The Review of Contemporary Scientific and Academic Studies
An International Multidisciplinary Online Journal
www.thercsas.com
ISSN: 2583-1380 Vol. 6 | Issue No. 4 | April 2026 Impact Factor: 6.53

Algorithmic Bias and AI: A Humanities Perspective

Dr. Aswathi M. P. (aswathimp@gmail.com), Associate Professor, Department of English,
KAHM Unity Women's College, Manjeri (Affiliated to University of Calicut), Malappuram, Kerala, India

 Crossref/DOI: https://doi.org/10.55454/rcsas.6.04.2026.001

Abstract: *Artificial intelligence plays a pivotal role in deciding the selection and governance in many critical domains such as education, employment, research, law and healthcare. This pervasive influence creates a crisis as they pose threat to the unbiased production of content. The phenomenon called algorithmic discrimination intensifies the inequalities in society creating a crisis in knowledge production and dissemination, particularly in humanities discipline. Spread in the pretext of objectivity, the data circulated will remain as partial truths or untruths. The paper explores the framework used by Humanities to interrogate the ethical, cultural and political landscape created by artificial intelligence. Drawing insights upon the data based on marginalization and theories such as feminism or cultural studies, the paper tries to unravel the digital injustice and cautions the harm that predictive technology does to the human mind. Grounded in the idea of inclusivity, the current paper challenges the notion of algorithmic neutrality by seeking answers to the following research questions: in what way does the social inequalities are addressed or identified by the algorithm, how can humanities identify the bias and create a critique of it, What are the historical, ethical and cultural dimensions of this issue and what contribution can be made from the interdisciplinary approach rooted in humanities to realign the principles of artificial intelligence and governance used by it. The methodology used for the study is basically textual analysis.*

Keywords: Algorithm Bias, Artificial Intelligence, Cyborg, Humanities

Article History: **Received: 08 April- 2026;** **Accepted: 15 April- 2026;** **Published/Available Online: 30 April- 2026**

1. Introduction

The post artificial intelligence era witnesses a different type of bias that the machine displays as if it sustains the unequal treatment of society based on class, colour, cate, gender or class. This is the phenomena which embarrasses the learning community, especially those from the humanities circle as this poses a threat or crisis to the intellectual vigour that the discipline has attained through its evolution. Addressing this bias is vital as the current system of governance rely greatly on the evaluation done by artificial intelligence which has its grip on all areas of life including healthcare, governance, education, marketing, employment, finance etc. Rather than diminishing the social disparities the algorithm which is working based on the data feed in a prejudiced manner producing results as outcomes of objective analysis. The reliance on this result will invariably hamper the possibilities of the attainment of egalitarianism in the society and hence is detrimental to the progress of humanity. The vulnerable ethnic groups, communities and individuals often fall prey to the newly emerged hazard. It is in this context of the bias in the machine that emerged the concept of responsible AI which provides the output in a rather unbiases way. To fulfil the emergence of such an ethical production of objective content, the humanities has to make a proactive intervention exploring the ethical, political, social and cultural implications of the bias of the machine and algorithm which is a systemic failure than the incidental error. A critical analysis of data histories, systemic inequalities initiated by humanities will catalyse such a move.

The assumption that the algorithms on online platforms work on an objective way is an absolute myth. The algorithms are primarily working based on the individual's choice to retain a particular content for watching. In addition to that the factors such as the collective choices of individuals is also playing a pivotal role in machine's selection of the content for gaze. The systemic priorities, gender, economic preferences, national agenda, monetary concerns etc. contribute to the bias of data driven technologies. In fact, algorithms largely depended on the availability of historical data and the institutions' take on how to use that to support the normal. The data which is used to train AI systems is the one which stabilizes the social hierarchies that support the masculine, white skinned, Euro-American citizens as the models based on which the results are largely produced.

As a powerful tool to interrogate the way how AI functions creating an illusion of objectivity, humanities utilises the tools of power and ethics, history and representation of identities. This is done analysing the social narratives underlying the choices of algorithms. Theoretical frameworks of cultural studies,

The Review of Contemporary Scientific and Academic Studies
An International Multidisciplinary Online Journal
www.thercsas.com
ISSN: 2583-1380 Vol. 6 | Issue No. 4 | April 2026 Impact Factor: 6.53

postcolonial theory and feminism are largely functioning as the media to catalyze this inquiry. In addition to exploring the modalities of the working of AI, humanities critically examine the politics and the power structures that channelize the biases in AI. Identifying essentialist models and rejecting their postulations presented as results or findings humanities makes a plea for the inclusivity and celebrates intersectionality. Theories such as feminism examine the manifestations of gender norms as portrayed by AI. Donna Haraway's *A Cyborg Manifesto* describes the complexities involved in the working of technology when she says:

> The cyborg is resolutely committed to partiality, irony, intimacy, and perversity. It is oppositional, utopian, and completely without innocence. No longer structured by the polarity of public and private, the cyborg defines a technological poll based partly on a revolution of social relations in the oikos, the household. Nature and culture are reworked; the one can no longer be the resource for appropriation or incorporation by the other. The relationships for forming wholes from parts, including those of polarity and hierarchical domination, are at issue in the cyborg world. (151)

According to Haraway, Cyborg is unpredictable and biased. It disrupts the system and takes sides by blurring many of the boundaries existing in the world. Along with that it creates new equations which may not be logical also. But at the same time, it imagines the rejection of essentialist models and strategies that consider human beings as mere data.

In *Epistemic Injustice: Power and the Ethics of Knowing* Mirand Fricker discusses testimonial and hermeneutical injustices which provide convincing explanations about the manifestation of harm in terms of knowledge that algorithms perpetrate by means of discrediting of testimony and overlooking collective interpretive sources. Bias in AI operates in various ways: at times it ignores human beings in terms of race or gender and does not consider this lack of consideration as lapse or lack of subjectivity. This homogenization only upholds the prerogatives of the dominant and negates the concerns of diverse groups and their plural issues. Even more dangerous versions of this bias lie in the prediction it makes making ethnicity as a ground for a person's possibility to become a criminal in future. "Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks." by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner published in *ProPublica* on May 23, 2016 discusses the partial vision of AI that anticipates the future involvement of blacks in crimes higher than the involvement of whites with no evidence to substantiate. Shockingly, testimonials and hermeneutic injustice were formed the basis for such predictions coming in the form of verdicts on human respectability of existence. While Youjin Kong's "(Un)Fairness in AI: An Intersectional Feminist Analysis" highlights gender bias in AI results , Timnit Gebru's work "Gender Shades" reiterates the gender bias and intersectionality connected with such a bias in it. After analysing the face recognition AI tools and algorithms provided by Microsoft and IBM, bias was seen as it found it easy to recognize white male faces compared to white female faces and white female faces in comparison with black female faces. Postcolonial theory also delves deep into this algorithmic bias as it makes a critique of the generalization of western priorities as normal and universal. Most of the times algorithms reiterate the specters of colonialism i.e. surveillance, control and manipulation. Ruha Benjamin's *Race After Technology* highlights the symbiotic relationship between society and technology to establish the argument that the technological determinism is false. Also, the book establishes the complexity connected with the pessimistic vision that technology perpetuates racism along with the utopian view that the technology establishes equality by supporting digital anonymity (Benjami 39).

2. Materials and Methods

The method used for this study is basically textual analysis.

3. Results

Large language models including GPT-2, GPT-3.5, GPT-4 by OpenAI, Llama 2 by Meta and Google Gemini are found to have been shown gender bias as reported in a UNESCO study titled, "Challenging systematic prejudices: an investigation into bias against women and girls in large language models" regarding regressive gender stereotyping made by these models. According to the report,

> more diverse, high-status jobs to men, such as engineer, teacher and doctor, while frequently relegating women to roles that are traditionally undervalued or socially-stigmatized, such as "domestic servant", "cook" and "prostitute" … -generated stories about boys and men dominated by the words "treasure", "woods", "sea", "adventurous", "decided" and "found", while stories about women made most frequent use of the words

https://doi.org/10.55454/rcsas.6.04.2026.001

The Review of Contemporary Scientific and Academic Studies
An International Multidisciplinary Online Journal
www.thercsas.com
ISSN: 2583-1380 Vol. 6 | Issue No. 4 | April 2026 Impact Factor: 6.53

"garden", "love", "felt," "gentle", "hair" and "husband". Women were also described as working in domestic roles four times more often than men in content produced by Llama 2 (1). The report states about the ethical flaws made by these language models citing examples of negative portrayals of transgender people and racial minorities.

The fundamental flaws of AI in many contexts are its oversimplified solutions to the complex issues. For example, in the selection of the right candidate from the resumes available for scrutiny, AI selects individuals based on their gender, ethnicity etc. This pitfall is the consequence of biased data which is used to train the system. The logic of efficient results produced with the use of artificial intelligence is under threat here. Human potential should be evaluated without reductionism considering the experience, ethics, context and narrative. The article titled, "AI, Gender Bias and Development" by Maria Burbach and Ilaria Mariotti published on UNDP website, it is argued that neural machine translation systems use masculine and feminine pronouns to reinforce gender stereotypes of individuals having superior status as a 'he' and inferior status as a 'she' (1). Extreme caution against these biases is what humanities perspective demand from the users. The LLMs also show biases regarding the scores given to the resumes of old women as they often rank the older women with equal qualification and experience of a male candidate below these male candidates with no proper justification as reported in *Nature* and in a report published by Standford university. In the health care sector, these biases based on skin colour and ethnicity often lead to negligence. As an instance, the skin cancer diagnosis is often failed using AI as these systems are producing results based on fair skinned images. The failure of accuracy factor in the case of darker skin led us to the apprehension that the AI models may classify the cancerous lesions wrongly as benign in darker-skinned individuals which resulted in lower possibility of survival. (Mehta 1). Again, a report from *Science* probes into an algorithm used to manage care for millions of patients in the U.S. which in a consistent manner ignored the health needs of Black patients because it relied on healthcare expenditure as a proxy for health status (Obermeyer et al. 447). Economic metrics are often mistaken to be the objective indicators when these AI tools as one observe this from the point of view of Humanities. The parameters such as access to healthcare, economy and insurance, discrimination on the front of diagnosis and treatment, systemic neglect are systematically underestimated to provide partial and hence faulty results. Humanities enquire how system should account for such inequities in the context of such technological disparities validated by machines. But we must always be cautious about whether a substantive intersectional fairness pattern is used instead of formal approach of dominant computer science based on statistical parity across subgroups. The advantage of such a policy framework is to extend such inquiries about intersectionality by asking questions about the possibility of algorithms in mitigating marginalization over attainment of formal equality.

4. Discussion

Humanities looks at these biases as deeply rooted in historical realities and sustained through diverse cultural and historical phenomena such as slavery, colonisation, world wars, surveillance and knowledge production. Humanities observe the rise of AI in the continuum. Surveillance at the times of AI, is similar though more sophisticated to the mechanism of control happened during colonial power. When looking through the lens of history, it has been observed that classification of population and predictive mapping were a part of establishing supremacy by learning about the subject. Simone Browne rightly argues in *Dark Matters* that the genealogy of surveillance must be traced back to the transatlantic slave trade and colonial archives, where Black bodies were catalogued and monitored (Browne 27). The analysis done in the light of humanities displays the evolution of such tools of categorisation and surveillance which manifest power in a rather subtle fashion. Invariably, the present-day technological innovations are rooted in the ideologies successfully worked in the past and later functioned as epistemes. Again, algorithms often encode able bodies ableist ideologies ignoring disabled individuals. This is effected through biases in the design and outcome or normalising ableism as efficiency which tantamount to systematic exclusion of disabled individuals as data.

It is crucial in this juncture to explore the implications of the working of these models in shaping the perception of humanity. This is to be analysed in the light of the popular use of AI as a second intellect where majority of human being stop thinking and reserve the task of reflection for AI to do. The reason behind such a human choice is the representation of AI in media as intelligent, objective, neutral and efficient. The biased data is presented in a positive empathetic and objective manner resulting in techno-utopian narratives that such representations feed into masking the irony it causes. At times media such as

https://doi.org/10.55454/rcsas.6.04.2026.001

The Review of Contemporary Scientific and Academic Studies
An International Multidisciplinary Online Journal
www.thercsas.com
ISSN: 2583-1380 Vol. 6 | Issue No. 4 | April 2026 Impact Factor: 6.53

movies address the ethical dilemma and complexity of representation of AI as depicted in movies such as *Ex Machina* or *Black Mirror.* These counter narratives offer a space for those who watch them to have a critique of the use of AI in the current times.

Now, the unique role of humanities lies in its potential to offer alternative way of development of AI with a reimagined focus on inclusivity, plurality, justice, and intersectionality. This is pivotal in forming policies that cater to this objective in areas such as education, governance and healthcare. The critical humanities component which is lacking in the data science curriculum of the times is to be redesigned including it. The software developers are trained to be aware of their ethical and moral responsibilities, the reflection of which is to be seen in the AI models that they make. Humanities can make a significant and sustainable contribution in the development of critical pedagogy for this purpose. As N. Katherine Hayles argues, "the humanities must intervene not only to critique but to transform technoscientific practices" (Hayles 85). In the field of AI education insistence is to be laid on the inclusion of algorithmic ethics, digital humanities and formation of a progressive society by means of technology. The change through this education envisaged is the creation of culturally sensitive and responsible scholarship. In addition to this, participatory design can yield more inclusive and just result, if these AAI models are trained using the ignored and omitted results incorporating insights from the feminist, postcolonial or any other methodology that aligns with the subaltern. Sasha Costanza-Chock argues for "design justice", a methodology that redesigns the process to centre the voices of the oppressed and challenge the status quo (Costanza-Chock 6). Humanities scholarship comes handy here by offering the framework and methodologies foregrounding dialogues against monologue and empathy against self-centered narratives.

5. Conclusion

The digital ambience that the humanities envisage is the one grounded in the concepts of empathy, equity, justice and inclusiveness. This can be made real by utilizing tolls that cater to the understanding of the myth of algorithmic neutrality, subtle operations of bias, prejudice and ignorance that govern the realm of AI. CARE principles i.e. collective benefits, authority to control, responsibility and ethics must be taken into consideration in the analysis to evaluate algorithmic bias as a decolonial issue of justice to know the digital epistemicide when the AI is trained using indigenous data without consent. Here, the productive role that the humanities scholars can play is to serve as the co-creators of these LLMs by engaging in redesigning system, making policies and educating the AI using the right data with a view to eradicate partiality and bring out the values of justice and equity.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine bias: There's software used across the country to predict future criminals. And it's biased against Blacks. ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code.* Polity Press.

Browne, S. (2015). *Dark matters: On the surveillance of Blackness.* Duke University Press.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research, 81*, 1–15. https://proceedings.mlr.press/v81/buolamwini18a.html

Burbach, M., & Mariotti, I. (2025, October 6). AI, gender bias and development. *United Nations Development Programme*. https://www.undp.org/eurasia/blog/ai-gender-bias-and-development

Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need.* MIT Press.

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

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

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001

Haraway, D. (1991). *A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century.* Routledge.

ISSN: 2583-1380 Vol. 6 | Issue No. 4 | April 2026 Impact Factor: 6.53

Hayles, N. K. (1999). *How we became posthuman: Virtual bodies in cybernetics, literature, and informatics.* University of Chicago Press.

Mehta, M. (2026, January 8). AI bias: 16 real AI bias examples & mitigation guide. *Crescendo*. https://www.crescendo.ai/blog/ai-bias-examples-mitigation-guide

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. https://doi.org/10.1126/science.aax2342

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown Publishing.

UNESCO. (2024). *Challenging systematic prejudices: An investigation into bias against women and girls in large language models. UNESCO Digital Library*. https://unesdoc.unesco.org/ark:/48223/pf0000388971

Conflicts of Interest: The author declares "No conflict of interest".