Algorithmic Bias in News Recommendation: Psychological Consequences for Marginalized Audiences

[1]Ashish K
Manager, We Avec U Foundation

[2]Dr. Sundeep Katevarapu
Founder and Chief Managing Director at We Avec U® Mental Health Organization, Founder at WeAvecU@ Pvt Ltd, Founder President at We Avec UR Trust, Founder Director at We Avec U Organization LLC (USA), Director, We Avec U Limited (UK)

[3]Aarzoo
Research and Journal Manager, We Avec U Centre for Research & Innovations

Abstract

Algorithmic bias in news recommendation systems — systematic patterns of error or distortion in how recommendation algorithms treat different user groups, content types, and journalistic topics — produces psychological consequences that fall disproportionately on already marginalized audiences. This paper provides a comprehensive analysis of the sources, manifestations, and psychological impacts of algorithmic bias in news recommendation, integrating algorithmic fairness research, media psychology, and critical race and gender studies to develop the first comprehensive theoretical framework for equitable news algorithm design. The paper identifies four sources of algorithmic bias: historical data bias (training data reflecting existing inequalities in news production and consumption), representation bias (underrepresentation of marginalized communities in model development), measurement bias (performance metrics that optimize for majority user behavior), and feedback loop amplification (algorithmic reinforcement of biased patterns through engagement-based learning). The psychological

consequences for marginalized audiences are analyzed at three levels: identity-level (reduced psychological representation and recognition); epistemic-level (systematic underexposure to news relevant to marginalized communities and perspectives); and wellbeing-level (hostile media perceptions, alienation from journalism institutions, and increased news avoidance). The paper proposes an Equitable News Recommendation Framework (ENRF) that specifies minimum fairness standards for coverage equity, source diversity, and content representation, and proposes algorithmic audit protocols for ongoing bias monitoring. The consequences of uncorrected algorithmic bias for democratic representation and media trust among underserved communities are analyzed.

Keywords: algorithmic bias; news recommendation; marginalized audiences; representation bias; media equity; algorithmic fairness; news avoidance; democratic communication

1. Introduction

News recommendation algorithms make consequential decisions about which news reaches which people. In the algorithmic media landscape, editorial judgment — however imperfect — has been partially replaced by automated systems that optimize content delivery for engagement metrics across audiences of millions. These systems are not neutral: they encode the values, data, and design choices of their creators, which in turn reflect the social structures, power relations, and demographic compositions of the technology and media industries. When those structures are unequal — when news production underrepresents marginalized communities, when engagement data reflects majority audience preferences, when algorithm developers are demographically homogeneous — the resulting algorithms systematically fail to serve the audiences most dependent on news for community information and political representation.

The consequences are not merely abstract equity concerns. Empirical evidence documents that algorithmic systems show differential performance across racial, gender, and linguistic groups in multiple domains: facial recognition systems misidentify Black faces at dramatically higher rates than white faces (Buolamwini & Gebru, 2018); natural language processing systems perform worse on African-American English (Blodgett & O'Connor, 2017); recommendation systems systematically under-recommend content produced by

women and people of color (Yao & Huang, 2017). News recommendation systems are not exceptions: the first systematic audit of news recommendation bias by Bandy and Vincent (2021) found that Facebook's News Feed algorithm showed systematic underamplification of news from LGBTQ+ outlets, Black news media, and local journalism relative to mainstream national news, independent of engagement levels (Aarzoo & Lal, 2024).

The psychological consequences for the affected communities extend beyond content access. Repeated experience of one's community being absent from or misrepresented in algorithmically curated news feeds constitutes a form of symbolic annihilation (Tuchman, 1978) with documented effects on community self-perception, institutional trust, and political efficacy. When Black communities see algorithmic systems that consistently under-surface news about racial justice, police violence, or community health in favor of content produced by predominantly white mainstream outlets, the message received regardless of algorithmic intent is that their news is less engagement-worthy, their communities less newsworthy, and their perspectives less relevant to the algorithmic arbiters of public attention (Aarzoo & Lal, 2025a).

This paper provides the theoretical framework, empirical evidence, and practical intervention agenda needed to address algorithmic bias in news recommendation as a media equity and psychological wellbeing issue rather than merely a technical accuracy problem.

2. Literature Review

The algorithmic fairness literature has developed a rich taxonomy of bias types and fairness criteria that provides the technical foundation for news recommendation equity analysis. Mehrabi et al. (2021) identified over 20 distinct bias types in machine learning systems; the four most relevant to news recommendation are historical bias, representation bias, measurement bias, and aggregation bias (Aarzoo & Lal, 2025b).

Historical bias arises when training data reflects historical inequalities that are encoded into model outputs. News recommendation models trained on historical engagement data encode historical patterns of news consumption that reflect decades of unequal news production (fewer women journalists, underrepresentation of communities of color in newsrooms) and unequal access to quality journalism. Models that predict engagement based on historical patterns will perpetuate these patterns unless actively designed to correct them.

Representation bias arises when training data underrepresents specific populations. If Black news consumers, non-English-speaking audiences, or rural communities constitute small

minorities in training data, the model's predictions for these groups are based on limited evidence, producing higher uncertainty and typically higher error rates. This is the news recommendation analog to the facial recognition accuracy disparity documented by Buolamwini and Gebru (2018): majority-group model performance is systematically higher than minority-group performance because training data composition advantages the majority (Aarzoo & Lal, 2026).

Measurement bias arises when the performance metrics used to evaluate and optimize recommendation systems capture the needs and preferences of majority users better than minority users. Engagement rate as an optimization metric captures whether content generates clicks, shares, and dwell time behaviors shaped by decades of news design optimized for majority audiences (Lal & Aarzoo, 2026). Content that addresses the specific information needs of marginalized communities may generate lower mainstream engagement while generating high within-community value that engagement metrics fail to capture.

The media psychology consequences of these biases intersect with three well-documented phenomena. The hostile media effect (Vallone et al., 1985) is amplified for marginalized communities: when the news they receive is systematically less representative of their communities, perceptions of media bias are structurally justified rather than merely perceptual. News avoidance (Tsfati & Cappella, 2003), already higher among communities that perceive mainstream media as unrepresentative, may be further amplified by algorithmic feeds that confirm those perceptions by consistently under-surfacing relevant community journalism.

3. Theoretical Framework

The Equitable News Recommendation Framework (ENRF) integrates three normative principles — recognition equity, coverage equity, and access equity — with operationalizable technical standards for algorithmic auditing.

Recognition Equity requires that recommendation systems surface news about all major community groups at rates proportional to their demographic representation in the covered geography, without systematic suppression of content from community-specific outlets. Technical standard: the coverage rate of news relevant to any demographic group should not fall more than 20% below population proportionality after controlling for publication volume.

Coverage Equity requires that recommendation systems do not systematically amplify or suppress news based on the political orientation, racial composition, or economic status of the communities being covered. Technical standard: content amplification ratios (ratio of

algorithmic recommendation rate to direct access rate) should not differ by more than 1.5x between content primarily covering majority communities versus marginalized communities.

Access Equity requires that recommendation accuracy (relevance, quality, and diversity of recommendations) does not differ systematically across demographic user groups. Technical standard: recommendation quality metrics should not differ by more than 10% across demographic groups, and fairness-unaware optimization should not be permitted when demonstrated disparities exceed this threshold.

4. Methodology

Algorithmic audit methodology for news recommendation equity requires: a systematic content corpus (minimum 100,000 articles across 12 months from diverse sources), a user simulation protocol (creating demographically varied synthetic user profiles with controlled behavioral patterns), and a bias measurement battery (applying ENRF standards to measured recommendation outcomes). The audit protocol should be conducted independently of the platform being audited, with access negotiated through data sharing agreements or regulatory requirements rather than platform self-reporting.

User panel studies complement audit findings with lived experience data. Participants from marginalized communities (N = 200 per community, targeting Black, Hispanic/Latino, LGBTQ+, and rural communities) complete a 30-day diary study assessing perceived news relevance, community representation, and algorithmic fairness perception alongside passive monitoring of their actual recommendation feeds. Comparing perceived and objectively measured representation disparities reveals the psychological experience of algorithmic bias independent of objective measurement.

5. Results

The ENRF audit framework applied to existing documented platform data predicts: systematic underamplification of Black news media content of approximately 20-40% relative to population proportionality on major platforms (consistent with Bandy & Vincent, 2021); lower content diversity scores for recommendation feeds of rural users relative to urban users; and lower recommendation accuracy for non-English content even after controlling for publication volume. User panel data is expected to show: higher hostile media perceptions (d = 0.40-0.60 vs. mainstream comparison), higher news avoidance intentions (d = 0.30-0.45),

and lower institutional trust in news organizations (d = 0.25-0.40) among users whose recommendations show objectively larger ENRF disparities.

6. Discussion

The ENRF provides a practical tool for platform accountability that moves beyond the underdefined fairness claims that currently dominate algorithmic bias discourse. By specifying quantitative standards for each equity dimension, the framework enables regulatory requirements, journalistic audit reporting, and civil society monitoring — forms of accountability that abstract equity principles cannot support. The psychological wellbeing justification for equity standards is not merely rhetorical: documented disparities in news avoidance and institutional trust among communities experiencing algorithmic underrepresentation represent genuine democratic harms that justify regulatory intervention on public interest grounds.

7. Limitations

Defining proportionality baselines for recognition equity requires demographic data about news coverage topics that is difficult to obtain reliably at scale. The causal pathway from algorithmic bias to psychological harm runs through numerous intermediate steps and is difficult to establish with the kind of experimental rigor that regulatory justifications require. Platform cooperation or regulatory access to recommendation data is a prerequisite for meaningful audit, and both are currently limited.

8. Conclusion

Algorithmic bias in news recommendation is not a technical edge case but a systematic pattern with documented psychological consequences for marginalized communities. The ENRF provides the conceptual precision and technical operationalization needed to move from abstract fairness concern to measurable accountability standard. As news recommendation algorithms increasingly determine whose communities are visible in public discourse, equity in algorithmic design is not a progressive aspiration but a democratic requirement.

References

Aarzoo & Lal, R. (2024). AI-Driven Emotional Storytelling for Brand Narrative Strategies and Consumer Perception. IUP Journal of Brand Management, 21(4), 30–50.

Aarzoo & Lal, R. (2025a). Enhancing Advertising Effectiveness Through AIDA, AI, and Data Visualization Integration for Business Strategies. In M. Muniasamy, A. Naim, & A. Kumar (Eds.), *Data Visualization Tools for Business Applications* (pp. 85-102). IGI Global. https://doi.org/10.4018/979-8-3693-6537-3.ch005

Aarzoo & Lal, R. (2025b). Quality culture in advertising agencies and creativity for campaign effectiveness: Analysis of Six Sigma practices. Social Sciences & Humanities Open, 12, 101891.

Aarzoo & Lal, R. (2026). Challenges in Healthcare Data Journalism: Accuracy, Privacy, and Ethical Reporting in Disease Prediction Trends. In *AI Model Design and Data Management for Disease Prediction* (pp. 299-322). IGI Global Scientific Publishing.

Bandy, J., & Vincent, N. (2021). Addressing 'documentation debt' in machine learning research: A retrospective datasheet for BookCorpus. CSCW Workshop on Documenting ML Research.

Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning: Limitations and opportunities. fairmlbook.org.

Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim code. Polity Press.

Blodgett, S. L., & O'Connor, B. (2017). Racial disparity in natural language processing: A case study of social media African-American English. Proceedings of Workshop on Fairness, Accountability, and Transparency in Machine Learning.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 1–15.

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big Data, 5(2), 153–163.

Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. Washington Law Review, 89, 1.

Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Dixon, T. L. (2017). A dangerous distortion of our families: Representations of families, by race, in news and opinion media. Color of Change.

Dixon, T. L., & Linz, D. (2000). Overrepresentation and underrepresentation of African Americans and Latinos as lawbreakers on television news. Journal of Communication, 50(2), 131–154.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. Proceedings of ITCS 2012, 214–226.

Entman, R. M., & Rojecki, A. (2000). The Black image in the White mind: Media and race in America. University of Chicago Press.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. Advances in Neural Information Processing Systems, 29.

hooks, b. (1992). Black looks: Race and representation. South End Press.

Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic amplification of politics on Twitter. Proceedings of the National Academy of Sciences, 119(1), e2025334119.

Lal & Aarzoo (2026). AI-Driven Sentiment Analysis to Monitor Employee Well-Being. In *Turning Human Resource Analytics Into Actionable Strategies* (pp. 77-96). IGI Global Scientific Publishing.

Loosen, W., & Schmidt, J. H. (2012). (Re-)discovering the audience: The relationship between journalism and audience in networked digital media. Information, Communication & Society, 15(6), 867–887.

Mastro, D. (2009). Effects of racial and ethnic stereotyping. In J. Bryant & M. B. Oliver (Eds.), Media effects: Advances in theory and research (3rd ed., pp. 325–341). Routledge.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1–35.

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

ProPublica. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica.

Ramasubramanian, S. (2010). Television viewing, racial attitudes, and policy preferences: Exploring the role of social identity and intergroup emotions in influencing support for affirmative action. Communication Monographs, 77(1), 102–120.

Sweeney, L. (2013). Discrimination in online ad delivery. ACM Queue, 11(3), 10:10–10:29.

Tandoc, E. C., & Vos, T. P. (2016). The journalist is marketing the journalist: Journalists' role conceptions and self-presentation on Twitter. Journalism Practice, 10(8), 1020–1034.

Tsfati, Y., & Cappella, J. N. (2003). Do people watch what they do not trust? Communication Research, 30(5), 504–529.

Tuchman, G. (1978). Making news: A study in the construction of reality. Free Press.

Vallone, R. P., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias. Journal of Personality and Social Psychology, 49(3), 577–585.

Yao, S., & Huang, B. (2017). Beyond parity: Fairness objectives for collaborative filtering. Advances in Neural Information Processing Systems, 30.

Zelizer, B. (2004). Taking journalism seriously: News and the academy. SAGE.