A DOCTRINAL STUDY ON MITIGATION OF ALGORITHMIC BIAS IN SOCIAL MEDIA PLATFORMS UNDER THE INDIAN CONSTITUTION AND THE INFORMATION TECHNOLOGY ACT, 2000

Avishikta Gupta, CHRIST (Deemed to be University)

ABSTRACT

This paper critically analyses the challenges of mitigating algorithmic bias in social media platforms. It argues that the existing statutes of the Information Technology Act, 2000, do not cover the complexities of AI-driven systems, which in turn violates the constitutional rights of an individual. The research further highlights how the fundamental right of freedom of speech and expression is being misused on social media platforms. The paper aims to build a nexus between the constitutional law and the Information Technology Act, thereby developing digital rights for people. Through the doctrinal and comparative analysis, this paper provides a framework to ensure technology aligns with constitutional morality and protects the rights of citizens in the digital age.

Keywords: algorithmic bias, constitutional rights, freedom of speech and expression, social media.

Introduction:

Standing today, social media is not just a mode of communication but a wider ambit that influences everyone. Complex algorithms heavily drive the pervasive influence of social media platforms. It has profoundly reshaped how we access social media, retrieve information, and interact with the world through it. This influence of social media holds the power to alter and create one's opinion on various world issues that can lead to communal hate, racism, and homophobia. Therefore, these powerful algorithms are not neutral; they affect and amplify societal biases and present their training data, which leads to algorithmic bias. The bias manifests in discriminatory outcomes, which in turn affect fundamental rights like equality and freedom of expression. The central idea of my research delves into the challenge of mitigating algorithmic bias in social media platforms. The two major challenges are: the lack of specific provisions to hold developers and platforms accountable for algorithmic bias under the Information Technology Act, 2000, and the misuse of freedom of speech and expression on social media, when amplified by algorithms, creates a new form of digital inequality that the current legal framework is ill-equipped to handle.

In India, where digital spaces are increasingly central to political participation, business, and cultural exchange, algorithmic bias threatens not only individual dignity but also constitutional guarantees of equality (Article 14), freedom of speech and expression (Article 19(1)(a)), and the right to life and personal liberty, including privacy (Article 21). Despite this, the Information Technology Act, 2000, and its subsequent rules offer only indirect protections, framing platforms as passive intermediaries rather than active shapers of discourse. This regulatory vacuum allows social media companies to misuse freedom of speech arguments to justify opaque algorithmic practices while escaping liability under Section 79 safe harbour provisions.

The urgency of this issue lies in its dual impact: it not only deepens social inequalities by amplifying existing prejudices but also undermines democratic participation and constitutional rights in the digital sphere. This paper, therefore, argues for the mitigation of algorithmic bias as a necessary step to curb inequality and discrimination, situating the problem within the Indian legal-constitutional framework while drawing upon comparative global jurisprudence and rights-based theories of justice.

Methodology:

This paper adopts a doctrinal legal research method, supported by a comparative and theoretical framework. It examines primary legal sources, including the Indian Constitution (Articles 14, 19, and 21), the Information Technology Act, 2000, particularly Section 79 on intermediary liability, the IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021[1], and the Digital Personal Data Protection Act, 2023. Judicial precedents such as Shreya Singhal v. Union of India[2], Justice K.S. Puttaswamy v. Union of India[3], and Anuradha Bhasin v. Union of India[4] are critically analysed alongside persuasive foreign cases like Gonzalez v. Google (2023)[5] and R (Bridges) v. Chief Constable of South Wales Police (UK)[6]. The study is further enriched by theoretical perspectives, including Social Contract theory, Foucault's surveillance framework, rights-based theories of justice, and utilitarian ethics, enabling a critical assessment of algorithmic bias through both constitutional principles and comparative global approaches.

Background:

The roots of algorithmic bias trace back to the early history of statistical modelling. Algorithms were created as a neutral set of instructions to solve mathematical or logical problems. The central idea of algorithmics is based on the statement - quality of output depends entirely on the quality of input data. Several factors lead to algorithmic bias on social media platforms:

Algorithmic bias arises from multiple sources, such as algorithms being trained on large datasets, and if those datasets are incomplete, unrepresentative, or already reflect societal prejudices, the system inevitably reproduces those biases. For example, an algorithm trained largely on media content from one region will privilege that region's perspective while sidelining others. Secondly, the architecture of the algorithm itself can embed bias. For instance, systems that rank content based on engagement may give preference to sensational or polarizing posts over accurate information, simply because they generate stronger user reactions. Lastly, since social media algorithms learn from the way individuals consume content, they can evolve in ways that reinforce user preferences. This often results in echo

---

[1] The Information Technology Act (intermediary Guidelines and Digital Media Ethics Code) Rules, 2021
[2] 5 SCC 1
[3] 10 SCC 1
[4] 8 SCC 1
[5] LLC, 598 U.S. 617
[6] 2020 EWCA Civ 1058

chambers, where people are repeatedly exposed to material that confirms their existing viewpoints, thereby deepening bias rather than challenging it.[7]

Algorithmic Bias as a violation of equality (Article 14):

Article 14 of the Indian Constitution guarantees "equality before the law" and[8] "equal protection of the laws", it stands as a fundamental bulwark against discrimination. Its connection to algorithmic bias is profound and multifaceted. At its core, algorithmic bias often results in disparate treatment or impact based on protected characteristics such as religion, caste, gender, socio-economic status, or even political affiliation.[9] When a social media algorithm, for instance, disproportionately amplifies hate speech against a minority community, suppresses content from certain political viewpoints, or unfairly targets individuals for scams based on their online behaviour, it inherently violates the spirit of equality.

Algorithmic biases on social media platforms often result in discriminatory outcomes based on gender, caste, religion, or socio-economic markers embedded; for instances, recommendation systems can make hate speech targeting minorities and silence counter-speech. In the same way, the automated moderation systems can also unfairly filter out content in vernacular languages over the English language, thus disadvantaging the speakers of the regions. This constitutes indirect discrimination, which is against the constitutional ethos of Article 14.

The case of E.P. Royappa v. the Supreme Court. State of Tamil Nadu (1974)[10] was of the view that arbitrariness is the opposite of equality. In cases where the algorithms are treated as black boxes and opaque in decision-making, their results cannot be explained by the use of reasonable classification or intelligible differentia. When a user is shadow-banned without being warned that an opaque algorithm marked his/her content, it is equivalent to arbitrary exclusion. Similarly, in Maneka Gandhi v. Union of India (1978)[11] 1 SCC 248, the Court stated that the laws or actions should not be arbitrary but should be just, fair, and reasonable. To this approach

---

[7] Understanding Algorithmic Bias in Social Networks, Dave barlett and matt DeSalle, April 17th, 2024 https://www.linkedin.com/pulse/understanding-algorithmic-bias-social-networks-simplyput-consulting-hk6xe/
[8] Answers to: Name and explain three laws from the constitution that protect citizens against the human rights violations of safe and healthy environments. https://www.classace.io/answers/name-and-explain-three-laws-from-the-constitution-that-protect-citizens-against-the-human-rights-violations-of-safe-and-healthy-environments
[9] The Constitution of India, 1950, Article14, 19(1)(a) and 21.
[10] 4 SCC 3
[11] Right to Work in India: An Analysis – Kalinga Plus. http://kalingaplus.kalingauniversity.ac.in/right-to-work-in-india-an-analysis/

in algorithmic governance, random censorship of the speech of users without justification or due process of law goes against the principle of equality. Comparatively, the UK Court of Appeal in R (Bridges) v. In the case of Chief Constable of South Wales Police [2020] EWCA 1058, it was found that the application of live facial recognition was not guaranteed any protection and was susceptible to discriminatory results, thus violating the equality and data protection principles. Even in India, despite not having an effect, this persuasive precedent highlights that algorithmic systems are to be subjected to rigorous scrutiny on equality.

Therefore, algorithmic bias has a direct negative impact on ensuring egalitarianism in digital arenas. Mitigation would entail platforms to perform algorithmic impact tests as well as introduce bias audits, and justify that the algorithmic processes comply with the command of non-arbitrariness and fairness in Article 14.

Article 19(1)(a): Free Speech vs Algorithmic Censorship:

Article 19(1)(a)[12] which is the right to freedom of speech and expression is at the centre of democratic discourse. The social media sites have turned into the marketplaces of ideas; however, the algorithms are the gatekeepers of visibility. Content is not open but is rather picked, promoted, or hidden according to metrics of engagement that promote sensationalism over discursive equilibrium.

The Shreya Singhal[13] case. Section 66A of the IT Act[14] was declared invalid by the Union of India (2015) as vague and overbroad, and that any restrictions on speech should be specific and precise. Although Shreya Singhal was discussing direct state censorship, its principles also apply to algorithmic censorship: suppression of speech, even if it is vague and arbitrary, by private actors in a public role, is constitutionally suspect.

In the Anuradha Bhasin case, the Court said that the internet is an essential means of practicing the freedom of speech in the Union of India (2020).[15] Indirectly, when platforms are limiting visibility by utilizing biased algorithms, they are effectively limiting the speech rights of a user, enabled by the internet. The U.S. Supreme Court took into account Gonzalez v. in the whole world. Algorithms' amplification was put into question by Google LLC (2023), although its

---

[12] Id 2
[13] 5 SCC 1
[14] The information Technology Act, 2000
[15] 3 SCC 637

study is inconclusive. Bringing up the case, petitioners claimed that YouTube has actively promoted extremist content through its algorithm, thus assisting in terrorism. The Court did not decide on the question of liability, but it was a good indication of the fact that algorithms are not always able to escape the banner of neutrality.

Therefore, Article 19(1)(a) is in danger of algorithmic suppression and amplification. Meanwhile, Article 19(2) allows adequate limitations to prohibit hate speech or incitement. It is difficult to tell when moderation is reasonable and when it is biased. Mitigation means transparency requirements - users need to be informed as to why their content was suppressed - and there should be appeal mechanisms to challenge algorithmic suppression.

Article 21 on Privacy, Identity & Dignity:

The right to privacy is captured by the wide-ranging provisions ensuring life and personal liberty in Article 21, Justice K.S. Puttaswamy v. Dignity, (2017)[16], and reputation. The violations of these in algorithmic bias are various:

- Profiling and Surveillance Capitalism – Algorithms deduce sensitive information (religion, sexuality, health) based on unproblematic behaviour, without their permission. This profiling might result in discriminatory targeting, which is against privacy and autonomy.

- Reputational Damage- Prejudiced algorithms can enhance defamatory messages, hurting the dignity and reputation of a person, both of which are aspects of Article 21.

- Psychological Impact- The experience of being either algorithmically suppressed or harassed creates a chilling effect, as it takes away the right to live with dignity.

The Court established a proportionality test in Justice K.S. Puttaswamy whereby any limitation of privacy needs to meet the test of legality, necessity, and proportionality. Social media that follow opaque algorithms to take advantage of the information in the hands of the user would not pass this test, particularly when the outcome is discrimination. Rights to access, correction, and purpose limitation are now established in the Digital Personal Data Protection Act, 2023 (DPDP Act). Even though it does not explicitly refer to bias, its clauses provide mechanisms

---

[16] 10 SCC 1

to require privacy-by-design algorithms that do not discriminate profiling. The same case as the Bridges (UK, 2020) ruled that facial recognition was poorly safeguarded, which infringed privacy and equality. This reveals a trend of judicial interpretation of algorithmic harms as privacy protection worldwide.

Therefore, mitigation is needed by ensuring that algorithms are transparent, users consent, and impacted to hold dignity, autonomy, and privacy.

Can Intermediary Liability be seen through Section 79 Information Technology Act, 2000?

The Information Technology Act 2000, in section 79, gives a safe harbour to the intermediaries and protects them against liability for the content posted by third parties, provided the intermediaries apply due diligence. Platforms used to be traditionally considered as passive conduits. Algorithms, however, transform them into active arbiters, and their liability posture is essentially altered. The IT (Intermediary Guidelines and Digital Media Ethics Code)[17] Rules, 2021, added due diligence, in that it should respect the constitutional rights of users. The new government jobs indicate that platforms might not have a safe harbour in case their algorithms are discriminatory or prejudiced. This provides a legal force on platforms to reduce the algorithmic bias.

Facebook Civil Rights Audit (2020) has a persuasive power proving that self-regulation is not effective because the internal corporate audits have revealed the discrimination of minorities by the algorithm. This argument may support the transfer of Section 79 requirements in India in the removal of reactive content to proactive algorithmic fairness.

According to Section 79, 'exemption from liability of intermediary in certain cases', that is, it provides a safe harbour. Therefore, mitigation of the same will require:

- Platforms need to assess the risk of discrimination in Algorithmic Impact Assessments (AIAs)

- Independent Audits- Algorithms controlled by third parties.

---

[17] The Information Technology Act (intermediary Guidelines and Digital Media Ethics Code) Rules, 2021

- Grievance Redressal - User contestation mechanisms.

Misuse of Freedom of Speech by the Media.

Article 19(1)(a) is the freedom of speech that forms the foundation of the democratic dialogue. However, on social media, not only do individuals exercise this right, but platforms do it through the algorithmic decisions they make. Misuse of the role as a digital intermediary by platforms occurs in two ways to an extent that is interconnected:

By amplifying harmful content through their efforts, the defendants impaired the credibility of the information and data they shared on the Twitter page. Over-amplification of Harmful Content. Algorithms are programmed in order to maximise engagement. By so doing, they commonly propagate hate speech, disinformation, and divisive speech. And this is where the problem of a paradox arises because platforms use the freedom of expression as an excuse to take no action, yet in reality, it is not neutral; it is rather biased and enables the promotion of harmful material to enrich themselves.

Facebook Civil Rights Audit (2020)[18] Auditors have determined that the algorithms at Facebook pushed out misinformation and methodical hate speech against minorities, whereas the executives of the company opposed any change in the name of Free speech and interaction. Similarly, in Gonzalez v. The petitioners (Google LLC, 2023) had claimed that the recommendation algorithm of YouTube was encouraging the recruitment of ISIS members to join it by suggesting their videos. The case did not establish an algorithmic liability ruling by the U.S. Supreme Court, although it highlighted how the largely so-called neutral position on free speech is fronted by commercial motivation to promote extremist content.

In the Indian case, when the platforms give communal hate an upward trend because of the algorithmic placement, it will be a direct infringement of Article 14 (equality) and Article 21 (dignity, reputation) of the impacted groups. The campaign orchestrated by the President aided in deterring peaceful demonstrators who might have opposed the curfew ordinances, yet infringed upon the rights of the politically dissenting people. The blacklisting of the President helped to dishearten peaceful protesters who could have gone against the curfew ordinances,

---

[18] Facebook's Civil Rights Audit – Final Report, Laura w. Murphy and Megan Cacace, Facebook civil rights Audit, Relman Colfax PLLC, Issue 1, July 8th, 2020, chapters-six and seven.

but violated the right to political opposition among the people.

At the opposite end, platforms tend to demote or shadow-ban critical opinions using vague community rules. Although not a direct state censorship, this type of suppression is practically the same as a suppression of dissent in the online community. Shreya Singhal v. Union of India (2020) under Section 66A of the IT Act (2015)[19] was declared unconstitutional by the Union of India (2015) on the ground that an imprecise restriction was a form of chilling free speech. It is the same issue today with algorithms, whereby individuals are punished without clear justification or even legal requirements. Anuradha Bhasin v. It was confirmed in that speech is central to internet access. Analogically, this content, such as biased suppression of visibility, can be contended to be the denial of speech rights in the online space.

Balancing Articles 19(1)(a) and 19(2):

Article 19(1)(a) is not absolute; it can be limited by reasonable limitations in competence by Article 19(2) (public order, decency, morality, security, etc.). Nevertheless, algorithmic moderation usually does not pass these constitutional tests:

The hate speech is being encouraged with an algorithm under the name of freedom of speech, even though it can pose a threat to the order in society.

The voices of dissent are tamed without being subordinated to any category of 19(2) ground, and as such, the restrictions are unconstitutional.

Echo Chambers and Filter Bubbles - Inequality of Information Access:

The news feeds and recommendations are personalized to maximize user engagement. This weakens pluralism and introduces a disparity in access to information. Anuradha Bhasin v. Supreme Court. Union of India (2020)[20] confirmed that free speech involves access to the internet. However, biased algorithms are successful in censoring the number of different points of view since the thoughts of a minority or opponents can rarely be seen. The freedom of expression and equality in online space is endangered when selected groups of people are disproportionately silenced or subjected to polarizing messages.

---

[19] Supra no 12
[20] Supra no. 13

Invisible Discrimination and Shadow Banning:

No better embodiment of arbitrary algorithmic discrimination is shadow banning, the act of making a user less visible without letting them know. The caste, gender, or political dissent posts of users of disadvantaged groups are likely to be blocked from other users. This goes against the broad meaning of Article 14 in E.P. Royappa v. State of Tamil Nadu and Maneka Gandhi (1974).[21] Union of India (1978), that do not allow arbitrariness. Unintelligible suppressed opaque is not intelligible differentia or rationally classified. In the rights view point, shadow banning undermines digital dignity by invisibilities of voices. Analogically, in the Bridges case (UK, 2020), the Court quashed the application of facial recognition based on the absence of safeguards, and the Court was aware of the discrimination risks. On the same note, algorithmic invisibility in India should be challenged on Articles 14 and 19(1)(a).

Algorithmic Policing and Predictive Policing:

Algorithms are prone to reproducing severe social biases, and algorithmic profiling tends to be based on predictions derived from data. These predictive algorithms can reproduce societal biases in India, where caste, religion, and gender are deeply ingrained in the social order. To illustrate, when algorithms label accounts as potential sources of misinformation, they will be biased in favour of vernacular content.

Justice K.S. Puttaswamy v. Union of India (2017)[22] has defined privacy as a primary right, which prevents intrusive profiling. However, social media profile users without their consent and breach Article 21 in terms of dignity and autonomy. Globally, Gonzalez v. The case of extremist recruitment with the help of algorithmic recommendation systems revealed by Google (2023) demonstrates the threatening consequences of uncontrolled profiling. This is the reason why India needs to come up with proactive protection against discriminatory profiling, in the form of algorithmic audits and compulsory data protection principles by the DPDP Act, 2023.

Comparative and Theoretical Approaches:

Normative theories and comparative jurisprudence additionally help understand how deep

---

[21] Supra no 9
[22] Supra no 14

algorithmic bias is and how urgent the mitigation should be. Cases that happen internationally include R (Bridges) v. Chief Constable of South Wales Police (2020) and Gonzalez v. Google (2023). The usage of automated systems without safeguards against bias is becoming more questionable by courts in the US, as demonstrated by the GDPR and Anti-Discrimination Directives of the European Union, which are not perfect but incorporate procedural rights, including the right to be able to challenge automated decisions. These events underscore the fact that India runs the risk of being left behind unless the IT Act and constitutional jurisprudence are changed to explicitly consider algorithmic harms.

In theory, platforms can be taken as digital sovereigns within the context of the Social Contract Theory: they are presented as users giving up their personal information in the hope of equal and just treatment. The realisation of discrimination through a biased algorithm is a betrayal of this contract as it fails to support neutrality. Applying Foucault's theory of surveillance[23]Algorithms should not be perceived as a passive tool, but as an active tool of discipline and control, which are used to push users to conformity and enhance state or corporate interests. Under a utilitarian viewpoint, platforms justify algorithmic curation, since this maximises total happiness by engaging. However, the actual outcomes are the propagation of false information, damage to reputation, and marginalization of the minority, which demonstrates that the losses are much greater than the alleged gains. Lastly, rights-based theories (Dworkin and Rawls)[24] Underline the fact that dignity, equality, and liberty are natural rights, which cannot be traded for efficiency or profit. The veil of ignorance suggested by Rawls is especially educative: no rational mind would accept a biased algorithm in the case when he or she was not aware of the possibility of being a part of a marginalised group in cyberspace.

Combined, these comparisons and theoretical prisms reinforce the argument that algorithmic bias supports inequality and discrimination. They also demonstrate why mitigation can no longer be a corporate goodwill issue or an issue to do with technical tweaking, but has to be legally established within the Indian constitutional provisions of Articles 14, 19, and 21 with clear responsibilities under the IT Act and any subsequent digital governance legislation.

---

[23] Harsh Mahaseth, Discipline and Punish: Understanding Foucault's theory of Panopticism, August 18th, 2024
[24] Rights based Approach, Funded by the European Union, 1st March, 2024

Conclusion and Recommendations:

The discussion in this paper confirms that the concern of algorithmic bias in social media platforms is not a technological failure but a fundamental constitutional and human rights issue. Algorithms have become the determinants of the reach and visibility of speech as well as credibility, becoming an effective way to influence who is listened to and who is silenced in the digital public sphere. Being biased, such systems strengthen the discriminatory process, enhance hate speech, and make the voices of the marginalised invisible, thus undermining the values of equality, free speech, and dignity.

The reason why mitigation is necessary is that the bias introduced by algorithms goes to the heart of how the Indian constitutional law should be. The violation of Article 14 occurs when the platforms use arbitrary, discriminatory suppression of users without being transparent or justified. The compromise of article 19(1)(a) occurs when the right to free expression is twisted to serve the purpose of hate speech being enhanced in the name of neutrality and dissent being suppressed without due process. Privacy and dignity are threatened by profiling, reputational harm, and psychological exclusion of people, thus interfering with Article 21. The constitutional basis for extending protection against algorithmic harms has a strong constitutional framework due to judicial precedents like E.P. Royappa[25] and Maneka Gandhi and others like Shreya Singhal and Anuradha Bhasin, as well as Puttaswamy. However, without explicit regulation, these rights are not implemented in cyberspace.

Similar trends reinforce this urgency. The case of Bridges that was decided on the constitutional level, courts must expressly accept the fact that algorithmic harms should be considered under the provisions of Articles 14, 19, and 21, subjecting platform practices to a proportionality and reasonableness test. by the UK Court of Appeals declared facial recognition illegal due to the absence of safeguards, EU GDPR provides rights against automated decisions, and a case in the U.S. Gonzalez v. established the same. Google initiated algorithmic liability debates. These illustrations prove that the world's democracies are approaching algorithmic accountability. India cannot afford to fall behind, especially with its rich and susceptible social structure, where biased amplification can easily breed communal disharmony or even inequality.

In this regard, mitigation needs a multi-pronged reform approach. On the constitutional level,

---

[25] Supra no 9

it is crucial that courts should expressly accept the fact that algorithmic harms should be considered under the provisions of Articles 14, 19, and 21, subjecting platform practices to a proportionality and reasonableness test. Legally, the Digital India Act of the future should have a mechanism of obligatory algorithmic impact reviews, audit services, disclosure of transparency, and the ability to challenge automated decisions, based on international standards. The provision in section 79 of the IT Act needs to be reformulated in such a way to ensure that the protection of safe harbour is limited to the provision of an algorithmic fairness in addition to reactive content takedown. Additionally, the Digital Personal Data Protection Act, 2023, must be made operational so that it incorporates fairness-by-design into privacy-by-design, which does not make data-driven profiling entrench discrimination. On the user side, there should be avenues of grievance redressal to enable people to sue algorithmic harms with group remedies to be taken if systemic bias applies to the whole community. In addition to this, digital literacy programs are needed to enable users to counter algorithmic discrimination.

Finally, algorithmic bias mitigation does not occur at the expense of limiting innovation but by integrating constitutional morality into digital governance. The platforms should not be allowed to prioritize engagement metrics to the detriment of equality, free speech, and dignity. The Indian constitutional vow of justice, liberty, equality, and fraternity should carry over into the algorithmic era. With legal reformation, judicial activism, and user empowerment, the social media sites could be made to serve the democratic purpose they were originally meant to serve, as opposed to being used as tools of inequality and discrimination.