Exploring the Ethical Challenges of AI and Recommender Systems in the Democratic Public Sphere

Khadiga Seddik[1] [0009-0000-0475-7631]

MediaFutures SFI, University of Bergen, Bergen, Norway
khadiga.seddik@uib.no

Abstract. The rapid integration of Artificial Intelligence (AI) and Recommender Systems (RSs) into digital platforms has brought both opportunities and ethical concerns. These systems, designed to personalize content and optimize user engagement, have the potential to enhance how individuals navigate information online. However, this paper shifts the focus to the ethical complexities inherent in such systems, particularly the practice of nudging, where subtle algorithmic suggestions influence user behavior without explicit awareness. Issues like misinformation, algorithmic bias, privacy protection, and diminished content diversity raise important questions about the role of AI in shaping public discourse and decision-making processes. Rather than viewing these systems solely as tools for convenience, the paper challenges the reader to consider the deeper implications of AI-driven recommendations on democratic engagement. By examining how these technologies can quietly influence decisions and reduce exposure to different perspectives, it calls for a reevaluation of the ethical priorities in AI and RSs design. The paper calls for creating a digital space that promotes independence, fairness, and openness, making sure AI is used responsibly to support democratic values and protect user rights.

Keywords: Recommender systems · Democracy · Public sphere · Ethical challenges · Nudging

1 Introduction

In today's digital age, the public sphere has undergone a significant transformation, thanks to the widespread adoption of AI and RSs. These technologies are used to make recommendations to users based on their preferences and past behaviors. While these technological advancements have undoubtedly brought significant benefits to the public sphere, such as personalized content and recommendations, and improved the efficiency and accessibility of the public sphere, they have also given rise to several ethical challenges. In particular, the ethical implications of AI and RSs have been a topic of much debate, with concerns being raised about their impact on the democratic nature of the public sphere. One of the major concerns is the use of nudging in RSs, which has the potential to influence people's decision-making processes, bringing into question the ethical implications of influencing users' choices. By analyzing the ethical implications of AI and RSs, this paper hopes to contribute to the ongoing debate about the role of these technologies in shaping our society and democratic values, ultimately advocating for a framework that prioritizes ethical considerations and protects user autonomy in a rapidly evolving digital landscape.

The paper begins by defining the public sphere and its essential role in democratic societies, highlighting the concerns raised by AI and RSs about their potential to undermine the principles of rational discourse, free expression, and accountability. It then examines the various ethical challenges posed by these technologies, including issues related to misinformation, filter bubbles, echo chambers, and algorithmic bias.

The paper then delves into the concept of "nudging", exploring its potential to shape user choices, guide user behavior, and even influence political preferences. It analyzes both the potential benefits and the ethical dilemmas of nudging, including the possibility of reinforcing existing power dynamics and restricting user autonomy.

Finally, the paper concludes by reflecting on the broader implications of AI and RSs for the future of the democratic public sphere, urging developers and policymakers to prioritize ethical considerations in the design and deployment of these powerful technologies.

2 The Public Sphere and Its Ethical Challenges

The public sphere refers to the space where individuals come together to discuss matters of common concern and engage in rational discourse. It is a crucial element of democratic societies as it allows citizens to express their opinions, engage in debate, and hold their leaders accountable. The concept of the public sphere was first introduced by German philosopher Jürgen Habermas in his book *The Structural Transformation of the Public Sphere* [8]. He argued that the public sphere emerged in Europe during the 18th century as a result of the rise of bourgeois society and that it played a vital role in the development of democratic institutions.

Habermas defined the public sphere as a space where individuals could come together to engage in rational discourse on matters of common concern. This space was characterized by certain key features, including freedom of expression, access to information, and the ability to participate in decision-making processes. Habermas's main focus is on the institutional changes that occurred in the seventeenth and eighteenth centuries, which led to the development of what we now recognize as public space.

In modern society, the public sphere takes on many different forms. It includes traditional media outlets such as newspapers, television, and radio, as well as social media platforms such as Facebook, Twitter, and YouTube. These platforms provide individuals with a means to express their opinions, share information, and engage in public debate. A study made by [5] claimed that the public sphere theory can be applied to the Internet because it has various unique features. Firstly, the Internet's structure allows for a diverse range of voices to be expressed, and the accessibility to create content is relatively easy, which enables people to publish personal blogs, social media profiles, upload videos, photos, and reports of current events to both social and professional news websites. Secondly, the structure of the Internet allows for feedback and exchange.

However, the public sphere is not without its challenges. One of the biggest challenges is the issue of media bias including gate-keeping bias, coverage, and statement bias [4]. Social media platforms have the potential to limit exposure to a range of diverse viewpoints and instead promote the formation of groups consisting of users with similar beliefs and opinions. These groups reinforce shared narratives, creating what is known as echo chambers [3], where individuals are only exposed to information that confirms their pre-existing beliefs. Additionally, social media platforms have been criticized for allowing the spread of misinformation. Another challenge is the issue of political polarization which has been exacerbated by social media. These platforms allow individuals to filter out opposing viewpoints and create their own personalized echo chambers.

There are several significant ethical challenges associated with the public sphere, including access and representation, where certain groups may be marginalized. Manipulation of public discourse occurs when various actors attempt to shape public opinion or suppress dissenting voices. The spread of misinformation and fake news can erode trust in institutions and undermine democratic processes. Additionally, ethical concerns involve privacy, freedom of speech, and balancing individual and collective interests.

The concept of the public sphere and its ethical challenges have become increasingly relevant in the age of technology. With the rise of AI and RSs, there is a growing concern about the impact of these technologies on the democratic public sphere. As news recommender systems become more widespread, they have the potential to shape public discourse and influence the way people perceive events and issues. In the next section, I will discuss the impact of these technologies on the democratic public sphere and explore the intersection between the public sphere and AI.

3 AI, News Recommender Systems, and the Democratic Public Sphere

During the past years, there has been a significant surge in interest in artificial intelligence. As a result, many people anticipate that AI will become increasingly used in everyday technologies and fundamentally transform societies in a significant and groundbreaking manner. However, there is an argument about whether AI is a threat or an opportunity to democracy.

On one hand, AI can enhance public discourse by offering personalized recommendations and filtering information based on users' interests and preferences. It also plays a crucial role in mitigating the spread of misinformation, helping to maintain the integrity of public discourse. Platform companies claim that implementing automated content filtering systems, which utilize algorithmic methods to identify harmful content, enables them to effectively regulate themselves. These claims are supported by evidence provided by corporations regarding the effectiveness of AI. For instance, Facebook reported in 2018 that its automated systems are capable of detecting and removing 99% of terrorism-related content and 52% of hate speech, nearly 100% of spam, 98.5%of fake accounts, and 86% of graphic violence-related removals [13].

On the other hand, there are concerns that AI-powered algorithms and recommendation systems can continue to propagate filter bubbles, echo chambers, and algorithmic bias, leading to the polarization of public opinion and the suppression of dissenting voices [6]. There is also a worry that AI could be used to manipulate public opinion and interfere in democratic processes, such as elections, by amplifying certain messages, suppressing others, or creating fake content. For example, an increasing body of research in computational social science has shown that AI bots are capable of influencing voter opinion, as well as launching attacks against journalists and undermining the credibility of political leaders [16].

Furthermore, the rise of interactive communication technology has transformed the internet into a primary source of news, owing to its availability around the clock. As a result, numerous news sources and agencies offer readers access to the latest news anytime and anywhere through online portals. To attract more traffic to their websites, these portals are increasingly utilizing AI and news recommender systems to enhance the user experience on their platforms. At the societal level, news recommender algorithms and overly personalized news recommendations can have negative effects on the general public by isolating people from counter attitudinal opinions and may lead to the filter bubble, echo-chambers, and political polarization [6]. It also affects users' behaviour in the long run, causing them to avoid counter-attitudinal information. This type of behaviour, at the societal level, poses a threat to democracy. Moreover, studies combining insights from the field of computer science with psychology have shown that diversity in recommendation sets increases user satisfaction [18].

The capacity to provide personalized and efficient recommendations to users is also responsible for some of the most significant worries regarding the influence of recommendation systems on democracy. If every user receives the news and information that aligns with their individual needs and preferences, it raises questions about the existence of a public space where diverse perspectives and ideas can converge. Critics caution that recommendations may be used in a way that restricts citizens' exposure to a range of viewpoints, creating a risk of limited access to different perspectives. In turn, there is a concern that the public sphere is slowly disappearing [9].

4 Exploring the Ethical Challenges of News Recommender Systems

Philosophy has extensively discussed moral principles and which principles may be correct throughout history. Ethics is a branch of philosophy that deals with moral principles, values, and theories of right and wrong conduct. Philosophers have explored various moral principles and ethical frameworks, seeking to develop systems of moral reasoning that can guide human behavior. Considering what we previously stated about RSs and their significant impact on a wide range of stakeholders, as they guide our preferences and steer our decisions, it is crucial to consider the ethical issues that arise from the extensive reach of these systems. These ethical issues may include inappropriate content, privacy concerns, social implications, and other related concerns.

In [14], the authors proposed that there is a widely agreed upon notion that two classes of variables are morally significant: actions and consequences. The authors further assume that by studying the behavior and impact of a RSs, it is possible to gain a comprehensive understanding of the ethical implications involved. Their proposal involves a taxonomy of ethical issues of RSs categorized along two dimensions: (a) the utility they contain and the potential for the RSs to negatively impact the utility of any of its stakeholders, (b) violate the stakeholders' rights. The adverse ethical impact of an action could occur either immediately or it could pose a risk for unethical consequences in the future. In the next section, I will discuss some ethical challenges addressed in the literature and their possible solutions.

4.1 Fake News and Misinformation

Some online news articles are known to be clickbait or fake news. Additionally, these articles may include hostile indicators or subliminal messages, as well as harmful or detrimental content, such as racism and hateful language [23]. Since news platforms deal with massive volumes of news articles every day, it's nearly impossible to filter out all fake news and harmful content. Therefore, it's important to take into account this ethical issue in the implementation and design of news recommender systems. Recommendation algorithms on news platforms significantly contribute to the spread of misinformation [7]. They have been criticized for unintentionally amplifying and distributing false information. For example, content-based recommendation algorithms recommend news similar to those they have previously expressed a liking for, and collaborative-filtering algorithms recommend news that is popular among similar users, those techniques play a role in establishing and strengthening echo-chambers. Furthermore, recommendation algorithms have a tendency to magnify biases, like homogeneity and popularity biases, in their attempt to furnish pertinent suggestions. These biases can restrict users' exposure to differing viewpoints and make users vulnerable to misinformation. When most recommendation algorithms evaluate the quality of the recommendations based on how personalized and accurate the recommendation items to users, this may increase the spread of misinformation.

However, Algorithms that encourage some level of cognitive dissonance, and metrics that aim to balance user satisfaction and discomfort, may be better suited for addressing and preventing misperceptions. One proposed solution to identify and differentiate the dissemination of false information is having datasets that contain items that have already been marked or recognized as misleading [7]. Another solution is to filter our news articles with short reading dwell time as this news is most probably clickbait [23].

4.2 Bias

Many prior works have discussed various types of bias in the field of machine learning in general and in RSs specifically. Li et al. [12] claim that since the training data and learning algorithm are the fundamental building blocks of machine learning systems, they are also the primary sources of bias. As a result, they categorized biases into two main types: Data Bias and Algorithmic Bias.

Data bias in RSs occurs when the data used to train the system is not diverse, incomplete, or contains inaccuracies, leading to biased recommendations. In news recommendations, data bias can also manifest as presentation bias, where the placement and size of news articles influence user click behavior [24]. On the other hand, the bias may come from the recommendation's algorithm itself, which is called algorithmic bias. Algorithmic bias may occur due to the use of certain optimization methods or biased estimators in an inappropriate manner.

Another type of bias is popularity bias [26] which refers to a type of bias that occurs when the system recommends news articles based solely on their popularity or the number of clicks/views they have received, rather than on their relevance or quality. This bias can lead to a self-reinforcing cycle where popular articles continue to be recommended, while lesser-known but potentially valuable articles are overlooked, resulting in a narrow or skewed view of the news.

Zhu et al. [26] introduced three factors that produce and affect the popularity bias: (a) audience size imbalance: It is commonly observed that items tend to follow a long-tail distribution, where a small number of items have a large audience size, while the majority of items have a smaller audience. (b) Model bias: a recommendation model assigns a higher rank to an item that has more clicks in the training data. This type of bias is the same as algorithmic bias mentioned in [12]. (c) Closed feedback loop: When the new models trained by data generated from the previous models, the popularity bias in the previous models will be inherited by the new models.

There are many ways introduced in the literature to mitigate bias in the recommendation process. Caton and Haas [2] developed a framework to mitigate bias in machine learning through three approaches: pre-processing, in-processing, and post-processing methods. Pre-processing involves cleaning and diversifying the data before training the model to reduce bias. In-processing methods integrate fairness metrics into the model's objective function during the algorithm development phase. Post-processing techniques, on the other hand, recompute scores or re-rank the recommendation list to address bias.

4.3 Fairness

Bias and fairness are closely linked in machine learning. Bias refers to systematic errors in a model or dataset, which can result in unfair treatment of certain groups. To ensure fairness, machine learning systems must identify and address bias in both the data and algorithms used. It is widely acknowledged in the field of fairness research that there is no universally agreed-upon definition of fairness, due to the fact that the requirements for fairness can vary depending on the specific context or situation. In news recommendation systems, unfairness can arise from biases related to sensitive user attributes. Users with similar sensitive attributes may display similar click behaviors [24]. If the model captures these biases, it may recommend news to certain groups while excluding others. This results in unfairness, as some users are unable to access news that matches their interests.

Wu et al. [25] proposed a framework for developing fair RSs that aim to mitigate bias and promote fairness. The framework is based on a multi-dimensional approach that includes pre-processing, in-processing, and post-processing methods to address different sources of bias. The framework takes into account various factors, including user attributes and item characteristics, to ensure that the recommendation results are fair and unbiased. It also incorporates fairness metrics to evaluate the performance of the RSs and to monitor for potential bias. While recommendation task can be recognized as a classification problem, some prior works have explored fair classification as a means to avoid discrimination in classification rule mining (e.g. [15]). This could involve developing algorithms that consider fairness criteria in their optimization process or taking post-hoc interventions to address discriminatory outcomes. The goal is to ensure that classification rules do not unfairly discriminate against certain groups of individuals based on their race, gender, age, or other sensitive attributes.

4.4 Privacy Protection

RSs collect and analyze personal data and user behavior's data to be able to provide personalized recommendations to each user, which makes privacy protection an important concern in RSs. The possibility of privacy leaks is high in the process of recommendation system work [10]. This is because RSs typically require access to large amounts of personal data in order to generate personalized recommendations, and this data can be sensitive and potentially identifiable. In [14], Milano et al. summarized the four stages in which privacy leakage can occur:

1. Data collection or sharing without explicit user consent.
2. Data storage, as storing data sets can increase the risk of their leakage to external agents or be targeted by de-anonymization attempts.
3. At the point where the system is capable of making inferences from the data, worries about privacy also emerge.
4. Collaborative filtering recommendation model: when the system creates the user model using the data collected on other users' interactions.

While Huang et al. described the stages of privacy leakage and their consequences from a different perspective in [10], they summarized the process of privacy leakage in three stages with the associated outcomes of each stage:

1. User modeling: The stage of obtaining user 's personalised preferences to create the user model. The collected user data consists of explicit data such as name, preferences, and keywords, and implicit data such as user behaviour on the website. During this phase, issues of privacy arise, such as unauthorized access, collection, monitoring, analysis, consolidation, transmission, and storage of data.
2. Calculation: in this stage, recommendation algorithm calculates similarity to generate the recommendation list. Privacy concerns in this stage include improper analysis, unauthorized transmission of data, and others.
3. Generating recommendation results: Privacy concerns in this stage include improper analysis, unauthorized transmission of data, and misleading recommendation.

A few studies have focused on the issue of maintaining privacy in news recommendation systems. One proposed solution was using federated learning [17], which involves training machine learning models on decentralized data. This approach ensures that the user's data remains on their devices, and the data is not shared with the server. Another solution is the algorithmic solution, which uses encryption to user's data. It also involves making the algorithms used in the RSs more transparent to users. This approach can be used to help users understand how their data is being used and ensure that the RSs are making recommendations that are fair and unbiased. Protecting user privacy in RSs often requires sacrificing some accuracy in recommendations. It's essential to analyze the trade-off between privacy and accuracy to find an optimal balance. Techniques like differential privacy or federated learning can introduce noise or limit data usage, affecting recommendation accuracy. Therefore, a careful assessment of this trade-off is vital to ensure user privacy while maintaining acceptable accuracy in recommendations.

5 Nudging in Recommender Systems

The development of RSs has brought many benefits to users, including personalized recommendations and improved user experiences. However, with the increasing complexity of these systems, there are ethical challenges that arise and impact the democratic public sphere.

In the previous section, I discussed some ethical concerns about issues such as filter bubbles, recommendation bias, spread of misinformation, and invasion of privacy. These challenges have prompted researchers and developers to consider the use of nudges in RSs. Nudges can be used to guide users towards certain recommendations while still preserving their freedom of choice. By applying nudges ethically, developers can help users make better-informed decisions while also mitigating some of the ethical challenges posed by RSs.

In this section, I will start by providing the general definition of *nudge*, followed by the definition of nudge in the field of recommender system. Finally, I will discuss some ethical challenged of nudge in recommender system and its impact on the democratic public sphere.

5.1 The Definition of Nudge

Before we go further, we need to know the clear definition of *nudge* in general. Thaler and Sunstein defined *nudging* in their book as *any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives* ([20], pp. 6). They presented the concept of nudging as a way to help people make better decisions in their daily lives. The authors argue that by understanding the way people think and behave, and by designing choice architectures that guide people towards better choices, we can improve outcomes in many areas of life.

The authors adopted the concept of *libertarian paternalists*. The libertarian part is based on the belief that individuals should generally be free to make their own choices and have the option to opt-out of undesirable arrangements if they wish. They emphasize that their approach is about preserving or increasing freedom of choice for individuals. When adding the term *libertarian* to the term *paternalism* it emphasizes the importance of preserving individual liberty. So, *libertarian paternalism* refers to an approach that is focused on preserving individual liberty and making it simple for people to make their own choices without being burdened or constrained. Libertarian paternalisms want to encourage people to make better decisions, but they do not want to limit or infringe upon their freedom. The main idea is to preserve individual autonomy while still guiding people toward better decisions through the use of nudges.

On the other hand, the paternalistic aspect centers around the notion that it is acceptable for private institutions and governments to make conscious efforts to guide individuals towards choices that will benefit their lives. This implies that institutions have a responsibility to use nudges to improve people's decision-making, even if it means limiting their choices in certain situations. Overall, it involves balancing the desire to preserve individual freedom with the need to guide individuals towards better decisions.

The fundamental principle of libertarian paternalism is to provide nudges that have the highest potential to assist individuals while causing the least possible harm. Thaler and Sunstein set what they called *a golden role* of libertarian paternalism: *offer nudges that are most likely to help and least likely to inflict harm* ([20], pp. 72). The authors suggest that nudges are necessary in situations where decisions are challenging, infrequent, do not provide immediate feedback, and are difficult to understand. Additionally, individuals are most likely to benefit from nudges in situations where decisions are delayed, or where individuals struggle to translate the relevant aspects of the situation into terms that they can easily understand ([20], pp. 73-76). This implies that nudges are most useful when individuals are faced with complex decisions that require guidance. Examples of such situations include choosing a retirement savings plan, selecting a health insurance plan, and making healthy food choices. In each case, individuals face complex decisions that can have long-term consequences for their well-being, and where nudges can help guide them towards better choices.

5.2 Nudging in the Context of Recommender Systems

The concept of nudging was initially explored in offline situations involving in-person decisions, focusing primarily on matters related to personal health or finances. As nudging gained in popularity and started to be used in the online environment, a new term emerged: *digital nudging*. It involves user interface elements of websites and software applications that affect choices of the users, the use of various design features, such as choice architecture, default settings, and social influence. As the main goal of RSs is to guide users to better choices and recommend items that users will most likely enjoy or items that the system think they are more suitable for the users to see, there are possibilities to merge personalized recommendations with "explicit" digital nudges [11]. For example, nudges in public service RSs can specifically manifest as default settings or the delivery of feedback on the recommended items they receive [21]. This feedback can be used to adjust the recommendations and make them more diverse over time.

Nudging can be used in RSs to benefit democracy by promoting exposure to diverse viewpoints which results in decreasing in selective exposure behavior and reducing filter bubbles. One approach is to use nudges that encourage users to explore content outside of their usual preferences or viewpoints and to engage with a broader range of content and viewpoints, which can help to promote a more informed and democratic society. There are different strategies and methodologies for using digital nudging in RSs which have been discussed in the literature. For example, [1] proposed a strategy for digital nudging in the recommender system called full homepage recommendation strategy. It is a nudging strategy which can be used to ensure access diversity by providing users with a wide range of content options on the homepage of a website or app. The idea behind this strategy is that the news recommender system pre-selects news articles for consumption based on a user's profile characteristics. And in consequence, it can help to reduce selective exposure by favoring the display of viewpoints that the user may be less familiar with or have not yet considered. Furthermore, the algorithm could also prioritize news articles that cover matters of public interest, rather than just reinforcing the user's existing preferences. By doing so, the system can help to ensure that users are exposed to a broader range of perspectives on issues that are relevant to society as a whole, rather than just their individual interests and biases.

However, this strategy has faced many criticisms, and some argued that it cannot be counted as nudge. Vermeulen and Judith argued that the pre-selected news article in the homepage limits individual choice - which contradicts with the definition of nudge - because users are not able to select content that has not been pre-selected for them [21]. The content that is not pre-selected is effectively de-selected, meaning that it cannot be consumed by the user. Users should have the ability to manage the parameters that inform the algorithmic selection and prioritization process through the user interface. This can help to ensure that users have more control over the content they see and can help to mitigate the potential negative effects of algorithmic filtering and bias.

On the other hand, if one of the choices, particularly the one associated with the diversity algorithm in our case, is pre-selected as the default option, then it constitutes a nudge within the choice architecture. However, this is only the case if individuals still have the option to opt out and choose an alternative option. As a result, Vermeulen and Judith didn't consider full homepage recommendation as a nudge on its own because it restricts user choice by pre-selecting content for consumption. Therefore, there is no such thing as *algorithmic nudging* in this context. Instead, the nudge is found in the default option, which is the pre-selection of the diversity-enhancing system for content display.

Other nudging mechanisms presented by Jesse and Jannach in their in-depth research [11] where they conducted a thorough literature review and identified 87 distinct nudging mechanisms. They then grouped these mechanisms into a new taxonomy, although not all of them had been previously applied in the context of RSs. Nevertheless, the authors proposed several potential methods for incorporating these mechanisms into RSs. They described the provision of recommendation as implicit nudge, as it incorporates various mechanisms that have been identified in the literature. Some of these applied mechanisms include *hiding nudge*, which involves hiding or de-emphasizing certain options or pieces of information and only present a specific subset of the available choices in order to nudge users towards selecting or prioritizing other options. Furthermore, *positioning nudge*, which means altering the placement or positioning of recommendations or other pieces of information or providing a particular ranking of the choices. For example, a recommender system might position certain recommendations more prominently on a user's screen or give some recommendations more salient placement or size in order to increase the likelihood that the user will select them. Alternatively, it might position recommendations in a way that encourages users to consider a wider range of options. Another approach is to increases the "ease and convenience" of selecting particular items. This involves manipulating the choice architecture to increase the likelihood that people will select certain options by making them more visible and accessible.

Although Nudging promotes positive outcomes such as healthier behaviors, increased access to diverse information, and decreased selective exposure behavior and filter bubble, the use of nudging in RSs raises ethical concerns regarding the extent to which it can influence user behavior, autonomy, and decision-making. This has led to a growing debate over the ethics of nudging in RSs. One of the key issues is the impact of nudging on the democratic public sphere, as it can potentially reinforce existing power dynamics and limit user autonomy. This raises important questions about the role of nudging in promoting democratic values and the responsibility of designers and developers in ensuring that nudging is used ethically in RSs. In this context, the next section will discuss the ethics of nudging and its implications for the democratic public sphere.

5.3 Ethical Challenges of Nudging

Although nudges are intended to assist individuals in making better decisions, this isn't always the case when it comes to practical application. Nudges are not always utilized in a manner that benefits the users. Weinmann et al. in [22] give an example of some low-cost airlines in Europe present non-essential options to their customers in a way that manipulates or guides them towards purchasing those options. This is an example of unethical nudging, as the airlines are taking advantage of customers' decision-making processes in order to increase their profits. While this approach may result in short-term gains for the company, there are potential negative consequences in the long run. Customers may become dissatisfied with the airline's practices and choose to take their business elsewhere, leading to a loss of goodwill. Negative publicity and even legal action may also result if the airline's actions are deemed to be deceptive or unfair. Overall, this situation highlights the importance of considering the ethical implications of nudges and the potential impact on both users and society.

Sunstein discussed several ethical challenges related to nudges and choice architecture in [19]. One ethical issue is that nudges can both safeguard the freedom of certain people and enforce specific actions on others. The author gave an example of the government that mandates large employers to implement automatic enrollment plans for retirement or health insurance, which nudges employees while forcing employers. Similarly, the government may require chain restaurants or movie theaters to disclose calorie information to customers, nudging them while forcing the restaurants. Other objections to nudges discussed in [19] include the issue of paternalism, where nudges are seen as a violation of individual autonomy and freedom of choice. However, as we discussed previously, others argue that individuals have the right to make their own decisions, even if they may not be in their best interest. Additionally, some nudges can force or pressure people to make a certain decision, even if they appear to give people a choice. other types of nudges can be seen as an insult to people's dignity and a way of treating them as if they were children.

As a rule, there are ethical considerations that must be taken into account when using nudging. Firstly, they must be transparent and communicated clearly to those who will be affected. Nudges should not not violate people's autonomy or freedom of choice. They should be designed to achieve their intended goals effectively without unfairly benefiting one group over another. Nudges should also show respect for human dignity and not be degrading or disrespectful. Additionally, the potential unintended consequences of nudges must be considered and minimized. Finally, those responsible for designing and implementing nudges should be held accountable for their actions. These ethical issues are also reflected in the use of nudging in RSs. The issue at hand is that designers cannot present options in a completely neutral way in a user interface, as they must choose a visual representation that determines the order of the options [11].

Another ethical concern also arises in the use of RSs, which aim to benefit both consumers and providers. While research mostly focuses on developing algorithms that assist users in finding relevant items, ethical considerations must take into account the utility of all stakeholders, including providers [11]. According to the categorization of ethical issues of RSs proposed by [14], any aspect of a recommender system that could negatively impact the utility of any of its stakeholders or impose such negative impacts constitutes a feature that is ethically relevant. Therefore, a recommender system that fails to consider the utility of all stakeholders does not operate ethically. Another popular issue is that RSs may filter out certain options for users based on personalized relevance, potentially infringing on the principle of freedom of choice.

Therefore, it is important for designers and developers of RSs to strike a balance between the principle of freedom of choice and the desire for relevance and personalization. It is also important for ethical considerations to take into account the utility of all stakeholders in the design and implementation of RSs.

6 Conclusion

In conclusion, the development of AI and RS has brought significant benefits to the public sphere, but it has also raised ethical concerns. One of the major ethical challenges in RS is the use of nudging. While nudges can help users make better decisions, they can also be used unethically to manipulate users and influence their choices. This is especially problematic in the context of the democratic public sphere, where the free flow of information is critical to ensuring a well-informed citizenry. As this essay has explored, nudging in RS has the potential to influence people's decision-making processes, which can impact the principles of autonomy, free of choice, and fairness in the democratic public sphere. The use of unethical nudges in RS can lead to short-term gains, but may also have long-term consequences. The designers of these systems must be mindful of these ethical challenges and work to ensure that users are not unfairly influenced by nudges. Ultimately, as we continue to grapple with the ethical challenges of AI and RS, it is important to remember that these technologies are tools that can be used for both good and bad. As such, it is our responsibility to use them ethically and ensure that they serve the best interests of society. The goal should be to create a democratic public sphere that is inclusive, transparent, and ethical, and that allows individuals to make informed decisions based on accurate and trustworthy information.

Acknowledgments. This work was supported by the NEWSREC Project (project number: 324835) and by industry partners and the Research Council of Norway with funding to MediaFutures: Research Centre for Responsible Media Technology and Innovation, through the Centers for Research-based Innovation scheme, project number 309339.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bernstein, A., de Vreese, C., Helberger, N., Schulz, W., Zweig, K., Baden, C., Beam, M.A., Hauer, M.P., Heitz, L., Jürgens, P., Katzenbach, C., Kille, B., Klimkiewicz, B., Loosen, W., Moeller, J., Radanovic, G., Shani, G., Tintarev, N., Tolmeijer, S., van Atteveldt, W., Vrijenhoek, S., Zueger, T.: Diversity in News Recommendation (Dagstuhl Perspectives Workshop 19482). Dagstuhl Manifestos **9**(1), 43–61 (2021). https://doi.org/10.4230/DagMan.9.1.43
2. Caton, S., Haas, C.: Fairness in machine learning: A survey. ACM Comput. Surv. **56**(7) (Apr 2024). https://doi.org/10.1145/3616865
3. Cinelli, M., Morales, G.D.F., Galeazzi, A., Quattrociocchi, W., Starnini, M.: The echo chamber effect on social media. Proceedings of the National Academy of Sciences **118**(9), e2023301118 (2021). https://doi.org/10.1073/pnas.2023301118
4. D'Alessio, D., Allen, M.: Media Bias in Presidential Elections: A Meta-Analysis. Journal of Communication **50**(4), 133–156 (01 2006). https://doi.org/10.1111/j.1460-2466.2000.tb02866.x
5. Edgerly, S., Vraga, E., Fung, T., Joon, T., Woo, M., Yoo, H., Veenstra, A.: Youtube as a public sphere: The proposition 8 debate (Oct 2009), https://aoir.org/wp-content/uploads/2020/02/IR10-Conference-Program.pdf, internet Research 10 - Annual Conference of the Association of Internet Researchers : Internet: Critical, IR 10 ; Conference date: 08-10-2009 Through 10-10-2009
6. Elahi, M., Jannach, D., Skjærven, L., Knudsen, E., Sjøvaag, H., Tolonen, K., Holmstad, , Pipkin, I., Throndsen, E., Stenbom, A., Fiskerud, E., Oesch, A., Vredenberg, L., Trattner, C.: Towards responsible media recommendation. AI and Ethics **2** (02 2022). https://doi.org/10.1007/s43681-021-00107-7
7. Fernández, M., Bellogín, A.: Recommender systems and misinformation: The problem or the solution? In: Tommasel, A., Godoy, D., Zubiaga, A. (eds.) Proceedings of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems co-located with 14th ACM Conference on Recommender Systems (RecSys 2020), Rio de Janeiro, Brazil, September 25, 2020. CEUR Workshop Proceedings, vol. 2758, pp. 40–50. CEUR-WS.org (2020), https://ceur-ws.org/Vol-2758/OHARS-paper3.pdf
8. Habermas, J.: The Structural Transformation of the Public Sphere. Studies in Contemporary German Social Thought, Polity Press (1992), https://books.google.no/books?id=e799caakIWoC
9. Helberger, N.: On the democratic role of news recommenders. Digital Journalism **7**(8), 993–1012 (2019). https://doi.org/10.1080/21670811.2019.1623700
10. Huang, W., Liu, B., Tang, H.: Privacy protection for recommendation system: A survey. Journal of Physics: Conference Series **1325**, 012087 (10 2019). https://doi.org/10.1088/1742-6596/1325/1/012087
11. Jesse, M., Jannach, D.: Digital nudging with recommender systems: Survey and future directions. Computers in Human Behavior Reports **3**, 100052 (2021). https://doi.org/https://doi.org/10.1016/j.chbr.2020.100052
12. Jin, D., Wang, L., Zhang, H., Zheng, Y., Ding, W., Xia, F., Pan, S.: A survey on fairness-aware recommender systems. Information Fusion **100**, 101906 (2023). https://doi.org/https://doi.org/10.1016/j.inffus.2023.101906
13. Marsden, C., Meyer, T., Brown, I.: Platform values and democratic elections: How can the law regulate digital disinformation? Computer Law and Security Review **36**, 105373 (2020). https://doi.org/https://doi.org/10.1016/j.clsr.2019.105373

14. Milano, S., Taddeo, M., Floridi, L.: Recommender systems and their ethical challenges. AI and Socity **35**(4), 957–967 (Dec 2020)
15. Pedreshi, D., Ruggieri, S., Turini, F.: Discrimination-aware data mining. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. p. 560–568. KDD '08, Association for Computing Machinery, New York, NY, USA (2008). https://doi.org/10.1145/1401890.1401959
16. Philip N. Howard, S.W., Calo, R.: Algorithms, bots, and political communication in the us 2016 election: The challenge of automated political communication for election law and administration. Journal of Information Technology & Politics **15**(2), 81–93 (2018). https://doi.org/10.1080/19331681.2018.1448735
17. Qi, T., Wu, F., Wu, C., Huang, Y., Xie, X.: Privacy-preserving news recommendation model learning. In: Cohn, T., He, Y., Liu, Y. (eds.) Findings of the Association for Computational Linguistics: EMNLP 2020. pp. 1423–1432. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.findings-emnlp.128
18. Raza, S., Ding, C.: News recommender system: a review of recent progress, challenges, and opportunities. Artificial Intelligence Review **55**, 1–52 (01 2022). https://doi.org/10.1007/s10462-021-10043-x
19. Sunstein, C.: Nudges and choice architecture: Ethical considerations. Yale journal on regulation **Yale Journal on Regulation (forthcoming)** (06 2015)
20. Thaler, R.H., Sunstein, C.R.: Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale University Press, New Haven, CT (2008)
21. Vermeulen, J.: To nudge or not to nudge: News recommendation as a tool to achieve online media pluralism. Digital Journalism **10**, 1–20 (02 2022). https://doi.org/10.1080/21670811.2022.2026796
22. Weinmann, M., Schneider, C., Brock, J.v.: Digital nudging. Business & Information Systems Engineering **58**(6), 433–436 (2016). https://doi.org/10.1007/s12599-016-0453-1
23. Wu, C., Wu, F., Huang, Y., Xie, X.: Neural news recommendation with negative feedback (01 2021). https://doi.org/10.48550/arXiv.2101.04328
24. Wu, C., Wu, F., Huang, Y., Xie, X.: Personalized news recommendation: Methods and challenges (2021). https://doi.org/10.48550/arXiv.2106.08934
25. Wu, C., Wu, F., Wang, X., Huang, Y., Xie, X.: Fairness-aware news recommendation with decomposed adversarial learning. Proceedings of the AAAI Conference on Artificial Intelligence **35**(5), 4462–4469 (May 2021). https://doi.org/10.1609/aaai.v35i5.16573
26. Zhu, Z., He, Y., Zhao, X., Caverlee, J.: Evolution of popularity bias: Empirical study and debiasing (2022). https://doi.org/110.48550/arXiv.2207.03372