AI & SOCIETY (2025) 40:5657–5666
https://doi.org/10.1007/s00146-025-02314-w

REVIEW

Engineering equity: designing diversity-aware AI to reflect humanity

Donghee Shin[1]

Received: 17 December 2024 / Accepted: 11 March 2025 / Published online: 25 March 2025

Abstract
Diversity plays a crucial role in recommendation systems. Enhancing the diversity of recommendations can expand users' perspectives, improve user experience, and support social equity. Developing diversity-aware AI is essential for creating systems that are adaptive, ethical, and capable of reflecting the complexity of human society. The necessity to create diversity-aware AI stems from the understanding that if AI is to mimic human intelligence in meaningful ways, it must surpass static, monolithic models that narrowly reflect only a portion of the human experience. AI must embrace diverse perspectives, adapting not only to the varying needs and backgrounds of users but also to changes in societal understanding. Creating more human-like AI requires focusing on the diverse reasoning and behavior of artificial agents and developing systems capable of dealing with such diversity is key to achieving more human-like AI. This study discusses the necessity of diversity in AI, arguing that it is essential for overcoming the limitations of static models, incrementally combining different components of intelligence, and expanding the notion of what constitutes intelligent adaptation.

Keywords Moral-aware AI · Diversity-aware news recommender system · Diversity in news recommendation · Faire-aware AI

1 Diversity-aware AI: navigating the future with inclusivity and fairness

In an increasingly digitized world, AI shapes countless aspects of daily life—from the algorithms guiding medical diagnoses to the recommendation engines powering social media. However, as AI's influence expands, so does the recognition that these systems can perpetuate and even exacerbate societal biases if not carefully designed (Cachat-Rosset and Klarsfeld 2023). Diversity-aware AI, a paradigm that advocates for the integration of diverse perspectives, values, and lived experiences, has emerged as a solution to these challenges (Zhou 2024). It is designed to promote a heterogeneous and balanced representation in their decision-making processes, outputs, and recommendations. Unlike traditional AI models that often optimize for accuracy, efficiency, or popularity—sometimes leading to homogenization or reinforcing existing biases—diversity-aware AI actively incorporates mechanisms to ensure that diverse perspectives, demographics, and content are fairly represented (Umbrello and van der Poel 2021).

In a diversity-aware approach, AI systems are calibrated to promote a heterogeneous mix in their recommendations or outputs (Campo-Ruiz 2025). In a recommendation system for music or movies, diversity-aware AI would not only prioritize items based on users' past preferences but would also introduce a range of content from artists of different backgrounds, genres, or cultures. In hiring algorithms, diversity awareness might involve creating a candidate pool that reflects a variety of backgrounds, experiences, and skills, rather than disproportionately selecting candidates from a narrow, homogeneous group. The underlying assumption is that exposure to diverse perspectives enriches the user experience and counteracts the reinforcement of echo chambers or filter bubbles (Noble 2018), which can arise when algorithms exclusively focus on optimizing for similarity and user engagement.

It adapts to various users and items while ensuring inclusivity in recommendations. At first glance, this may seem contradictory. AI is designed to provide personalized and customized services while still requiring diversity. The key lies in the importance of diverse data and varied perspectives in developing effective AI systems (Shams et al. 2025).

✉ Donghee Shin
don.h.shin@ttu.edu

[1] Texas Tech University, Lubbock, United States

Diversity in data sources and the programming teams behind AI technologies ensures that the systems can understand and serve a wide range of users. This richness of input leads to more accurate recommendations, a better understanding of different cultures, and ultimately more effective solutions that resonate with a broader audience (Lin et al. 2024).

The principle of diversity in AI is not just an ethical ideal but a practical necessity for developing systems that can truly adapt (Achon et al. 2024). Human intelligence is characterized by its ability to adjust to new information, unfamiliar situations, and different viewpoints. Traditional AI models, by contrast, often rely on static representations of problems and data, constraining their ability to adapt to new scenarios. As such, diversity awareness becomes essential, as it facilitates a "model change" paradigm (Du et al. 2021). This means that rather than AI being bound by one set model, it must continually revise its own understanding in light of different perspectives and data inputs that may alter its perception of the problem space (Baumer 2017).

Diversity-aware AI incorporates principles from fairness-aware AI and value-sensitive algorithms. While such AIs ensure that AI systems reflect a broad range of perspectives and representations, fairness-aware AI prioritizes the prevention of bias and discrimination within AI systems to ensure equitable treatment across all demographic groups (Jui and Rivas 2024). The focus of fairness-aware AI is to create systems that do not systematically disadvantage any group based on protected attributes like gender, race, or socioeconomic background. While diversity-aware AI seeks to enhance variety and representation, fairness-aware AI addresses issues of justice and equal opportunity, working to prevent algorithms from unfairly benefiting or harming specific groups. In this sense, diversity-aware AI is an inclusive concept encompassing other value-aware AI such as fairness, transparency, trust, and ethics (Zhao et al. 2024). While diversity awareness may intersect with these values, it does not inherently guarantee or subsume them. Fairness, for instance, focuses on equitable treatment and outcomes, which may sometimes be in tension with diversity goals that emphasize broad representation. Transparency, which ensures that AI decisions are explainable and understandable, is a separate concern from diversity, as an AI system can be diverse but still operate as a black box (Crawford and Paglen 2021). Similarly, trust and ethics are broader normative concepts that involve user confidence, accountability, and moral considerations that extend beyond diversity itself. Rather than viewing diversity-aware AI as an overarching principle that absorbs these values, it is more accurate to recognize it as one of several interdependent ethical considerations, each requiring its own nuanced approach in AI design and governance.

Designing diversity-aware AI actively promotes and showcases variety across different demographic and identity groups. The aim is not simply to prevent discrimination but to actively enhance the representation of traditionally underrepresented or marginalized groups in various AI-driven applications. Diversity enables AI to expand beyond the "static model" approach, providing the groundwork for a system that can evolve alongside its changing environment (Currin et al. 2022). Without a diversity-aware foundation, AI systems run the risk of perpetuating biases, reinforcing stereotypes, and missing nuances that would otherwise be apparent to human decision-makers. These pitfalls not only hinder the utility of AI but can also result in harmful, exclusionary impacts on the very users AI is meant to serve. Thus, diversity awareness is a requirement that underlies any model change, encouraging systems to explore the boundaries of current understanding and pushing AI toward more human-like adaptability (Evans et al. 2022).

2 Diversity as a catalyst for adaptive, open-ended intelligence

Developing diversity-aware AI involves rethinking intelligence itself (Chen and Sundar 2024). Human intelligence is characterized by its ability to synthesize and adjust to new perspectives, which often arise incrementally through exposure to varied experiences and viewpoints (Hermann 2022). Similarly, diversity-aware AI must adopt an "open-ended" model of intelligence, where the system's understanding of tasks, values, and even its operational goals evolves over time through exposure to diverse data (Shin 2025). This approach not only enhances the adaptability of AI systems but also fosters a type of intelligence that is more responsive to complex, evolving environments (Jesse and Jannach 2021). For instance, an open-ended AI in a healthcare setting could improve its treatment recommendations by learning from data that reflects a wide array of patient backgrounds, medical histories, and treatment outcomes. However, it is important to distinguish diversity-aware AI from general-purpose AI. A diversity-aware system does not necessarily mean an AI capable of performing multiple distinct tasks across medical domains. Rather, it ensures that within a given task—such as diagnosing conditions from X-rays or recommending treatments—it leverages diverse and representative data to improve accuracy and inclusivity. In some contexts, such as dermatology or genomics, demographic diversity in training data is essential to avoid biased outcomes. However, in other domains like radiology, demographic attributes may have minimal impact, and the primary concern is ensuring sufficient diversity in anatomical variations, disease manifestations, and imaging conditions. Thus, diversity-aware AI aims to enhance robustness and fairness within its scope of application rather than serving as an all-encompassing general-purpose AI. Such an AI would recognize that a single, universal model is insufficient for addressing the nuances of individual cases. Instead, it would incrementally incorporate insights from different populations, medical practices, and treatment modalities to refine its recommendations. This incremental synthesis of diverse perspectives represents a fundamental shift from traditional models, allowing AI to develop a richer, more context-sensitive understanding that approximates human adaptability (Crawford and Paglen 2021).

The goal of diversity-aware AI is to ensure that AI systems are not only effective and efficient but also fair, inclusive, and reflective of the pluralistic societies in which they operate (Lin et al. 2024). At its core, diversity-aware AI is concerned with recognizing and respecting the diversity of human experiences, identities, and social contexts (Li and Liu 2021). This includes but is not limited to, aspects such as race, gender, socioeconomic status, culture, and language. Traditional AI models, primarily trained on data that reflects a limited subset of these dimensions, often fail to generalize well to diverse populations, leading to skewed results and reinforcing existing inequalities. For instance, facial recognition technologies have been shown to perform significantly better for certain demographic groups than others, often struggling with accuracy in identifying individuals from minority backgrounds. Such disparities underscore the urgency of designing AI that consciously accounts for diversity to avoid reproducing structural biases. The path to diversity-aware AI involves a multi-layered approach.

First and foremost, inclusive datasets are essential for diversity-aware AI (Yu et al. 2024). In many cases, biases in AI arise because the data used for training and validation do not represent the full spectrum of human experiences (Holstein et al. 2019). When AI models are trained predominantly on data from specific demographics, such as North American or European populations, their accuracy for people outside these demographics suffers. Expanding data collection to incorporate broader and more representative samples helps to mitigate this issue. However, it is essential to approach this task ethically, ensuring that data collection respects privacy rights and does not exploit marginalized communities (Aguirre et al. 2016).

Moreover, the algorithms themselves must be designed to detect and adjust for biases rather than reinforce them (Chen and Sundar 2024). This requires transparency in model development, as well as the implementation of fairness measures that can identify and mitigate unfair patterns. Techniques such as adversarial debiasing and fairness-aware regularization are increasingly used to counteract inherent biases in data, but these technical solutions are just one part of a larger equation. Designing diversity-aware AI also requires ongoing human oversight and accountability. Interdisciplinary teams, including experts in social sciences, ethics, and law, should be involved in the development process to ensure that these models are not just technically sound but also socially responsible (Hanna et al. 2020).

Additionally, the importance of stakeholder engagement in building diversity-aware AI cannot be overstated (Heitz et al. 2022). Engaging with communities that AI will impact is essential to understanding the nuances of how it might affect different groups. This participatory approach enables AI designers to gain insights into the specific needs, concerns, and values of these communities (Bastian et al. 2021). For example, when designing AI tools for healthcare, input from diverse patients and medical professionals can help developers anticipate how algorithms may impact various populations differently and guide them to design solutions that improve outcomes for all. The result is not only more equitable AI but also technology that enjoys greater public trust and legitimacy (Currin et al. 2022).

Diversity-aware AI also has a broader, societal value. By embedding diversity considerations into AI systems, organizations can create products that resonate with a global audience, fostering greater inclusion and understanding across cultural and geographical divides (Werder et al. 2024). For example, language processing models that accommodate linguistic diversity—such as dialects, nonstandard grammar, or multilingual inputs—can help break down barriers in communication, allowing more people to access technology in ways that are meaningful to them. In an age where AI mediates so many facets of human interaction, diversity-aware AI represents a powerful means of promoting cultural inclusivity and empowering underserved communities (Cachat-Rosset and Klarsfeld 2023). However, pursuing diversity-aware AI presents significant challenges. There are technical limitations, as current methodologies for fairness and debiasing are still in nascent stages. Additionally, the need for vast, representative datasets clashes with privacy concerns and the logistical difficulties of collecting data from diverse populations. The field is also constrained by broader societal challenges, such as systemic biases that AI alone cannot address. For example, achieving truly equitable healthcare outcomes through AI is complicated by underlying disparities in access to medical resources. While AI can play a role in addressing these issues, it must operate within a broader framework of social reform (Zowghi and Mahmud 2024).

To make the claim that diversity-aware AI contributes to more equitable societies more robust, diversity-aware AI should addresses: (1) How diversity-aware AI mitigates existing systemic biases**,** such as addressing historical underrepresentation in hiring, lending, or healthcare, thereby promoting broader social inclusion (van Esch et al. 2024), (2) The conditions under which diversity-awareness aligns with equity**,** particularly when it is coupled with fairness-aware and transparency-driven AI approaches that ensure balanced decision-making (Jora et al. 2022), and (3) Potential risks and necessary safeguards**,** recognizing that diversity promotion should not come at the cost of fairness, transparency, or merit-based considerations, and discussing ways to navigate these trade-offs effectively (Evans et al. 2022).

3 The importance of diversity-aware AI: a path to equitable algorithms

As AI pervades nearly every aspect of life, concerns around its fairness, inclusivity, and ethical use have come into sharper focus. Diversity-aware AI—AI that is explicitly designed to acknowledge and incorporate the range of human experiences, identities, and backgrounds—is a response to these concerns, aiming to create equitable and trustworthy technology for all (van Esch et al. 2024). Diversity in AI is not merely a checkbox or an ethical ideal, but a practical necessity for creating systems capable of learning, adaptation, and realistic human interaction (Loecherbach et al. 2020). Traditional AI models often work within a "static model" paradigm, which relies on a predefined dataset and follows fixed protocols for interpreting and responding to new inputs. While effective within limited contexts, this static approach restricts an AI's ability to adapt to new circumstances, perspectives, or users. Such models can perpetuate biases, reinforce stereotypes, and struggle to interpret situations outside of narrowly defined norms (Roche et al. 2023). The importance of diversity-aware AI cannot be overstated, as it stands to prevent harm, enhance innovation, and ultimately contribute to a more just and inclusive society. The central reason for prioritizing diversity-aware AI is to prevent harmful biases from perpetuating or amplifying inequalities. AI systems are only as fair as the data and assumptions on which they are built (Mattis et al. 2022). When training data is skewed toward certain demographics, such as majority racial or socioeconomic groups, the resulting AI models may perform poorly for others, leading to adverse outcomes. For example, AI algorithms in healthcare have been shown to misdiagnose or inaccurately assess risk for certain populations due to underrepresentation in the training data. These biases are not merely technical flaws; they have real-world consequences that disproportionately affect already marginalized communities. Diversity-aware AI, by incorporating representative data and using bias-detection techniques, can work to prevent these discriminatory outcomes, ensuring that AI systems serve all people equitably (Chauhan and Kshetri 2024).

Beyond preventing harm, diversity-aware AI is essential for fostering innovation and expanding the scope of AI applications (Møller 2023). When AI is designed with diversity in mind, it is more adaptable to various contexts and environments, enabling it to address a wider range of challenges. For instance, natural language processing (NLP) systems that consider linguistic diversity are better equipped to understand and serve multilingual populations, improving accessibility for non-native speakers or people from diverse dialect backgrounds. This broader applicability not only enhances user experience but also stimulates innovation, as developers uncover new use cases and solutions for previously overlooked communities. The inclusive design of AI systems thus drives the field forward, broadening the reach and potential of AI technology (Drabiak 2024).

Diversity-aware AI also plays a crucial role in building trust between technology providers and the communities they serve. Public trust in AI has been shaken by high-profile instances of bias, from facial recognition systems with higher error rates for minority groups to recruitment algorithms that inadvertently favor certain demographics over others. These incidents undermine confidence in AI's fairness and raise concerns about its influence on decision-making processes that directly impact people's lives (Zaid et al. 2022). By developing diversity-aware AI, organizations can demonstrate a commitment to ethical and inclusive practices, which can help rebuild trust. Transparent, diversity-focused design and evaluation practices signal to the public that AI systems are developed responsibly, with an emphasis on respecting individual rights and reducing bias (Roche et al. 2023).

Moreover, diversity-aware AI has the potential to address broader social inequalities (Yin et al. 2023). In fields like employment, education, and criminal justice, AI is increasingly used to make or inform decisions with profound impacts on individuals' lives. Without attention to diversity, these systems can unintentionally reinforce existing biases within these sectors, exacerbating disparities rather than alleviating them. However, with a diversity-aware approach, AI can become a tool for promoting equity (Yeung 2017). For instance, fairer algorithms in hiring can open opportunities for historically underrepresented groups, while diversity-aware AI in education can provide tailored resources that cater to diverse learning needs. As AI becomes intertwined with society, it has the potential not only to reflect but also to shape social structures, making its role in equity vital (Jora et al. 2022).

Another significant aspect of diversity-aware AI is its alignment with ethical principles. The ethical use of AI is increasingly becoming a priority for stakeholders worldwide, including policymakers, tech companies, and consumers. A diversity-aware approach to AI aligns with principles of fairness, respect for human dignity, and social responsibility. It recognizes the importance of designing technology that honors diverse perspectives and experiences, acknowledging that inclusivity is an ethical imperative in any society that values equity (Heitz et al. 2022). As AI ethics guidelines proliferate, diversity awareness is emerging as a cornerstone for responsible AI, making it integral to the future regulatory landscape and guiding companies in sustainable AI practices (Zowghi and Mahmud 2024).

Diversity-aware AI is important because it represents a more inclusive vision of technological progress. The digital revolution should be for everyone and ensuring that AI systems consider diversity embodies this principle. Technology shapes societies in powerful ways, and diversity-aware AI ensures that all groups have a voice and presence in the digital future. It enables individuals from various backgrounds to participate fully and equitably in an AI-driven world, reducing digital divides and fostering a more inclusive global community (Jang et al. 2022). By embracing diversity-aware AI, we take a step toward a world where technology advances without sacrificing fairness or equity, paving the way for a digital landscape that uplifts and empowers all of humanity (Møller 2023).

4 Ethical and operational challenges in diversity-aware AI

Diversity-aware AI seeks to create more inclusive and representative systems by recognizing and supporting the needs of diverse user groups. These systems aim to respect cultural, social, and demographic differences, fostering equity across various applications. However, designing AI that meaningfully incorporates diversity is complex, raising significant ethical and operational challenges related to bias, transparency, privacy, and practical implementation.

4.1 Bias in data and algorithmic decision-making

One of the most pressing concerns in diversity-aware AI is bias within training data (Shin 2025). AI systems learn from vast datasets that often reflect historical prejudices and institutional inequalities (Søraa 2023). While diversity-aware AI is intended to counteract these biases, there is a risk that poorly designed interventions may inadvertently reinforce or amplify them. For example, a diversity-aware hiring algorithm that aims to improve gender balance might still rely on biased historical data that favors certain genders or job profiles, leading to unintended discriminatory outcomes (Yu et al. 2024). Similarly, overcorrecting for diversity could result in reverse discrimination, where the pursuit of diversity results in the exclusion of certain groups, creating an ethical tension between inclusivity and fairness (Shin 2025). Addressing these biases requires careful dataset auditing, fairness-aware learning techniques, and ongoing evaluation to prevent perpetuating existing inequalities.

Algorithmic bias also arises from model design choices and developer assumptions. Predictive models, particularly in high-stakes areas such as healthcare or criminal justice, may unintentionally produce inequitable outcomes for underrepresented groups. Predictive policing tools trained on historically biased crime data can reinforce racial disparities by targeting communities that were historically over-policed. Bias can emerge not only from the data but also from the assumptions and design choices of developers, who may unconsciously introduce their own perspectives into the model (Yu et al. 2024). Bias-mitigation strategies such as adversarial debiasing and fairness-aware learning are being explored, but they often involve trade-offs that can affect model accuracy or introduce new complexities into the system.

4.2 Challenges in defining and measuring diversity

A fundamental challenge in diversity-aware AI is determining what constitutes "diversity" in algorithmic decision-making. Diversity encompasses a wide range of dimensions, including race, gender, socioeconomic background, language, culture, and disability status. Attempting to quantify and operationalize diversity within AI models is inherently complex (Yu et al. 2024). There is a risk of reducing diversity to a set of predefined demographic categories, which may fail to capture intersectional identities or lived experiences (Cachat-Rosset and Klarsfeld 2023). Additionally, AI systems that rely on rigid diversity metrics may engage in tokenism, where representation is prioritized without meaningful inclusion or consideration of underlying structural inequalities (Li and Liu 2021).

4.3 Transparency, explainability, and accountability

Many AI systems operate as black boxes, making it difficult for users to understand how diversity considerations influence decision-making. This lack of transparency can lead to distrust in AI recommendations, particularly in high-stakes areas such as hiring, healthcare, and finance (Kim and Pasek 2020). If individuals cannot verify why an AI system made a particular decision, they may question whether diversity was incorporated fairly or if it was used in a way that compromised merit-based evaluations (Loecherbach et al. 2020). Developing explainable AI (XAI) solutions that clarify how diversity-aware algorithms function remains a significant challenge (Lin et al. 2024).

4.4 Privacy and ethical considerations

To ensure diversity, AI systems often require access to demographic and personal data. However, collecting and using such information raises serious privacy concerns, particularly regarding compliance with regulations such as GDPR, FERPA, and COPPA (Aguirre et al. 2016). Users may be reluctant to disclose sensitive attributes such as race or disability status, and improperly handling such data could lead to privacy breaches or misuse (Zowghi and Mahmud 2024). The challenge lies in balancing the need for diversity awareness with ethical data governance, ensuring that user autonomy and data security remain protected (Shin 2025).

4.5 Risk of reinforcing stereotypes

Diversity-aware AI models may inadvertently reinforce stereotypical representations if they rely on historically biased training data. For example, an AI-powered recommendation system designed to showcase diverse cultural content might unintentionally promote clichéd or oversimplified depictions of different groups (Crawford and Paglen 2021). Similarly, AI-generated educational materials intended to reflect diverse perspectives may end up reinforcing stereotypical narratives, rather than fostering genuine inclusivity (Chen and Sundar 2024). Preventing such issues requires careful curation of training data and continuous auditing of AI outputs.

4.6 Trade-offs between personalization and diversity

Many AI-driven platforms, such as news recommendation systems and content curation engines, rely on personalization to enhance user experience. However, balancing personalization with diversity poses a fundamental challenge. While diversity-aware AI aims to expose users to a broader range of perspectives, excessive diversification could reduce the relevance of recommendations, leading to decreased user engagement (Mattis et al. 2022). On the other hand, prioritizing personalization may result in filter bubbles that reinforce existing viewpoints rather than broadening exposure (Jesse and Jannach 2021). Striking a balance between user preferences and diversity goals remains a key challenge in AI design (Sax 2022).

4.7 Institutional resistance and implementation barriers

Even when diversity-aware AI is designed effectively, institutional barriers can hinder its adoption. Organizations may resist implementing diversity-focused AI due to cost concerns, lack of technical expertise, or reluctance to change existing workflows (Jang et al. 2022). Additionally, some stakeholders may perceive diversity-aware AI as a politically sensitive or controversial issue, leading to resistance in integrating such systems into decision-making processes (Heitz et al. 2022). Without clear incentives and regulatory frameworks, diversity-aware AI may struggle to gain widespread adoption (Jürgens and Stark 2022).

While diversity-aware AI presents a promising approach to addressing biases and promoting inclusivity, it is not without its risks. Algorithmic bias, privacy concerns, transparency limitations, and the challenges of defining diversity all pose significant obstacles. To ensure that diversity-aware AI serves its intended purpose, ongoing oversight, ethical AI development, and interdisciplinary collaboration are necessary. Developers must strike a careful balance between fairness, effectiveness, and user trust, ensuring that AI remains a tool for equity rather than an unintended source of new biases.

5 Case study: diversity-aware recommendation systems

Recommender systems (RS) play a critical role in shaping public opinion by acting as algorithmic gatekeepers of online content (Scheffauer et al. 2023). While they enhance user experience through personalization, concerns have emerged about their role in fostering misinformation, filter bubbles, and constrained perspectives (Knudsen 2023). Users often fall into algorithm-driven "rabbit holes" where they are exposed to reinforcing viewpoints rather than diverse perspectives (Møller 2023). To counteract these issues, researchers have explored solutions such as AI-driven nudges and algorithmic auditing to promote exposure to diverse content (Heitz et al. 2022).

News platforms utilize algorithmic nudges for content recommendations, but these nudges, while optimizing engagement, can inadvertently create echo chambers and partisan personalization (Cardenal et al. 2019; Bryanov et al. 2020). The relationship between RS and AI illustrates how hyper-personalized news, driven by behavioral and contextual data, can constrain rather than enhance diversity in news exposure (Jesse and Jannach 2021). As RS increasingly influence media consumption, there is a growing need to design systems that allow users to access a broader spectrum of perspectives beyond algorithmically reinforced biases (Sax 2022). Diversity-aware AI has emerged as a safeguard to ensure algorithmic personalization aligns with journalistic and societal values (Mattis et al. 2022).

5.1 Side effects of algorithmic personalization

Algorithmic personalization, while enhancing relevance, raises concerns about selective exposure, filter bubbles, and manipulation (Helberger 2019; Møller 2023). As RS increasingly tailor content based on past consumption, they create self-reinforcing cycles that limit user exposure to alternative viewpoints (Bastian et al. 2021). This tension between personalization and diversity poses a challenge: too much personalization leads to over-specialization, while excessive diversity risks reducing relevance (Sax 2022). Journalistic RS seek to balance these trade-offs through algorithmic nudges that promote diverse perspectives while maintaining engagement (Jürgens and Stark 2022). Platforms like CNN, BBC, and The New York Times integrate machine learning to personalize content while attempting to uphold journalistic integrity. However, commercial and engagement-driven algorithms often prioritize click-through rates over diverse content exposure, raising ethical concerns (Loecherbach et al. 2020). Research has shown that algorithmic nudges can increase news diversity, particularly when users face content overload or struggle to discover diverse perspectives on their own (Sonoda et al. 2022). Beyond accuracy, RS design must consider dimensions such as diversity (exposure to varied viewpoints), inclusion (representing different human perspectives), and equity (ensuring balanced news coverage) (Sax 2022). Achieving this balance requires transparency in AI design, user agency in content selection, and explainability in recommendation processes.

5.2 Nudging toward media pluralism and news diversity

The growing reliance on RS has amplified concerns about their role in shaping democratic discourse (Helberger 2019). News diversity remains a foundational principle in media, ensuring access to a broad spectrum of perspectives (Baden and Springer 2017). Algorithmic gatekeeping, while enhancing efficiency, can limit viewpoint diversity and create information cocoons (Cardenal et al. 2019). The challenge lies in designing RS that upholds public values while enabling user-centered personalization (Heitz et al. 2022). Algorithmic nudges offer a potential solution by embedding diversity-aware mechanisms in RS, ensuring that recommendations do not merely reinforce past behaviors but actively introduce a wider range of perspectives (Evans et al. 2022). However, despite increasing research in this area, the operationalization of diversity-aware recommendations remains underexplored. Future work should focus on designing AI systems that guide users toward diverse content ethically and transparently while protecting journalistic values.

5.3 A two-step flow of gatekeeping: why RS need more than personalization

RS operate on personalization principles, but over-personalization can lead to dissatisfaction and polarization (Mitova et al. 2022). Users may grow frustrated with highly tailored content and seek counter-attitudinal perspectives for balance (Monzer et al. 2020). This dynamic suggests a two-step flow of gatekeeping: first, AI personalizes content; second, algorithmic nudges guide users toward diverse viewpoints. Unlike traditional one-step gatekeeping, where editors filter news, this two-step model sees users co-construct their content experience with AI-driven nudges. This interactive process helps mitigate the risks of self-reinforcing personalization and supports exposure to diverse viewpoints (Currin et al. 2022). Personalization paradoxically increases users' desire for alternative perspectives, making diversity-aware RS crucial in preventing algorithmic echo chambers (Aguirre et al. 2016). To address these challenges, RS should incorporate explainability, adaptability, and mechanisms that allow users to control their content exposure (Evans et al. 2022). The goal is to create a system where users engage with diverse perspectives while maintaining the benefits of personalization. Future research should explore how algorithmic nudges can act as intermediaries between readers and news organizations, promoting transparency while balancing personalization with diversity. While diversity-aware AI can enhance news exposure and counteract filter bubbles, it must be carefully balanced with safeguards that maintain news accuracy, mitigate misinformation risks, and respect user preferences. Integrating credibility filters, transparency in recommendation logic, and user-driven customization can help navigate these trade-offs effectively.

6 Conclusion

Diversity in AI remains a challenge as algorithms personalize our news feeds, make news recommendations, and target algorithmic curations (Heitz et al. 2022). Scholars and experts echo the need for diversity and transparency in AI and algorithmic technologies in general (Jürgens and Stark 2022). This study highlights how diversity nudges affect users and steer algorithms to enact choice architectures and nudges that influence user behavior relating to diverse news. While the various principles of algorithmic nudges account for recent developments in AI media, an essential proposition in algorithmic nudges is that users should understand the logic of algorithms (transparency), engage in the personalization process (two-tier processes), and remain in control of such nudges (Zaid et al. 2022). Although algorithmic gatekeeping is increasingly pervasive and embedded in many media services, it also creates unintended consequences where journalistic value and moral responsibility for their nudges cannot be suitably attributed to any particular editorial or contextual factors. Algorithmic nudging should enable users to make better news selections and consumption decisions by facilitating informed cognitive processes, extending engagement to construct data, and augmenting users' literacy to utilize insights from the data. It is critical to pay close attention to the complicated societal context within which algorithmic nudges are used and deployed to prevent algorithmic nudges from progressing past the limited perspective of traditional nudging as a simple user interface in AI environments. Equally important is to design algorithmic nudges in diversity-aware AI as a platform for marketplace ideas and transparent user-centered mechanisms that contribute to overcoming users' emotional, cognitive, and psychological limits when they make decisions and perform actions that can promote discourse and, ultimately strengthen democratic reflection in algorithmic media. To benefit from AI systems, users must engage with various views and diverse opinions in ways that translate into pro-social media effects. Thus, our study has valuable implications for how we can redefine personalization conceptually and design diversity-aware AI practically to promote diverse news consumption.

7 Where do we go from diversity-aware AI?

The future of diversity-aware AI is not just about enhancing algorithmic performance—it is about reshaping the very foundation of how technology interacts with and represents human diversity. As AI continues to influence critical areas such as healthcare, education, governance, and media, its ability to adapt to diverse cultural, social, and individual contexts will determine its long-term success and ethical viability (Chauhan and Kshetri 2024). Diversity-aware AI must move beyond static models and rigid demographic categorizations to embrace more fluid, contextual, and intersectional understandings of identity and inclusion (Achón et al. 2024).

However, realizing the full potential of diversity-aware AI requires a paradigm shift in how we design, deploy, and govern these systems. Bias mitigation and transparency must become foundational elements of AI development, ensuring that models are not only fair but also accountable to the communities they impact. Emerging advancements in explainable AI, fairness-aware learning, and adaptive intelligence offer promising pathways, but they must be accompanied by institutional commitments to ethical oversight and interdisciplinary collaboration. No single discipline can tackle the challenges of diversity-aware AI alone—integrating expertise from computer science, social sciences, law, and ethics will be essential to developing systems that are both technologically robust and socially responsible.

At the same time, the global AI ecosystem must recognize that diversity-aware AI is not a mere technical challenge but a societal imperative. Regulatory frameworks, corporate policies, and public discourse will shape the trajectory of AI's impact on inclusion and equity (Zowghi and Mahmud 2024). As AI regulations evolve worldwide, ensuring that diversity-awareness aligns with fairness, transparency, and accountability will be crucial. Governments, organizations, and research communities must work together to establish guidelines that foster ethical AI while preventing the reinforcement of systemic biases.

Diversity-aware AI is about more than just building better algorithms—it is about shaping a future where technology serves all of humanity equitably. If developed responsibly, it has the potential to bridge social divides, promote inclusive decision-making, and empower marginalized communities (Crawford and Paglen 2021). The challenge ahead is not just to design AI that recognizes diversity but to ensure that it actively contributes to a more just, fair, and inclusive society (Shin 2025). The choices we make today in AI design, policy, and governance will define whether diversity-aware AI fulfills its promise or perpetuates existing inequalities. The responsibility lies with all of us to build AI systems that not only reflect the world as it is but also help create the world as it should be.

Author contribution Single author. Dr. Shin wrote the entire article.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

Achon L, Souza A, Hume A, Cernuzzi L (2024) A diversity-aware recommendation system for tutoring. AI Commun 37(1):1–23. https://doi.org/10.3233/AIC-230434

Aguirre E, Roggeveen A, Grewal D, Wetzels M (2016) The personalization-privacy paradox. J Consum Market 33(2):98–110. https://doi.org/10.1108/JCM-06-2015-1458

Baden C, Springer N (2017) Conceptualizing viewpoint diversity in news discourse. Journalism 18(2):176–194. https://doi.org/10.1177/1464884915605028

Baumer E (2017) Toward human-centered algorithm design. Big Data Soc 4(2). https://doi.org/10.1177/2053951717718854

Bastian M, Helberger N, Wijermars M (2021) Safeguarding the journalistic DNA: attitudes towards the role of professional values in algorithmic news recommender designs. Digital J 9(6):835–863. https://doi.org/10.1080/21670811.2021.1912622

Bryanov K, Watson B, Pingree R et al (2020) Effects of partisan personalization in a news portal experiment. Publ Opin Q 84(S1):216–235

Cachat-Rosset G, Klarsfeld A (2023) Diversity, equity, and inclusion in artificial intelligence: an evaluation of guidelines. Appl Artif Intell 37(1):2176618. https://doi.org/10.1080/08839514.2023.2176618

Campo-Ruiz I (2025) Artificial intelligence may affect diversity: architecture and cultural context reflected through ChatGPT, Midjourney, and Google Maps. Human Soc Sci Commun 12:24. https://doi.org/10.1057/s41599-024-03968-5

Cardenal S, Aguilar-Paredes C, Cristancho C, Majó-Vázquez S (2019) Echo-chambers in online news consumption. Euro J Commun 34(4):360–376. https://doi.org/10.1177/0267323119844409

Chen C, Sundar SS (2024) Communicating and combating algorithmic bias: Effects of data diversity, labeler diversity, performance bias, and user feedback on AI trust. Human–Comput Interact. https://doi.org/10.1080/07370024.2024.2392494

Crawford K, Paglen T (2021) Excavating AI: the politics of images in machine learning training sets. AI Soc 36(4):1159–1171

Chauhan PS, Kshetri N (2024) The role of data and artificial intelligence in driving diversity, equity, and inclusion. IEEE Access 12:37829–37843

Currin C, Vera S, Khaledi-Nasab A (2022) Depolarization of echo chambers by random dynamical nudge. Sci Rep 12:9234. https://doi.org/10.1038/s41598-022-12494-w

Drabiak K (2024) AI and machine learning ethics, law, diversity, and global impact. Bioethical Inquiry 16(1):1–17. https://doi.org/10.1259/bjr.20220934

Du Y, Ranwez S, Sutton-Charani N, Ranwez V (2021) Is diversity optimization always suitable? Toward a better understanding of diversity within recommendation approaches. Inf Process Manage 58(6):102721. https://doi.org/10.1016/j.ipm.2021.102721

Evans R, Jackson D, Murphy J (2022) Google News and machine gatekeepers. Digital J 11(9):1682–1700. https://doi.org/10.1080/21670811.2022.2055596

Hanna A, Denton E, Smart A, Smith-Loud J (2020) Towards a critical race methodology in algorithmic fairness. In: Proceedings of the 2020 conference on fairness, accountability, and transparency (FAT* 2020), pp 501–512

Holstein K, Wortman Vaughan J, Daumé III H, Dudik M, Wallach H (2019) Improving fairness in machine learning systems: what do industry practitioners need? In: Proceedings of the 2019 CHI conference on human factors in computing systems, pp 1–16

Heitz L, Lischka J, Birrer A, Paudel B, Tolmeijer S, Laugwitz L, Bernstein A (2022) Benefits of diverse news recommendations for democracy. Dig J 10(10):1710–1730. https://doi.org/10.1080/21670811.2021.2021804

Helberger N (2019) On the democratic role of news recommenders. Digital J 7(8):993–1012. https://doi.org/10.1080/21670811.2019.1623700

Hermann E (2022) Artificial intelligence and mass personalization of communication content. New Media Soc 24(5):1258–1277. https://doi.org/10.1177/14614448211022702

Jang W, Chun J, Kim S, Kang Y (2022) The effects of anthropomorphism on how people evaluate algorithm-written news. Digital J. https://doi.org/10.1080/21670811.2021.1976064

Jesse M, Jannach D (2021) Digital nudging with recommender systems. Comput Human Behav Rep 3:100052. https://doi.org/10.1016/j.chbr.2020.100052

Jora RB, Sodhi KK, Mittal P, Saxena P (2022) Role of artificial intelligence in meeting diversity, equality, and inclusion (DEI) goals. In: 2022 8th international conference on advanced computing and communication systems (ICACCS), pp 1687–1690. https://doi.org/10.1109/ICACCS54159.2022.9785266

Jürgens P, Stark B (2022) Mapping exposure diversity. J Commun 72(3):322–344. https://doi.org/10.1093/joc/jqac009

Jui TD, Rivas P (2024) Fairness issues, current approaches, and challenges in machine learning models. Int J Mach Learn Cyber 15:3095–3125. https://doi.org/10.1007/s13042-023-020

Kim D, Pasek J (2020) Explaining the diversity deficit. Commun Res 47(1):29–54. https://doi.org/10.1177/0093650216644647

Knudsen E (2023) Modeling news recommender systems' conditional effects on selective exposure. J Commun. https://doi.org/10.1093/joc/jqac047

Li B, Liu L (2021) Counteracting bias amplification through fairness-aware deep learning. IEEE Trans Neural Netw Learn Syst 32(8):3441–3455

Lin Z, Guan S, Zhang W et al (2024) Towards trustworthy LLMs: a review on debiasing and dehallucinating in large language models. Artif Intell Res 57:243. https://doi.org/10.1007/s10462-024-10896-y

Loecherbach F, Moeller J, Trilling D, van Atteveldt W (2020) The unified framework of media diversity. Digital J 8(5):605–642. https://doi.org/10.1080/21670811.2020.1764374

Mattis N, Masur P, Möller J, van Atteveldt W (2022) Nudging towards news diversity. New Media Society. https://doi.org/10.1177/14614448221104413

Mitova E, Blassnig S, Strikovic E, Urman A, Hannak A, de Vreese CH, Esser F (2022) News recommender systems. Ann Int Commun Assoc: 1–30. https://doi.org/10.1080/23808985.2022.2142149

Møller L (2023) Designing algorithmic editors. Digital J. https://doi.org/10.1080/21670811.2023.2215832

Monzer C, Moeller J, Helberger N, Eskens S (2020) User perspectives on the news personalization process. Digital J 8(9):1142–1162. https://doi.org/10.1080/21670811.2020.1773291

Noble S (2018) Algorithms of oppression: how search engines reinforce racism. NYU Press

Roche C, Wall PJ, Lewis D (2023) Ethics and diversity in artificial intelligence policies, strategies and initiatives. AI Ethics 3:1095–1115. https://doi.org/10.1007/s43681-022-00218-9

Sax M (2022) Algorithmic news diversity and democratic theory. Digital J 10(10):1650–1670. https://doi.org/10.1080/21670811.2022.2114919

Scheffauer R, Goyanes M, Gil de Zúñiga H (2023) Social media algorithmic versus professional journalists' news selection. Journalism. https://doi.org/10.1177/14648849231117980

Shams RA, Zowghi D, Bano M (2025) AI and the quest for diversity and inclusion: a systematic literature review. AI Ethics 5:411–438. https://doi.org/10.1007/s43681-023-00362-w

Shin D (2025) Debiasing AI: rethinking the intersection of innovation and sustainability. Routledge

Sonoda A, Seki Y, Toriumi F (2022) Analyzing user engagement in news application considering popularity diversity and content diversity. J Comput Social Sci 5:1595–1614. https://doi.org/10.1007/s42001-022-00179-3

Søraa RA (2023) AI for diversity. CRC Press. Routledge

Umbrello S, van de Poel I (2021) Mapping value sensitive design onto AI for social good principles. AI Ethics 1(3):283–296. https://doi.org/10.1007/s43681-021-00038-3

van Esch P, Cui Y, Heilgenberg K (2024) Using AI to implement diversity, equity and inclusion (DEI) into marketing materials. Austral Market J 32(3):250–262. https://doi.org/10.1177/14413582241244504

Werder K, Cao L, Ramesh B, Park EH (2024) Empower diversity in AI development: Diversity practices that mitigate social biases from creeping into your AI. Commun ACM 67(12):31–34. https://doi.org/10.1145/3676885

Yeung K (2017) Hyper nudge: big data as a mode of regulation by design. Inf Commun Soc 20(1):118–136

Yin K, Fang X, Chen B, Sheng L (2023) Diversity preference-aware link recommendation for online social networks. Inf Syst Res 34(4). https://doi.org/10.1287/isre.2022.1174

Yu X, Gao Z, Zhao C, Qiao Y, Chai Z, Mo Z, Yang Y (2024) Diversity-aware unbiased device selection for federated learning on non-IID and unbalanced data. J Syst Archit 156:103280

Zaid B, Biocca F, Rasul A (2022) In platforms we trust? J Broadcast Electron Media 66(2):235–256. https://doi.org/10.1080/08838151.2022.2057984

Zhao Y, Wang Y, Liu Y, Cheng X, Aggarwal C, Derr T (2024) Fairness and diversity in recommender systems. ACM Trans Intell Syst Technol. https://doi.org/10.1145/3664928

Zhou S (2024) A value and diversity-aware news recommendation systems: can algorithmic gatekeeping nudge readers to view diverse news? J Mass Commun Q. https://doi.org/10.1177/10776990241246668

Zowghi D, Mahmud S (2024) AI for all: identifying AI incidents related to diversity and inclusion. AI Ethics

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.