Recommending news in traditional media companies

Jon Atle Gulla[1] | **Rolf Dyrnes Svendsen[2]** | **Lemei Zhang[1]** | **Agnes Stenbom[3]** |

Jørgen Frøland[4]

[1] Department of Computer Science, NTNU, Norway

[2] NxtMedia Lab, Norway

[3] Schibsted, Sweden

[4] Polaris Media, Norway

Correspondence
Jon Atle Gulla, Department of Computer Science, NTNU, Norway.
Email: jon.atle.gulla@ntnu.no

Abstract

The adoption of recommender systems in online news personalization has made it possible to tailor the news stream to the individual interests of each reader. Previous research on commercial recommender systems has emphasized their use in large-scale media houses and technology companies, and real-world experiments indicate substantial improvements of click rates and user satisfaction. It is less understood how smaller media houses are coping with this new technology, how the technology affects their business models, their editorial processes, and their news production in general. Here we report on the experiences from numerous Scandinavian media houses that have experimented with various recommender strategies and streamlined their news production to provide personalized news experiences. In addition to influencing the content and style of news stories and the working environment of journalists, the news recommender systems have been part of a profound digital transformation of the whole media industry. Interestingly, many media houses have found it undesirable to automate the entire recommendation process and look for approaches that combine automatic recommendations with editorial choices.

HEAD

Most news organizations today use algorithmically personalized news services on their news sites. Even though the basic technology is more than 20 years old (Negroponte 1995), there are fundamental technological challenges that have prevented widespread adoption until recent years. With the technological advances in Artificial Intelligence (AI) and Big Data, however, news organizations are now able to personalize news from massive news streams to the interests of each individual reader. This has been important to keep the news outlets relevant to readers as the World Wide Web has increased the depth and breadth of media content. According to the Pew Research Center, about 82% of US adults in 2019 read news online on

a regular basis, and about 57% used mobile devices as their primary news device (Walker 2019). Similarly, a Statista survey of 28 European countries in 2017 showed that 61% of European news consumers read news online (Statista 2019). Scandinavians were early adopters of online news services, as witnessed by the 90% of Norwegians and 85% of Swedes that consumed news on the Internet.

News personalization means that the media outlet is adapting its news content and presentation to individual users' known or inferred preferences. The underlying technological solution is often referred to as a *news recommender system*, which is normally embedded into a comprehensive media platform that integrates journalistic work and decisions with large-scale information processing capabilities.

Early personalization attempts allowed users explicitly to indicate their preferences and interests and filter news stories accordingly. This was done with newspaper newsletters, email newsletters, RSS feeds and SMS messages in the early 2000s, and it was also used to arrange the layout and content of simple home pages and choose among particular news feeds (Thurman 2019). Modern techniques rely on algorithms to monitor and analyze user behavior and context, interpret news stories and infer the users' preferences directly from their behavior. For example, they monitor how much time the reader spends on a page, how he/she scrolls down the news page, and where and when the reader shows interests in which topics. Big Data architectures and machine learning techniques have allowed the systems to collect vast amounts of data, build mathematical representations of users and stories, and predict each individual user's interest in each available news story. It may seem that news personalization has become a game of numbers that have made editors' and users' own choices unnecessary and maybe even superfluous. News personalization is thus well suited for global large-scale news outlets, but recent experiments indicate that the technology may also be useful in small markets like Scandinavia, where user bases are smaller and content production is limited.

In spite of the advanced algorithms and the fine granularity of user data, there are many open questions about news personalization's effectiveness and desirability. Recommender systems try to formalize very subjective notions like editorial decisions, serendipity, and social relevance, and describe human behavior in quantifiable terms (Helberger, Karppinen, and D'Acunto 2018). Media outlets' roles and responsibilities in society may extend beyond just satisfying users' immediate demands. Even if the technology can predict user preferences, it is not clear how newsrooms should make use of these predictions and reconcile users' preferences with editorial judgment and journalistic standards (Zamith 2018). Whereas users may prefer new articles that confirm what they have just read, editors may find it more useful from a long-term perspective to recommend high-quality articles that challenge the users' views or bring in new perspectives. Also, there is an intrinsic relation between news personalization and targeted digital advertising that may lead to conflicts between commercial and editorial considerations (Bodó et al. 2019). Even if someone is reading articles about popular ski resorts, he/she may not want to be exposed to alpine ski ads on every article page.

In this article we first discuss the news organizations' online business models and their initial interests in recommender systems. The various technologies used in news personalization are then presented, followed by an analysis of the news domain and its suitability for algorithmic recommendations. News personalization is intrinsically very complex due to the whole dynamics of the media industry as well as stakeholders' conflicting goals. After a discussion of personalization services among medium-sized Scandinavian newspapers, we turn our attention to two aspects that have received considerable attention in recent years. First, user privacy issues and new regulations are gradually forcing the recommender systems to be more transparent and controlled by their users. Second, the news organizations may need to consider potential long-term effects of personalized news like filter bubbles, echo chambers and polarization of readers. From a technology point of view, news personalization can be regarded as one of many AI-driven applications that continue to transform the media industry. In the last section we take a broader look at how artificial intelligence affects not only news organizations' strategies and business models, but also their journalists and the whole news production process.

CHANGING BUSINESS MODELS

Historically, media houses have published news and feature articles that subscribing or non-subscribing readers find interesting enough to pay for. Since a newspaper provides a channel to a large number of users, it has attracted advertisers that are willing to pay for placing ads in the paper. With the introduction of online news services, the dynamics of this relationship between media houses, readers and advertisers are affected in various ways, and new companies have emerged that complicate the revenue streams in the sector. The transition from print to digital started almost 30 years ago, and companies have over the years re-assessed their business models as online news and extracted user data are leveraged to increase reader engagement and create new sources of revenues.

Traditionally, newspapers cover their expenses with a mixture of news stand and shop sales, subscription fees, and advertising revenues:

- Paper copies. Newspapers have traditionally been published as mass production of texts and visuals on low-grade printed paper. Sales of single copies from newsstands and shops are common, though there are also ad-supported printed newspapers that are distributed for free.
- Subscriptions. The subscription business model is based on customers paying a recurring price at regular intervals for access to a product. The subscription model is popular in a myriad of businesses and has proven useful in online publishing. Public subscriptions have traditionally been considered a guarantee of independence

from advertisers who may try to influence editorial governance.

- Ads. Before the introduction of online news, a newspaper typically generated 70%–80% of its revenues from advertising and the remainder from sales and subscriptions (Mensing 2007). Print ads are ads that run in local or national, daily or weekly news publications. The professional sales ads were later supplied with classified ads that are smaller ads for all kinds of direct sales between people and also include professional sales of jobs, properties and autos.

The rise of data-supported online publishing, search engines, digital platforms, and social media have disrupted the traditional newspaper business and deeply affected their business models. There are now almost no news outlets left that do not maintain a digital presence.

Online news consumption and production. In the old days, news stories were produced in a linear fashion. Reporters conceived a story idea through laborious research and data analysis, cultivated and interviewed sources, packaged the information into a draft, and worked with editors to finalize their one-off story. Only after the story was published much later were the readers able to access to the information (Marconi 2020). With online reporting, news gathering and content processing are supported by data analytics and automated workflows, and distribution is channeled to desktops and mobile units in various formats. The information flow is much higher, and there is a two-way mutually dependent interaction between news outlets and audience.

Free access vs paywalls. Early adopters of online publishing were concerned about their positioning in a new market and prioritized free access to news content as a software service. The objective was to maximize reach with large volumes of content and increase their revenues with automated advertising models. However, many advertisers have gradually moved their ads from traditional news sites to Internet portals like Google and Facebook. As the competition for ad revenues grew from large-scale international platforms like Google and smaller Internet sites dedicated to particular topics, many companies started looking for new ways of monetizing their content. Paywalls, which require users to register and log in, have enabled companies both systematically to collect more data about their readers and offer paid content. Soft paywalls may require the users only to register or give them access to a certain number of stories in a certain time frame (metered paywall). A more common variation of the model is the freemium model, in which a first tier of content is free, but access to premium features is limited to subscribers or people who pay. Publishers tend to allow free use of some lower quality content, whereas access to the greater part of their unique production is restricted. Unfortunately, the availability of free news from aggregators and social media has kept rates for paid content low, challenged subscription loyalty, and increased churn. The media companies also suffer from many traditional advertisers that are shifting their business to new media platforms.

Ad-driven business model. Digital traffic in terms of audience reach was a priority for media companies entering the online business. Increased click rates drive ad revenues and often reflect an increasing number of readers. Media companies profit from users visiting their sites, but they can also extract a share when other platforms are visited. A survey by the British ad federation ISBA indicates that a web site typically receives 51% of an ad payment, 27% goes to intermediaries like Google and 15% is impossible to trace (Clement 2020). Digital behavior leaves traces that provide new insight into users' interests and can be used to construct detailed user profiles for advertising purposes. A company like Google has been very successful in this digital marketplace. Google ad income has risen from 28 billion dollars in 2010 to 135 billion dollars in 2019 (Nylen 2020). In comparison, the total global newspaper revenue was estimated to 108 billion dollars in 2019. With the technology companies' dominance in the market, it seems unlikely that ads will be such an important source of revenue for media companies in the future. In the US, for example, the total ad revenues of newspapers have been in steady decline from 67 billion dollars in 2000 to 16.12 billion dollars in 2013 and 11.52 billion in 2019 (Statista 2021).

Converting users. In recent years media companies have turned to digital subscriptions for increasing revenues and funding news production. Disruptive players like Spotify and Netflix have already shown that users may be willing to pay for content in the music and cinema industry. With digital subscription models the quality of journalistic products is important to attract new subscribers and keep them interested in paying for news stories. Central to the model is the tracking of user behavior to understand users' engagement and unlock their interests into paying loyalty. Click rates are still part of this user tracking, though it seems that reading times are as important to fully understand users' true engagement.

NEWS RECOMMENDATION STRATEGIES

Non-personalized/non-contextualized recommender systems are the simplest type of recommendation approaches. They do not take any personal information into account and tend to provide the same recommendation results for all users, at least for a predefined geographical area. The recommendations can either be manually selected by online news providers, based on the popularity of
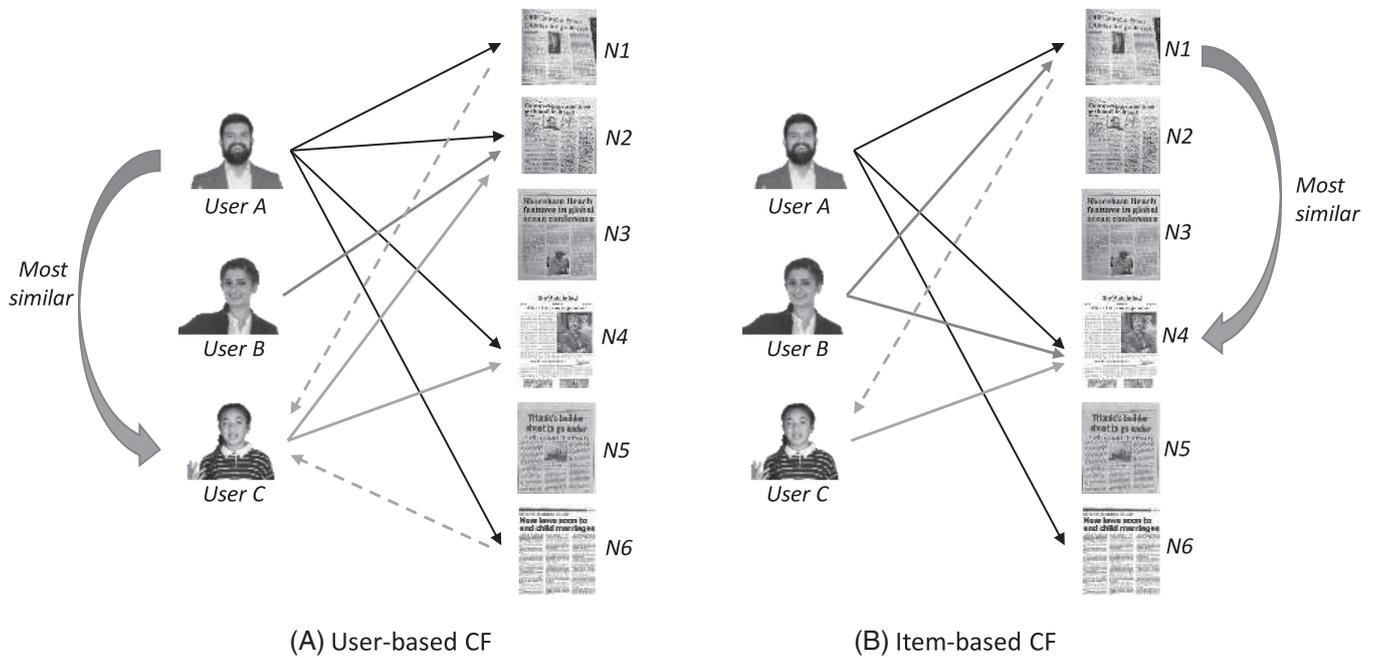
FIGURE 1 User-based and item-based collaborative filtering

items, or the recommendations can be the top-N newly arrived items (Hiralall 2011). In the news domain, a typical example of non-personalized/non-contextualized recommendation is the news curation of front page stories and headlines. Traditional news curation relied solely on expert human editors, who used to select news stories that should be consumed by the readers – a process known as journalistic gatekeeping (Shoemaker, Vos, and Reese 2009). With the proliferation of online news aggregators and social media news feeds, many media organizations have introduced algorithms to reduce manual labor by considering recency, relevance, and diversity (Chakraborty et al. 2017). Recency measures the difference between the curation time and the publish time of the news story, i.e. the age of the story. Intuitively, the importance of news will decrease with longer publication times as older news tend to fade away rapidly. Relevance can be referred to as the importance or the impact of a news story by e.g. crowd-driven measurements of popularity, such as the number of people who read or liked the news. Some news recommender systems also suggest relevant news articles to online consumers based on the similarity of the news articles with the news they are reading. Diversity can be measured by the curated news stories' coverage of different topics or categories. Chakraborty et al. (2018) developed a news recommendation framework, Samar, for editors curating front-page stories while considering recency, relevancy, and diversity in a non-personalized manner.

Collaborative filtering (CF) is one of the most common filtering techniques for recommending news stories based on other users' opinions or properties. It was initially introduced in a recommender system called Tapestry (Goldberg et al. 1992), which was designed to recommend or filter a set of documents for a particular user. The fundamental idea behind CF is group intelligence, which means that people collaborate to help each other perform filtering by recording their reactions to documents they read. For example, if Tom and John both prefer to read a piece of news regarding a football match, the other news stories read by John might be something Tom would like to read since they are likely to have some similar reading preferences. This approach is called user-based CF, since the underlying idea is to compare users' interests and predict based on similar users. Since user *A* is found to be the most similar to user *C* in Figure 1A, the system recommends two new articles to user *C* that user *A* liked to read. These are the articles *N1* and *N6* that are shown with dashed lines in the figure. In item-based CF, the idea is to find similar articles – not on the basis of content – but on the basis of the extent users have liked the articles. In Figure 1B, user *C* has enjoyed reading article *N4*, which is found to be the most similar to article *N1*. The item-based CF systems then recommends article *N1* to user *C*.

CF suffers from several drawbacks in the news domain. First, a problem is often referred to as the first-rater problem (Das et al. 2007; Good et al. 1999) which is problematic for news recommendation since online news services strive to present the most updated information to users in a timely manner. CF-based methods need to wait several hours to collect enough clicks to recommend incoming news stories to users, resulting in undesirable time lags

between breaking news and recommendations. Multi-arm bandit models have however been used with some success to deal with cold-start problems (e.g. Felicio et al. 2017). Liu, Dolan, and Pedersen (2010) observed that CF tends to recommend entertainment news stories to users that never clicked on entertainment stories before due to the popularity of entertainment news stories and the high number of clicks generated from the neighborhood users/items.

Whereas collaborative filtering approaches use only user-item rating matrices to form recommendations, *content-based recommender systems* are designed to recommend items that are described with attributes that carry some similarities with a user profile. If John reads a piece of news related to the US presidential election campaign one day, chances are high that he might want to read other stories with more details about the campaign. In such cases, John's reading history is sufficient to discover meaningful preferences, and recommendations are made by analyzing the content of news stories he has read rather than other users' reading histories. Content-based news recommender systems recommend news stories that are similar to what the users have read in the past and are particularly useful for the most recent news as well as for news stories that are not so popular among other readers.

Even though content-based methods are effective with respect to cold start items and data sparsity problems, there are limitations that hamper its use in news personalization. The novelty and serendipity of content-based systems are relatively low because the methods only look at the user's already read news stories and thereby miss the opportunity to include topics or stories that add something entirely new.

Hybrid approaches combine CF-based and content-based methods and have been used with considerable success in the news domain. If the CF- and content-based models are first built separately, a technique like adaptive weighted average may be used to combine them (Melville and Sindhwani 2011). Other mixing strategies include switch back and forth, feature combination, cascade, meta-level hybrid, and feature augmentation (Athalye 2013; Borges and Lorena 2010; Burke 2002; Claypool et al. 1999). An example of a hybrid news recommender system is the Google News webpage. As an anonymous user, you are shown headlines that are the same for all users in the default area (e.g. the United States) on the top section of the homepage. If you sign in with Google, it also shows personalized news and local news chosen on the basis of personal settings as well as search and reading histories.

With the surge of *deep learning techniques* in collaborative filtering and content-based recommendation, neural network-based recommender systems have attracted increased attention and achieved impressive performances (Raza and Ding, 2020; Zhang et al. 2019). Deep neural networks are efficacious in learning underlying explanatory factors and useful representations from input data. In addition to reducing the efforts of handcrafted feature design, deep learning-based approaches can easily make use of heterogeneous content information like text, images, audio, and even video.

Several recent recommender systems discard long-term user profiles and try to recommend stories solely on the basis of what the users have already read or clicked in their current sessions. Some users may prevent the use of profiles, or the profiles are so general that they are of little value for personalization. Many news outlets also do not track users with their identifiers for a long time period. Even though cookies and browser fingerprints can help provide more historical user data, most users are not subscribers and have only one or two sessions that can be identified and used for recommendation. In a news outlet like Adresseavisen, which is the third biggest news portal in Norway, subscribers take up only about 20% of the users in their logs (Zhang, Liu, and Gulla 2019a). In such cases, it is necessary to take the structure of the ongoing session into account to capture richer information. To this end, session-based recommender systems (SBRS) are proposed to learn user consumption patterns and user preference shifts from one consumption to another. Here, consumption can be referred to as any user-item interaction behavior such as purchasing, visiting, clicking, and reading.

Different from conventional collaborative filtering and content-based methods, SBRS take a session as the basic unit for recommendation and comprehensively explore the information embedded within one session (intra-session) as well as from one to another (inter-session). In the news domain, a session can be regarded as a sequence with multiple news stories read in one event. Recent studies verify the effectiveness of news recommendation in session-based settings. In Zhang et al. (2019b) the authors consider inter- and intra-session relationships for modeling user preferences on news articles. An attention mechanism is adopted to capture the users' main purpose within the current session, like checking out some entertainment news stories or a particular sports event, and a recurrent neural network is employed to predict which stories the user will click on next in a sequential fashion. The work proposed in De Souza Pereira Moreira, Jannach, and Da Cunha (2019) leverages multiple types of contextual information extracted from news stories, such as news textual content, publishing date, category and author, to learn news representations for the deep learning-based recommender system. SBRS can be applied to a wide range of applications due to the availability of session data that is much richer than the data normally required by recommender systems (Wang et al. 2019).
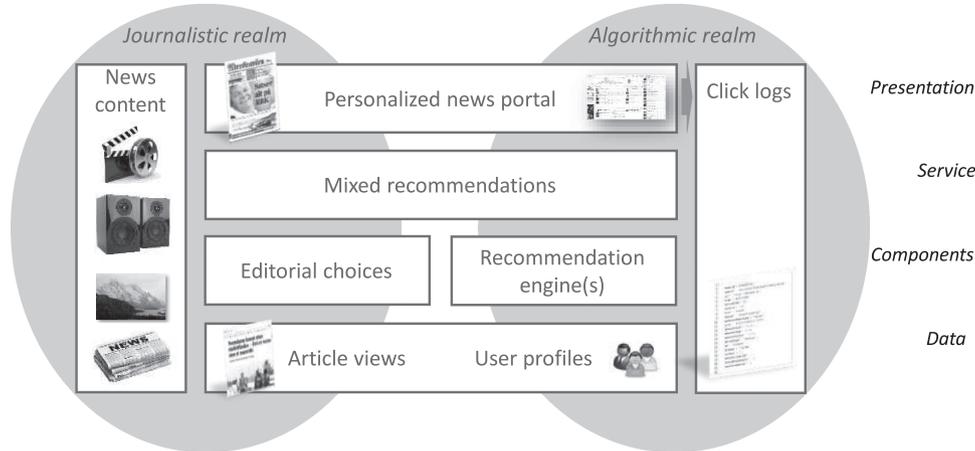
FIGURE 2 News personalization platform

THE RISKS OF AUTOMATIC NEW RECOMMENDATION

Personalized news platforms allow news to be tailored to individual users at scale, but they are also part of a larger machinery that relates to the media's role and responsibilities in society. Editors may still be needed for bringing in relevant news that are not directly related to users immediate reading histories, and they can assess how the stories fit into the news sites' overall presentation and content strategies. Automatic recommendation is hampered by the stories' short life spans, textual variation, and inconsistent quality, as well as by the users' behaviors that are often difficult to analyze in terms of goals or information needs.

While conveying news to interested readers, the news platform serves as a social adhesive that provides a set of common experiences and guards against cultural and political fragmentation. As Sunstein (2001) pointed out, media outlets carry some societal responsibilities that may be influenced by their approach to personalization. On this background it may be too short-sighted to optimize recommendation strategies with respect to prediction accuracy without taking into account the media's impact on society at large. There is no absolute guarantee that automatic recommendations are meaningful or important to users, in spite of all the data available and the sophistication of the models used. In practice, commercial news outlets try to balance the efficiency of automatic recommendations with some editorial considerations. Even though editors do not know the preferences of individual users, they have the expertise and experience to promote stories of importance to society at large, or in general recommend stories of great quality or depth that relate to the present story. We have seen that some media outlets mix a number of editorial recommendations into the list of automatic ones. Other news sites keep editorial and algorithmic recommendations apart and present two lists of recommendations to their readers. In any case, these editorial considerations call for rather complex hybrid recommender systems that need procedures or weighting systems for combining various automatic techniques as well as for balancing the resulting automatic recommendations with recommendations manually associated with news stories or news contexts. Interestingly, a 2016 survey of over 53,000 online news readers in 26 countries suggested that readers would rather have their stories picked automatically than by editors or journalists (Newman 2016).

From the architecture of personalized news outlets in Figure 2, we see how journalistic and algorithmic parts come together in commercial platforms. Whereas the algorithmic recommendation engines are the focus of most research prototypes, commercial systems depend on large-scale real-time monitoring and processing of news stories and user behavior, are integrated with multimedia user interface environments with numerous offerings and tap into journalistic systems for news production and editorial decisions. After the automatic recommendations are generated, editors may choose to combine these with manual suggestions into a mixed list of recommendations. At the presentational layer, there are options about where and how to list recommendations, what they should look like, and to what extent they should be accompanied by graphical elements or other attention-grabbing features. Understanding the interplay between these components and evaluating the effects of personalization are often tedious and error-prone, since it is difficult to isolate the recommendation strategies from other parts affecting their performance.

A particularly sensitive issue in news personalization is recency or timeliness. As opposed to items like books or movies in other domains, news stories have short life spans and are normally only relevant for a few days after publication. Since experiments show that news stories are rarely viewed after 2–3 days, it is tempting to downplay older news by means of time-decay models, graph-based models or popularity-based models (Raza and Ding 2020). However, as shown in Figures 3 and 4, the life span of news stories depends on a number of factors. On the one hand, different media outle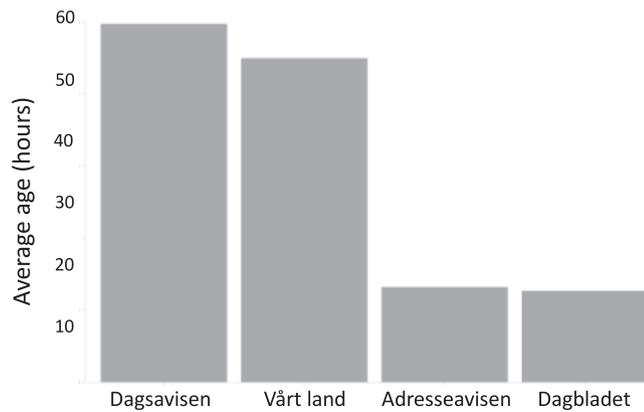ts focus on different types of news and cater to different types of audiences. Figure 3 shows the average age of clicked news stories from four different Norwegian newspapers (Svendsen, Gulla, and Frøland 2019). The analysis is based on user logs from 1 August to 1 November 2014 for all four newspapers. Dagbladet and Adresseavisen are general news sites that publish the latest news on a wide range of topics to a broad audience. On average, news stories on these two sites are read a little bit more than half a day after publication. Stories on Dagsavisen, with its political emphasis, and Vårt land, with its religious perspective, appeal to smaller audiences that sympathize with the newspapers' profiles. Since their stories are accessed over a longer time period with the average click coming more than 2 days after publication, recency does not seem to be such an important issue for all news outlets. On the other hand, also within a news outlet there are substantial differences of life spans from one news category to another. As indicated in Figure 4 from Adresseavisen, health-related articles live almost four times longer than general news stories.

Another complication is the analysis of news content. Unstructured news stories of typically 200 to 300 words may use irony, sarcasm, exaggerations, and similar rhetorical devices to make their content more interesting and convincing. Since movie and book descriptions tend to be structured, factual, and quality-checked, recommendation services at Amazon and Netflix can build fairly
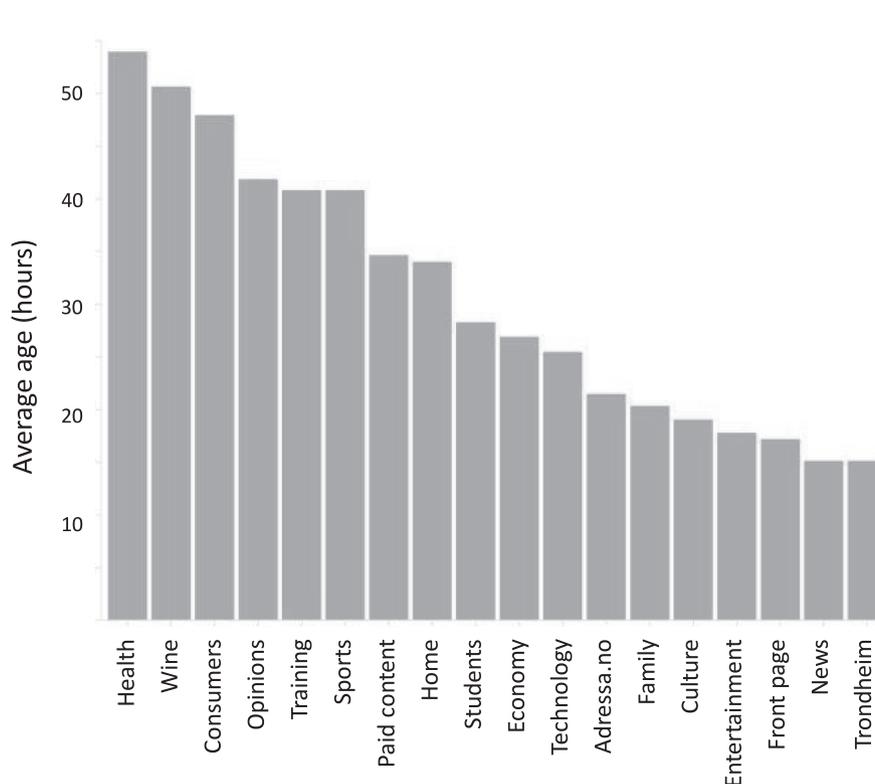
FIGURE 3 Average age of clicked news stories from four Norwegian newspapers

FIGURE 4 Average age of clicked Adresseavisen news stories sorted by news category

accurate content descriptions for their content-based recommendation. In the media sector, however, more sophisticated linguistic techniques are needed, and the final content representations are not likely to be as accurate as in other domains. Consequently, many news recommender systems have resorted to analyzing content in terms of news stories' entities, like person names and geographical locations, which amounts to about 25–30 in a typical Norwegian news story (Gulla et al. 2017). In recent years media outlets have also faced an abundance of fake news or highly biased news. Fact checking organizations help sort out untrustworthy or unbalanced news stories, but there is also a growing number of algorithmic techniques for fake-news detection and media-bias detection (Zhou and Zafarani 2020). The uncertain quality of news stories complicates the recommendation process, as news providers normally want to boost the most trustworthy and balanced stories.

Finally, there are also aspects of news readers that are challenging to deal with in traditional recommender systems. The recommender systems monitor users' click behavior and store relevant properties of their context and behavior as they are browsing the news and reading selected stories. In news personalization, explicit feedback is normally not available and the systems must use less accurate substitutes like reading time to infer user's interest in a story. Also, a user session involves reading around 6–8 stories, which is rather short for an accurate extraction of user's preferences and interests. What makes the situation worse, is that news readers tend to change their interests even within a session and move from one topic to another. In addition to users' long-term profiles, some news recommender systems address this with the introduction of short-term profiles for users' current preferences. Experiments also show that many readers follow certain topical paths when they browse the news, which can be explored used process mining techniques (e.g. Epure et al. 2016). An analysis of Dagbladet's user logs show that their readers are almost four times more likely to move from political to international news than in the other direction. Even though this may be partly explained by the layout of the news site, there is enough variation among the users to suspect that individual user preferences also play a role. Since there is only a 0.7% chance that a reader of international news chooses a sports story afterwards, it may not be very useful to recommend sports news to someone reading about international affairs. A final aspect in user modeling is the influence of users' contexts. It seems that time, users' geographical location and other contextual factors have an effect on the preferences of news readers. Son, Kim, and Park (2013) argue for example that news recommender systems ought to emphasize local news to mobile users. The analysis of logs from four German and Norwegian news sites in Kille, Özgöbek, and Lommatzsch (2018) shows among other things that reading frequencies depend on the time of day, financial news are rarely consulted after 6 pm, health- and family-related news dominate in the evenings, and for unknown reasons Norwegians read news earlier in the morning than German readers.

CASE STUDIES FROM TRADITIONAL MEDIA COMPANIES

Commercial news personalization platforms combine state-of-the-art recommendation strategies with architectures for large-scale data processing. Some commercial news outlets deal with millions of unique visitors that view large amounts of news stories and leave detailed traces for the underlying news recommender systems. These outlets are good candidates for full-scale AI-driven news personalization, and many case studies report on successful deployment among these large international companies. For media companies with smaller user bases, the technology adoption has been less ambitious and the experiences are more mixed.

A full-fledged personalized experience is implemented for example in the Washington Post (Graff 2015). Both content-based recommendation and collaborative filtering are used to recommend stories to more than 52 million unique users per month. Besides extracting prominent entities from the news stories and adding them into user profiles, they also assign stories and users to 130 news categories that are used together with the entities to recommend stories from the relevant category. The hybrid recommender system is part of the larger *Post Recommends* system that has seen a 95% year-over-year increase of click through rates since the introduction of their new personalized news service. Unique visitors were up 65% year-over-year.

Personalization is not only used for news content selection. In 2018 the Wall Street Journal moved away from static paywall models and introduced an individualized paywall on top of their news personalization service (Arns 2018). They observed that the amount of article views needed to convert a reader varied and was often influenced by long-term interests and context. Analyzing the readers' propensity to subscribe from their user logs, they built a recommender system that would find the ideal paywall model for each individual reader. Whereas some readers were exposed to soft paywalls and could read a substantial number of stories before receiving a subscription offer, others met a hard paywall after a few clicks. Together

with Cxense, a company offering personalization services to almost 200 online customers globally with more than 2 billion users, the Wall Street Journal deployed a system with dynamic paywalls that has helped them attract 25% more subscribers.

As the user base grows, the relevance of personalization increases and more sophisticated techniques are introduced. Google News, which had approximately 150 million unique monthly visitors in the US in 2018, aggregates and recommends news from more than 50,000 news sources (News Media Alliance 2019). Advanced hybrid techniques are used with no editorial control and with an emphasis on click rates and advertising revenues. It is interesting to note, however, that personalization is not restricted to large-scale news aggregators and global news outlets, but are also deployed by media companies that have a few thousand subscribers and daily publish a few hundred news stories. This is often the case for Scandinavian news outlets that do not have the resources to build large personalization teams and carry out systematic experiments with a plethora of recommendation strategies. The model has rather been to start out with simplified approaches and add more sophistication as competence and experience grow.

Scandinavian news outlets have often introduced two simplifications to their commercial personalization platforms (Svendsen, Gulla, and Frøland 2019). Rather than personalizing with respect to individual user profiles, many companies recommend to a user segment that typically represents a group of users with some characteristic similarities like *age* and *gender*. Both Amedia, a Norwegian newspaper company with 80 local news outlets and 1.96 million readers, and NRK, the state broadcasting company in Norway, have successfully applied segment-based news personalization on their websites. Another simplification has been to represent news content in terms of a limited number of topical categories and ignore the actual words – or entities – of the stories. NRK, for example, has experimented with a content-based recommender engine, in which both content and users were represented as vectors of 81 news categories like *animals*, *football*, and *international politics*.

In 2017 NRK deployed a new hybrid personalization service for the TV programs available on their web site (Holmstad 2017). With only anonymous users and an archive of more than 100,000 programs, of which 10% make up 90% of user visits, NRK decided to combine collaborative filtering and content-based recommendation in a way that did not require any stored user profiles. For the collaborative filtering part, each program was represented as a vector of 20 dimensions that reflected their usage pattern. The content-based recommender engine extracted prominent phrases from the descriptions of each program that could be used to compute content similarities between programs. When the user had just finished viewing a particular TV program, he/she was recommended new programs that shared some usage characteristics or content characteristics with the current program. Normally at the bottom of a program page NRK would list other programs to watch, and users would select one of them in about 12% of the cases. After replacing this list of programs with recommendations from their personalization engine, users clicked on a recommended program in 20% of the cases. Also, recommended programs were watched to the end 58% more often than non-recommended programs, though it should be added that even with personalization users leave the programs before the end in 97.5% of the cases.

Another interesting personalization experiment was carried out by the iTromsø news outlet for their mobile news readers in 2015 (Svendsen, Gulla, and Frøland 2019). The whole front page was personalized, though editors were given the opportunity to replace the top story if needed and some serendipitous stories were added to the mix. This resembles the multi-armed bandit approach used by for example Cañamares, Redondo, and Castells (2019). An analysis of the three last months before personalization and the 3 months after its introduction showed a 23% increase of unique visitors and a 9% increase of click rates. Table 1, which gives a year-over-year comparison with either three or 5 months, reveal some other interesting observations from iTromsø. Reading times increased by 28% to 28 s for the front page and by 15% to 62 s for the news stories themselves. In the 5 months period after introducing personalization, the number of mobile readers was 20% higher and session times were 9% longer than in the same 5 months period 1 year earlier. Total news consumption for this 5 month period in 2015 was 120,833 h, up 31% from 2014. Since there is a chance, though, that the focus on click rates and reading times have influenced also the choice of events to report from and the style of writing, it is still early to conclude about the long-term effects of automatic news recommendation in iTromsø.

For smaller news outlets it may be interesting to investigate to what extent personalization can revitalize old news content. In 2017 Altaposten first compared click rates when recommending news stories that were up to 2 days old with recommending stories that were up to 4 weeks old (Svendsen et al. 2017). Each configuration was run over 3 weeks and showed substantially increased click rates for both mobile and desktop readers when adding more and older content, and desktop readers with 42% increase seemed most enthusiastic about the old content (see Table 2). A similar experiment was run afterwards, comparing news that were at most 12 weeks old with news that were up to 1 year old. The increased click rates of 23% for mobile users and 37% for desktop users were surprisingly high and may suggest that for some news outlets news stories are not as

TABLE 1 Results from fully personalized front page at iTromsø

Evaluation metrics	One year before personalization	After personalization	Change
Click rates (measured over 3 months)			
Traffic on front page	22,238 clicks/day	24,516 clicks/day	+10%
Reading times (measured over 3 months)			
Front page	22 s	28 s	+28%
News stories	54 s	62 s	+15%
Media consumption (measured over 5 months)			
No. of mobile readers	878,000 users	1,049,000 users	+20%
Session time	6 min 19 s	6 min 55 s	+9%
Total media consumption	92,500 h	120,833 h	+31%

TABLE 2 Recommendations make older content more relevant in Altaposten

Experiments over 3 weeks period with news personalization	Change of click rates	
	Mobile readers	Desktop readers
News data set expanded from 0–2 days old to 0–4 weeks old	+17%	+42%
News data set expanded from 0–12 weeks old to 0–52 weeks old	+23%	+37%

short-lived as has been assumed in the past. Another explanation may simply be that as the users ran out of recent news, they started to explore older ones that were now convenient to access. If that is the case, the immediate increase of click rates may be short-lived and users may gradually get frustrated by the outdated nature of the news storied suggested to them.

Overall it seems that Scandinavian media companies are embracing personalization technologies, and they are carefully testing out different techniques and scenarios. Most personalization experiments on Scandinavian news outlets have had promising results, though the improvements are moderate and it is sometimes difficult to know exactly what causes the improvements. Companies are often not in a position to test out every configuration, and several techniques may be bundled together before A/B tests are carried out.

The experiments above also illustrate Scandinavian newspapers' belief in some editorial control. Personalization technologies and AI in general are intended to support editorial processes and decisions, rather than replacing them. Reading times are often preferred to click through rates, as they seem to be more relevant for user conversion. Leading media companies in Scandinavia like Schibsted, Polaris Media, and Amedia all emphasize the need to increase the number of subscribers using personalization or other technologies. This contrasts with Google News' large-scale ad-driven personalization approach and consti-

tutes what Bodó (2019) refers to as the *"news logic of personalization"*.

USER PRIVACY ISSUES AND IMPLICATIONS FOR NEWS RECOMMENDATION

News outlets collect and analyze vast amounts of personal data to build detailed user profiles for news personalization. The data, which include demographic details, geographical locations, time zones, device types, news stories read, categories, and key words, among many other features, trace the behavior of the users and are stored for later analysis and recommendation. As the user logs grow in size, they provide insight into not only the users interests and preferences, but often also their movements and activities. To some extent the data may even be used to infer sensitive user data about health, ethnicity, sexual orientation and political views. The example in Figure 5 from the Adresseavisen news outlet shows how a simple entry in such a user log may look like. The event and the user are given internal identifiers. The time and date of the event is recorded, and we can see that the user spent 23 s reading the front page of Adresseavisen. The user used Chrome on Windows to access the page from a small town called Verdal in the Nord-Trøndelag county. The log reveals the sequence of web pages that he/she has been reading in this

FIGURE 5   An event extracted from an Adresseavisen user log on the Cxense platform

session. Important to the construction of user profiles, the log also links to the actual news content, which is analyzed separately in terms of entities mentioned and other relevant features.

Surveys show that users are aware of some of the concerns of news personalization. When more than 38,000 people across 19 European countries were asked about the possible negative impact of personalization algorithms, around 46% thought that more personalized news mean that their privacy are at greater risk (Newman 2016). It is interesting to note, though, that the same people found other aspects of news personalization more problematic. About 56% of the respondents worry that they miss out on important information, while 53% think they may miss out on challenging viewpoints. In a German survey about personalization in general, 82% of 1,065 respondents claim that they were concerned about their data privacy, and 57% objected to personalized news feeds on social media (Kozyreva et al. 2020). Despite these concerns, users seem to think that the benefits of news personalization outweigh its risks, and even users that are conscious about user privacy do little to protect their data and easily adopt personalized services (Thurman 2019). When readers of iTromsø were given the chance to turn off personalization, only 0.9% did and only 3.5% cared to have a look at iTromsø's logging of their reading behavior. In comparison, a survey conducted for the British government from 2010 shows that 9% of the 1,012 respondents were actively refusing all Internet cookies, whereas 45% would only accept selected cookies (Lancefield et al. 2011).

With new regulations in place like the EU General Data Protection Regulation (GDPR), users' data protec-

tion rights are strengthened. Since GDPR's journalism provision does not seem to apply to news personalization (Eskens 2019), users may exercise their rights to control the processing of their personal data for personalization. They may withdraw their consent to personalization or change their personal profiles used by the recommender systems. When users take control of their own profiles, news personalization platforms need to be more transparent about their profiles and allow users to manage the way personalization is applied to them.

Today there is a discrepancy between users' worries about user profiling and their overall acceptance of profiling in news recommendation, but also between the legal need for transparency and the closed machine learning models used in practical solutions. Since most users accept the situation with all the data collected and processed in news recommender systems, content providers only slowly open up their recommendation strategies and user profiles for user inspection. News personalization platforms are gradually changing, though the changes reflect new legal frameworks rather than users' expressed concerns.

SOCIETAL IMPACT OF NEWS PERSONALIZATION

Except for informing citizens, the media companies are expected to create a diverse public forum where ideas and opinions in a democratic society can be articulated, encountered, debated, and weighed (Helberger 2019). User engagement is central to this mission, as well as trust and transparency among readers and media companies. News outlets' personalization strategies have gradually changed as the competition from large-scale platform companies has intensified. Click rates, reading times, and similar metrics are still important when new features are deployed and evaluated, though there is an awareness that such metrics may not be able to capture the full complexities of media's roles in society.

The long-term effects of algorithmic news recommendation have been discussed in recent years and are often framed with concepts like *filter bubbles* (Pariser 2011) and *echo chambers* (Sunstein 2001). There is a concern that people on the Internet gravitate towards content and people that reinforce their existing views. Avoiding contradictory views and alternative theories, they see only fragments and are not exposed to views or data that challenge their own perspectives. News personalization may be instrumental in creating these filter bubbles or echo chambers by steering readers towards news stories that are like the stories they have already read and appreciated, or towards stories that have been positively received by other people like them. If this is the case, personalized news outlets have failed in

providing a diverse environment for sharing information and debating important matters in society from different perspectives.

Echo chambers emerge when groups of people choose to connect with each other and exclude and discredit outside sources. Filter bubbles refer to some intellectual isolation that can result from systems separating people from information that disagrees with them. Even though filter bubbles and echo chambers have caught public interest and are sometimes taken for granted, it is difficult to formulate a precise definition of the phenomena and verify the extent of the problem (Weinberger 2017). There are separate communities that share important ideas or identities, but the question is to what extent they are detached from other views. The communities may be polarized, but not necessarily disconnected from society (Brun 2019). According to Dubois and Blank (2018) such a disconnection is unlikely in an online environment where people are regularly faced with things they disagree with and can search and retrieve information from a vast number of news sources. A study from UK, USA, Germany and Spain suggests that users of web search read a balanced number of left-leaning and right-leaning online news stories (Fletcher and Nielsen 2018). News organizations are not so worried about filter bubbles and feel that their personalized news services include appropriate strategies for addressing the threat (Bodó 2019).

A concern is the non-transparent nature of recommender systems (Diakopoulos and Koliska 2017). When users do not understand why certain stories are recommended and others not, they may gradually doubt that they are given a fair presentation of the matters and suspect that media companies try to guide them in particular directions. The prevalence of fake news makes the situation even more problematic (Newman et al. 2018), as readers may both fear that they are not presented all relevant news and the presented stories may not even be true. These are aspects of personalized news platforms that need to be taken seriously by media companies to avoid an erosion of trust and relevance in society.

Both political authorities and publishers are looking for means and methods to maintain media's position in society as a provider of fact-seeking journalism and high-quality public debates. Fact-checking organizations supplement news organizations in many countries, and features for adding more diverse and serendipitous news stories to the recommendations have been suggested (Helberger 2018). There are now also promising works on preventing bias in recommender systems, ensuring fairness across users and taking multiple parties take into account when recommendations are generated (e.g. Abdollahpouri and Burke 2019; Leonhardt, Anand, and Khosla 2018). The Council of Europe has addressed the need to enhance users' exposure to the broadest possible diversity of media content (Bodó 2019). Further development of recommender systems may address many of these issues, though it may be necessary also to evaluate the fine line between algorithms and human judgment.

CONCLUSIONS AND FUTURE DIRECTIONS

Scandinavian news organizations mostly follow what is referred to as a "news logic of recommendation". There is a limited user base and content base, and the platforms are careful about the user data they collect. Conversion may often be more important than click rates, and editorial control is still considered important. News personalization experiments from Scandinavian news outlets indicate some improvements in click rates and reading times, but the results are moderate and sometimes not conclusive. The increased focus on privacy and transparency has led to the development of techniques like session-based recommendation strategies, though it is still unclear how efficient they are in the commercial news domain. With the complexities of modern news platforms, it is difficult both to evaluate individual strategies on live traffic and investigate the dependencies between platform components. It is also an open question to what extent click rates and reading times are appropriate evaluation metrics for platforms that serve such important roles in society at large.

AI technologies are at the core of news personalization. But AI is equally relevant for other critical newsroom efforts such as monitoring news consumption, analyzing users' behavior, or even producing news content. The developments within the field of recommender systems are not occurring in silos, but as part of an ongoing digital transformation of the media industry. News media organizations are in many cases turning into *technology* companies, either by producing or at scale utilizing advanced technologies.

One trend that we see is the ambition to embed journalistic values into algorithmic news solutions. At the Swedish public broadcaster, Sveriges Radio, the digital news team has developed a "public service algorithm" recommending content based on editors' opinions of the importance and life span of the news story mixed with values specifically answering to its public service mission (Zachrisson 2020). Those values include how well a story conveys unique voices, if it covers blank spots, if the audio storytelling is compelling and if it is recorded out in communities. A similar commercial solution is found at the Nordic media group Schibsted, where a personalization algorithm is optimized on three metrics: the journalistic mission, engagement, and conversion. In addition to this,

some media organizations have published ethical guidelines steering their AI and machine learning development, including the BBC and Bayerischer Rundfunk.

Looking to the future, we see how news organizations may increasingly—and more routinely—practice hybridization. As previously discussed, we consider how recommender systems call for balance between automated and manually assessed recommendations. This idea spans beyond recommender systems. Diakopoulos (2019) uses the term *hybridization* to describe how media organizations are increasingly *"combining algorithms, automation, and people to conduct news work"*. In this process, the media industry is tasked with pairing human and machine expertise. Guzman and Lewis (2020) discuss how technologies such as automated-writing software, conversational agents, or social bots are made to perform specific tasks within communication processes which were formerly associated with human beings and effectively function as communicators themselves (rather than mediators of human communication). As AI technologies advance in sophistication, the media industry – and the people in it — must acquire new skills and techniques to leverage them in accordance with the publishing mission.

On that topic, the media industry is facing diverse challenges related to the ongoing technological shift. How to meet the many challenges mentioned in this article – including fundamental questions related to news credibility and an increasing digitization and prediction of human experience – is being managed in various ways. Studying Norway specifically, Maasø, Sundet, and Syvertsen (2007) showed how privileges such as press subsidies, VAT exemptions, and special legal protections are awarded to the media under the premise that the benefiting media organizations will adhere to expectations - both non-binding and binding - reaching beyond those traditionally put on businesses. How these expectations are to be met as AI becomes part of the newsroom is a topic explored both within and beyond the Scandinavian region.

As noted by Carlson (2015), the implementation of intelligent systems in newsrooms across the world is generating a "technological drama" over the potential, limitations and potential dangers of these new tools. While an over-belief in the abilities of these emerging technologies is undesirable, too much pessimism about AI in the media sector can be distracting and even detrimental. Instead, the journalism profession could be better served by a proactive approach toward AI developments, with potential measures including high level evaluations of strengths, weaknesses, ethical questions, and potential future applications of AI for the news media field (Broussard et al. 2019). In any case newsrooms are gradually transforming into interdisciplinary teams that need to understand user preferences and accept usage data as an unerring currency for editorial strategies.

The technological and economic conditions of the production, distribution and consumption of news are changing. AI is central in news personalization, but we now see AI being introduced into a number of tools for news production and consumption. Journalistic processes already benefit from AI-driven tools, and AI may enable entirely new ways of communicating with readers and analyzing events and news-related data. There is no doubt that AI is driving important changes in the media industry, and personalization is only one of many interesting applications of the technology. In the long run there may be a whole range of AI applications that turn out to be instrumental to the digital transformation of media companies and their future profitability.

REFERENCES

Abdollahpouri, H., and R. Burke 2019. "Multi-stakeholder Recommendations and its Connection to Multi-sided Fairness." *Workshop on Recommendation in Multistakeholder Environments (RMSE), 13th ACM Conference on Recommender Systems (Recsys), Copenhagen.*

Arns, T. 2018. *The New Paywall is Dynamic: How the Wall Street Journal Grew Subscriptions by 25 Percent*. 27 April. Oslo, Norway: Cxense ASA. .

Athalye, S. 2013. *Recommendation System for News Reader*. MSc thesis, San Jose, CA, USA: Fac. Computer Science, San Jose State Univ , 2013, https://doi.org/10.31979/etd.xn48-6q4j.

Bodó, B. 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): 1054–75, https://doi.org/10.1080/21670811.2019.1624185.

Bodó, B., N. Helberger, S. Eskens, and J. Möller 2019. "Interested in Diversity: The Role of User Attitudes, Algorithmic Feedback Loops, and Policy In News Personalization." *Digital Journalism* 7(2): 206–29.

Borges, H. L., and A. C. Lorena 2010. "A Survey on Recommender Systems for News Data." In *Smart Information and Knowledge Management*. Berlin, Germany: Springer, 129–51.

Broussard, M., N. Diakopoulos, A. L. Guzman, , R. Abebe, M. Dupagne, and C. Chuan 2019. "Artificial Intelligence and Journalism." *Journalism and Mass Communication Quarterly* 96(3): 673–95.

Burke, R. 2002. "Hybrid Recommender Systems: Survey and Experiments." *User Modeling and User-Adapted Interaction* 12(4): 331–70.

Cañamares, R., M. Redondo, and P. Castells, 2019. "Multi-armed Recommender System Bandit Ensembles." In *Proceedings of the 13th ACM Conference on Recommender Systems (Recsys)*. Copenhagen. https://doi.org/10.1145/3298689.3346984.

Carlson, M. 2015. "The Robotic Reporter." *Digital Journalism* 3(3): 416–31, https://doi.org/10.1080/21670811.2014.976412

Chakraborty, A., S. Ghosh, N. Ganguly, and K. P. Gummadi 2017. "Optimizing the Recency-Relevancy Trade-Off in Online News Recommendations." In *Proceedings of the 26th International Conference on World Wide Web*, 837–46.

Chakraborty, A., M. Luqman, S. Satapathy, and N. Ganguly 2018. "Editorial Algorithms: Optimizing Recency, Relevance And Diversity For Automated News Curation." In *Companion Proceedings of The Web Conference 2018*, 77–8.

Claypool, M., A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin 1999. "Combining Content-Based and Collaborative Filters in an Online Newspaper." In *Proceedings of ACM SIGIR Workshop on Recommender Systems*.

Clement, J. 2020. Google: Ad revenue 2001–2018. Technical report. Statista, 5 February. Hamburg, Germany.

Das, A. S., M. Datar, A. Garg, and S. Rajaram 2007. "Google News Personalization: Scalable Online Collaborative Filtering." In *Proceedings of the 16th international conference on World Wide Web*.

De Souza Pereira Moreira, G., D. Jannach, and A. M. Da Cunha 2019. "Contextual Hybrid Session-based News Recommendation with Recurrent Neural Networks." *IEEE Access* 7: 169185–203. https://doi.org/10.1109/ACCESS.2019.2954957.

Diakopoulos, N. 2019. "Towards a Design Orientation on Algorithms and Automation in News Production." *Digital Journalism* 7(8): 1180–4.

Diakopoulos, N., and M. Koliska 2017. "Algorithmic Transparency in the News Media." *Digital Journalism* 5(7): 809–28.

Dubois, E., and G. Blank 2018. "The Echo Chamber is Overstated: The Moderating Effect of political Interest and Diverse Media." *Information, Communication & Society* 21(5): 729–45.

Epure, E. V., J. E. Ingvaldsen, R. Deneckere, and C. Salinesi 2016. "Process Mining for Recommender Strategies Support in News Media." In *Proceedings of the 10th International Conference on Research Challenges in Information Science (RCIS)*, 1–2.

Eskens, S. 2019. "A Right to Reset Your User Profile and More: GDPR-Rights for Personalized News Consumers." *International Data Privacy Law* 9(3): 153–72.

Felicio, C. Z., K. V. R. Paixão, C. A. Z. Barcelos, and P. Preux 2017. "A Multi-Armed Bandit Model Selection for Cold-Start User Recommendation." In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization. (UMAP'17)*, 32–40.

Fletcher, R., and R. K. Nielsen 2018. "Automated Serendipity." *Digital Journalism* 6(8): 976-89, https://doi.org/10.1080/21670811.2018.1502045.

Goldberg, D., D. Nichols, B. M. Oki, and D. Terry 1992. "Using Collaborative Filtering to Weave an Information Tapestry." *Communications of the ACM* 35(12): 61–70.

Good, N., J. B. Schafer, J. A. Konstan, A. Borchers, B. Sarwar, J. Herlocker, and J. Riedl 1999. "Combining Collaborative Filtering with Personal Agents for Better Recommendations." In *Proceedings of the 16th national conference on Artificial intelligence and the 11th Innovative applications of artificial intelligence conference innovative applications of artificial intelligence*.

Graff, R. 2015. "How the Washington Post used Data and Natural Language Processing to Get People to Read More News." *Knight lab*, Northwestern University, June 3.

Gulla, J. A., L. Zhang, P. Liu, Ö. Özgöbek, and X. Su 2017. "The Adressa Dataset for News Recommendation." In *Proceedings of the International Conference on Web Intelligence*, 1042–48.

Guzman, A. L., and S. C. Lewis 2020. "Artificial Intelligence And Communication: A Human-Machine Communication Research Agenda." *New Media & Society*, 4 July. https://doi.org/10.1177/1461444819858691.

Helberger, N., K. Karppinen, and L. D'Acunto 2018. "Exposure Diversity as a Design Principle for Recommender Systems." *Information, Communication & Society* 21(2): 191–207.

Helberger, N. 2019. "On the Democratic Role of News Recommenders." *Digital Journalism* 7(8): 993-1012, https://doi.org/10.1080/21670811.2019.1623700.

Hiralall, M. 2011. "Recommender Systems For E-Shops." *Business Mathematics and Informatics paper*. Amsterdam: Vrije university.

Holmstad, Ø. 2017. "Hvor godt virker algoritmedrevne anbefalinger i NRK TV? ["How well do algorithmically driven recommendations work at NRK TV?]. Evaluation report. NRK beta, 18 May.

Kille, B., Ö. Özgöbek, and A. Lommatzsch 2018. "Defining Contextual Factors for News Consumption." In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management 2018 Workshops*, 6th International Workshop on News Recommendation and Analytics.

Kozyreva, A., S. Herzog, P. Lorenz-Spreen, R. Hertwig, and S. Lewandowsky 2020. *Artificial Intelligence in Online Environments: Representative Survey of Public Attitudes in Germany*. Berlin: Max Planck Institute for Human Development. https://doi.org/10.17617/2.3188061.

Lancefield, D., M. Ambler, M. Rauber, and R. Patel 2011. *Research into Consumer Understanding and Management of Internet Cookies and The Potential Impact of the EU Electronic Communications Framework*. UK: PricewaterhouseCoopers LLP , April.

Leonhardt, J., A. Anand, and M. Khosla 2018. "User Fairness in Recommender Systems." In *WWW'18 Companion. The 2018 Web Conference Companion*, Lyon. https://doi.org/10.1145/3184558.3186949.

Liu, J., P. Dolan, and E. R. Pedersen 2010. "Personalized News Recommendation Based on Click Behavior." In *Proceedings of the 15th international conference on Intelligent user interfaces*, 31–40.

Maasø, A., V. S. Sundet, and T. Syvertsen 2007. "Fordi de fortjener det: Publikumsdeltakelse som strategisk utviklingsområde i mediebransjen [Because They Deserve It: Audience Participation as a Strategic Development Area in The Media Industry]." *Norsk medietidsskrift* 14(2): 126–54.

Marconi, F. 2020. *Newsmakers – Artificial Intelligence and the Future of Journalism*. Columbia University Press 2020.

Melville, P., and V. Sindhwani 2011. "Recommender Systems." In *Encyclopedia of Machine Learning and Data Mining*. Boston, MA, USA: Springer, 829–38.

Mensing, D. 2007. "Online Revenue Business Model Has Changed Little Since 1996." *Newspaper Research Journal* 28(2): 22–37. https://doi.org/10.1177/073953290702800202.

Negroponte, N. 1995. *Being Digital*. New York: Alfred A. Knopf.

News Media Alliance, 2019. Google Benefit from News Content. *Economic Study*, Arlington, US: News Media Alliance .

Nylen, L. 2020. "Google Dominates Online Ads – And DOJ May Be Ready To Pounce." *Politico*. 6 April 2020.

Pariser, E. 2011. *The Filter Bubble: What the Internet is Hiding from You*. New York: Penguin.

Raza, S., and C. Ding 2020. *A Survey On News Recommender System – Dealing With Timeliness, Dynamic User Interest And Content Quality, And Effects Of Recommendation On News Readers*. arXiv:2009.04964.

Newman, N. 2016. Digital News Report. Reuter Institute for the Study of Journalism.

Newman, N., R. Fletcher, A. Kalogeropoulos, D. A. L. Levy , and R. K. Nielsen 2018. Digital News Report. Reuter Institute for the Study of Journalism.

Shoemaker, P. J., T. P. Vos, and S. D. Reese 2009. "Journalists as gate-keepers." *The handbook of journalism studies* 73.

Son, J.-W., A.-Y. Kim, and S. Park 2013. "A Location-Based News Article Recommendation wit Explicit Localized Semantic Analysis." In *Proceedings of the 36th international ACM SIGIR conference on Research and dvelopment in information retrieval*, 293–302.

Statista. 2019. *Online News Consumption In European Countries 2017*. Hamburg, Germany: Statista Research Department. Technical report. Statista, Statista Research Department 10 April. .

Statista. 2021. Advertising Revenue of U.S. Newspapers 2013–2019. *Statista Research Department*. Hamburg, Germany: Technical report. Statista, Statista Research Department 10 April. .

Sunstein, C. 2001. *Republic.Com*. Princeton, NJ: Princeton University Press.

Svendsen, R. D., J. A. Gulla, and J. Frøland 2019. "Anbefaling av nyhetsinnhold i praksis [Recommendation of News Content in Practice]." *Norsk medietidsskrift* 26:21 March. https://doi.org/10.18261/ISSN.0805-9535-2019-01-04.

Thurman, N. 2019. "Personalization of News." In *The International Encyclopedia of Journalism Studies*, edited by T . Vos , and F. Hanusch . Wiley-Blackwell.

Walker, M. 2019. *Americans Favor Mobile Devices Over Desktops and Laptops for Getting News*. Pew Research Center, November 19.

Wang, S., L. Cao, Y. Wang, Q. Z. Sheng, M. Orgun, and D. Lian 2019. *A Survey on Session-Based Recommender Systems*. arXiv preprint arXiv:1902.04864.

Weinberger, D. 2017. *Pointing at the Wrong Villain: Cass Sunstein and Echo Chambers*. Los Angeles Book of Reviews, 20 July.

Zachrisson, O. 2020. *An Algorithm for Empowering Public Service News*. 28 September. London School of Economics, Department of Media and Communications.

Zamith, R. 2018. "Quantified Audiences in News production." *Digital Journalism* 6(4): 418–35.

Zhang, L., P. Liu, and J. A. Gulla 2019a. "Dynamic Attention-Integrated Neural Network for Session-Based News Recommendation." *Machine Learning* 108(10): 1851–75. https://doi.org/10.1007/s10994-018-05777-9.

Zhang, S., L. Yao, A. Sun, and Y. Tay 2019b. "Deep Learning Based Recommender System: A Survey and New Perspectives." *ACM Computing Surveys* 52(1): 1–38.

Zhou, X., and R. Zafarani 2020. "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities." *ACM Computing Surveys* 53(5).

AUTHOR BIOGRAPHIES

Jon Atle Gulla is a professor of information systems at the Norwegian University of Science and Technology (NTNU) since 2002 and the director of the Norwegian Research Center for AI Innovation. He has a PhD in computer science from 1993 and holds three MSc degrees in computer science, linguistics and management. His research is on natural language processing and semantics in the context of recommender systems, search engines and conversational systems.

Rolf Dyrnes Svendsen is the communication manager of the Norwegian Research Center for AI Innovation (NorwAI) and media advisor for editorial and commercial networks and innovation projects. He was previously the leader of the NxtMedia Lab in Norway. His 40 years of media experience includes work as a journalist, holding multiple editor positions and heading editorial and commercial software development for a major Norwegian media group.

Lemei Zhang is a PhD candidate at Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. Her research focuses on user modelling and recommender systems. Zhang has conducted research in the area of user behavior analysis based on textual data streams and context-aware recommendation algorithms in collaboration with major media companies.

Agnes Stenbom is an industrial PhD candidate at the Royal Institute of Technology in Stockholm, Sweden, where she conducts research on AI as a transformative force in journalism. In addition to her research engagements, Stenbom leads the work with Responsible AI at the Nordic media group Schibsted.

Jørgen Frøland is a project manager at Polaris Media ASA, where he is responsible for the initiative on personalization technologies and is part of the effort to transform Polaris Media into a data-driven organization. He has a bachelor's degree in informatics and a MSc degree in management from the Norwegian University of Science and Technology. His research is on the practical use of AI in news recommendation and journalism in general.