-making process? – Who are the different people you negotiate with about the NRS, and how often are you in contact? – Who are other key stakeholders involved in the decision-making processes and the development of RS?
Evaluation of NRS	– What metrics are used to evaluate the performance of the NRS currently? – Can you reflect on the correspondence between the evaluation metrics and the goals you aim at achieving with NRS? – How many different dashboards are used to evaluate the NRS performance? Which metrics are included in those dashboards? – Who sees those dashboards? – Are there any metrics in those dashboards that don't match with your goals? – Are there any other metrics you would like to add to the dashboards?

Table 5. Logic-specific Section [Mitova et al., 2023]

Topic	**Guiding Questions**
Journalistic logic	– To what extent are the following values considered when designing NRS: accuracy, impartiality, discoverability, diversity, serendipity? – How are journalists dealing with NRS in their day-to-day practices? – How do journalists deal with the metrics and data insights in dashboards? – Can you reflect on the impact of data insights, metrics, and dashboards on news production and distribution? – How do you balance data insights with journalistic independence and safeguarding the role of gatekeeping and agenda-setting?
Commercial logic	– To what extent are the following values considered when designing NRS: accuracy, diversity, transparency? – How does your team deal with the metrics and data insights in dashboards? – Has the day-to-day practice in your team changed since the introduction of NRS and personalization? – How does your team perceive the value of NRS and personalization?
Tech logic	– To what extent are you balancing accuracy and beyond-accuracy values when designing NRS? – Have you experienced any difficulties in the development of NRS? – Can you share how the day-to-day practices in your team proceed, specifically for the NRS and personalization progress? – How does your team perceive the value of NRS and personalization?

Table 6. Strategy Section for all Respondents

Topic	Guiding Questions
Digital strategy [Volberda et al., 2021]	**Pillar 1: Cognitive barriers** – What difficulties (concerns, resistance, or hesitancies) toward the implementation of NRS have you experienced in your own team? In other teams? – What concerns, difficulties, or tensions were raised with the implementation of NRS between different teams? – How have these issues been managed and resolved? – To what extent do these concerns influence decision-making around NRS? **Pillar 2: Routines and Workflows** – Has your team been affected by the implementation of NRS? – Have other teams been impacted by the implementation of NRS? – What new roles have emerged within news organizations (including "bridging" roles)? – How has collaboration between editorial teams, technical departments, and B2C changed? – Have any structural changes occurred within [own team] or other departments to facilitate implementation of NRS? **Pillar 3: Organizational dynamics** – Are there any evident shifts in power or management structures within the organization attributed to the implementation of NRS? – How has the adoption of NRS (and other personalization efforts) introduced new working methodologies/workflows?
User perspective	– How do you perceive the user perception of NRS? – To what extent is transparency towards users about algorithmic curation considered in NRS discussions? – When referring to a favourable user experience: how would you measure that? – When referring to "being relevant for the user": how would you measure that? – When referring to meaningful or "good" engagement: how would you measure that?
Future work	– How do you expect NRS and personalization to develop in the coming year? – What are your predictions for NRS and personalization in news distribution in the next 5 years? – In your ideal world: how would personalization be implemented in the news flow?

B. Coding Diagram

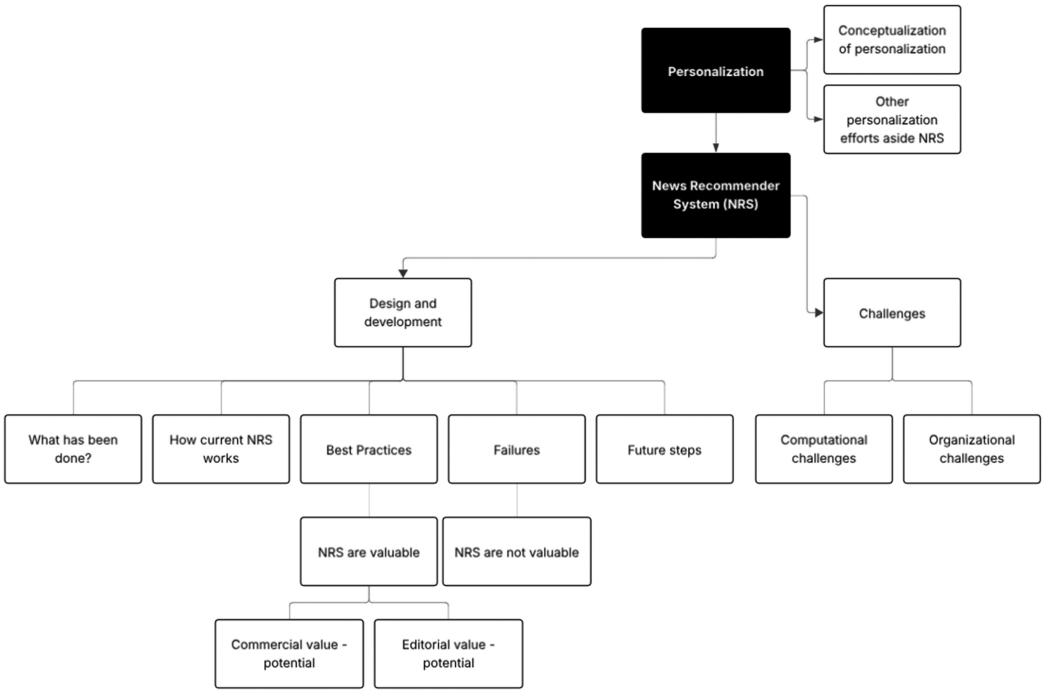


Fig. 1. Coding diagram part 1.

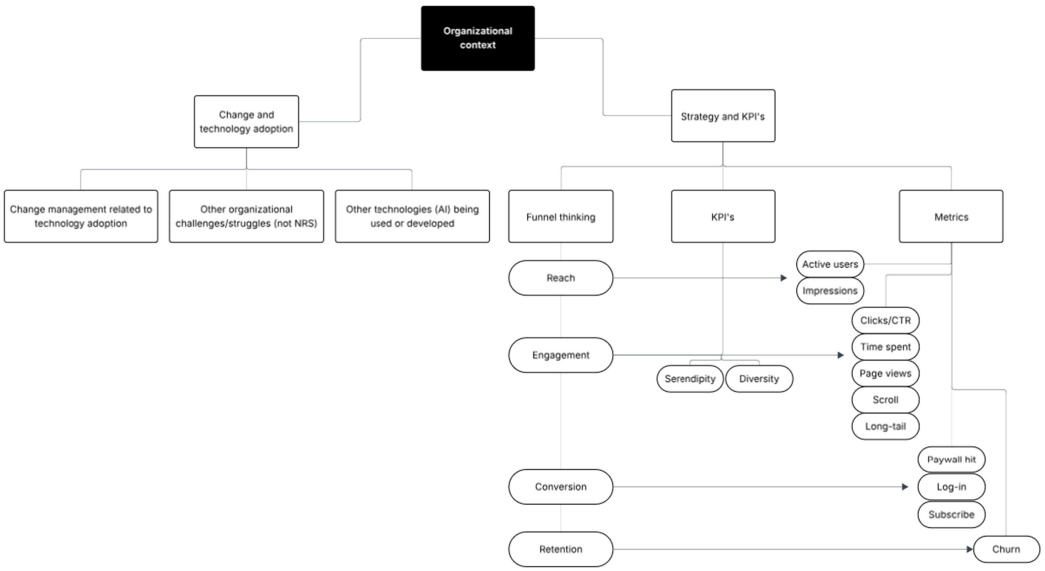


Fig. 2. Coding diagram part 2.

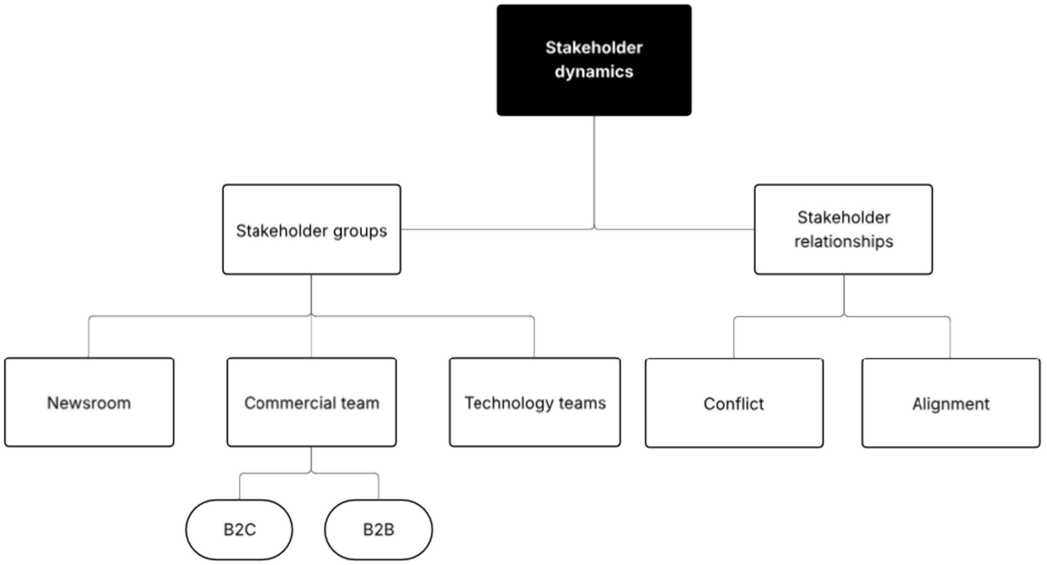


Fig. 3. Coding diagram part 3.

Received 14 March 2025; revised 11 November 2025; accepted 16 November 2025