

CODED HIERARCHIES: RACIAL AND CASTE DISCRIMINATION IN SOCIAL MEDIA

Dr. Prashant Dhondiba Kasabe¹, Mr. Martin Subhash Khairnar²

¹Associate Professor, JET's Zulal Bhilajirao Patil College, Dhule (MS).

Email: kasabeprashant44@gmail.com

²Research Scholar, JET's Zulal Bhilajirao Patil College, Dhule (MS).

Abstract

Earlier, the means of communication with the people were yagna, home-havan, bhajan, kirtan, bhagavatsaptah, sermons, newspapers, radio, etc. After World War II, television was included in it. All these are traditional media for disseminating information. Now, internet-based infotech and biotech technology-based devices have come, which are called social media. In reality, they are not social media but global corporate media. They are big corporate companies and their operations are mainly conducted worldwide through servers in a city in America or China. Let us call them global media. These global media have created a monopoly in delivering information to the people. In India, Facebook, WhatsApp, YouTube, Google Plus are the only internet platforms each. In India, 'Koo' has not been able to replace Twitter and 'Telegram' has not been able to replace WhatsApp. Like TV, global media is also a mine for spreading 'fake news'. BBC had held a seminar on 'Beyond Fake News' at IIT Delhi on 13th November 2018. In it, Facebook's Manish Khanduri said about propaganda, 'If it violates our Community Standards, we remove it'. For example, we have a very strict policy on content related to porn, violence, rape or murder, and such content is removed immediately; but when someone calls some content as propaganda for a political party, we enter the realm of thought and speculation. In the same programme, Google's Erin J. Lew said that they are trying to stop fake news. On the other hand, regarding action against fake news, Twitter's Vijaya Gade said, 'We don't have any policy that can stop you from lying on Twitter. Even if we give people the option to lie, what can we do about it? Because it is not against our policy'. This means that the global media is not in the mood to stop fake news and propaganda. It was because of fake news and propaganda that the case of Kanhaiya Kumar, a PhD student at Delhi's Jawaharlal Nehru University, was made. Due to the propaganda of love jihad, on December 8, 2017, Shambhulal Regar alias Shambhu Bhavani killed Afrazul, who had gone to work as a labourer from Malda in West Bengal, by hitting him on the head with a hammer. Shambhulal Regar alias Shambhu Bhavani was a victim of the fake agenda created with the help of biased algorithm set in our global media. Thousands of ordinary people have been transformed into murderers and rioters in this way. From the perspective of the impact of propaganda on Shambhulal Regar, the attack on Afrazul's head with a hammer was revenge for the 'atrocities committed against Hindus', this was an attack against the conspiracy to 'trap Hindu girls in the net of love and convert them to Muslims'. This was revenge for the 'Islamic conspiracy' to separate Kashmir from India. This was revenge for the conspiracy of Muslims to 'increase the population by having four marriages and having many children'. This Shambhu, while not being an activist of any Hindutva organization, brutally murdered a poor labourer. Millions of human robots have been created in this way. These human robots do not have the ability to think for themselves and have the power of reason. These human robots act by believing the lies they hear in propaganda. The rapid evolution of social media

and Artificial Intelligence (AI) was initially viewed as a democratizing force that would dismantle traditional social barriers. However, this paper argues that these technologies have instead created "Coded Hierarchies" digital environments where historical racial and caste-based discrimination are not only preserved but also implemented and automated as a natural tendency. By referring to Dr. B.R. Ambedkar's 20th-century critique of biased administrative "machinery" to 21st-century algorithmic "machinery," the study illustrates how systems of exclusion are scaled through code. The paper analyzes the Frances Haugen disclosures to highlight the catastrophic neglect of non-Western markets and the "detection gap" in caste-based hate speech. Furthermore, it incorporates the perspectives of Dr. Vijender Chauhan to critique the "Savarna-bias" inherent in Large Language Models (LLMs) and digital hiring tools. The findings suggest that the "neutral" design of Silicon Valley platforms acts as a "Digital Manu," reinforcing dominant-group supremacy. The paper concludes that true digital justice requires a shift from technical colour-blindness to a "caste-conscious" and "race-aware" design philosophy that prioritizes the safety of marginalized communities over platform engagement metrics.

Keywords: Coded Hierarchies, Racial and Caste Discrimination, Social Media, Etc.

► *Corresponding Author: Dr. Prashant Dhondiba Kasabe*

Introduction

Dr. Babasaheb Ambedkar argued in the book, *What Congress and Gandhi have Done for the Untouchables* that "legal reforms and constitutional guarantees (like Article 17) were insufficient if the machinery responsible for enforcing them—the police, judiciary, and administration—remained under the control of upper-caste Hindus who harboured traditional prejudices. He believed that because these officials shared the prejudices of the Hindu community, they often acted in collusion against the interest of the untouchables, leading to under-reporting of crimes, failed investigations, and a denial of justice". It was the prediction of the visionary Sociologist Dr. Babasaheb Ambedkar in the mid of the 20th Century. Since the inception of the social media/corporate global media, its complete authority has been shouldered, grabbed and controlled by the upper-caste corporate power holders. These platforms are solely owned by them, that's why whatever data is created, provided to these corporate sectors, it is used for the caste and racial discrimination at worldwide. The colonial mindset is presiding over all the Asian and African Countries. When we dive deep into the world of social/global media handles, we come across algorithmic bias is working at its best to create caste discrimination and racial discrimination.

In reality Algorithmic bias means systematic, unfair, unjust and repeatable errors in computer systems that produce discriminatory outcomes, often disadvantaging specific groups based on caste (upper-castes), race, gender, colour (e.g. facial recognition trained mostly on white coloured faces) age or socio-economic, religious, and regional factors. These biases arise from flawed training data, poor model design, or human prejudices embedded in AI algorithms. The tech industry is founded on the concept of "Technological Neutrality"—the belief that algorithms are mere mathematical formulas, inherently free from human prejudice. This perspective assumes that if a machine makes a decision, it must be objective. However, this research asserts that technology is a social product. It is shaped by the priorities of its creators, the limitations of its training data, and the cultural contexts of its headquarters.

The Digital Translation of the Varna System

In the Indian context, the internet has become a new site for the exercise of the Four Varna system. While physical untouchability is a punishable offense under the Indian Constitution, "Digital Untouchability" thrives. This is observed when Dalit-Bahujan-Adivasi (DBA) voices are systemically deprioritized by recommendation engines. The "Coded Hierarchy" ensures that the digital public square remains dominated by those who already possess historical social capital. If the surnames such as Jadhav, Patil, Kulkarni, Mochi, Sawant, Salve, Deshpande, etc. are given to the AI and asked to offer the jobs to them, at this data AI gives jobs of IT sector to the surnames like Deshpande, Kulkarni. The job of sweepers is offered to surnames such as Patil, Jadhav, Sawant, etc. and the job of sewage cleaners is given to Mochi, Salve, etc. This kind of biases are found in the AI tools. As far as Indian culture is concerned upper caste Brahmins, Kshatriya, Vaishyas will get good quality jobs and salary also. But Shudras and Ati-shudras/ Untouchables like SC's, ST's, OBC's will be given the jobs which are under-rated in the social context. This is the new Four Varna System is staunchly and rigorously following in the AI modelled global media handles.

The Intersection of Global Hierarchies

Though race and caste are distinct, their digital lives are strikingly similar. Both rely on the act of categorization to maintain power. In the United States, this manifests as the "New Jim Code," where technology is used to manage and marginalize Black populations. Amazon developed an AI recruiting tool that was trained on resumes submitted to the company over a 10-year period, which were predominantly from men. Consequently, the algorithm taught itself that male candidates were preferable, downgrading resumes containing words like "women's".

The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) software, used to predict recidivism (re-offending) in the US, was found to be twice as likely to falsely flag Black defendants as high-risk compared to white defendants. An algorithm used on over 200 million people in US hospitals to predict which patients needed extra medical care was found to heavily favour white patients. It wrongly equated higher healthcare costs (associated with better access) with higher needs, neglecting the needs of Black patients with lower access to care. Facial recognition technologies have demonstrated significantly higher error rates when identifying people of colour, particularly black women, leading to wrongful arrests.

Algorithms can perpetuate historical biases in lending, with studies showing that AI systems for mortgages have routinely charged minority borrowers higher interest rates than white borrowers for similar loans. Social media algorithms have been shown to disproportionately flag content from minority users as offensive, while missing hate speech targeting those same groups. A UK exam grading algorithm during the pandemic lowered scores for students from state schools while boosting scores for those at private schools, reflecting systemic socio-economic biases. Algorithms like the "five-factor model" tests, used to screen applicants for, have been alleged to discriminate against individuals with mental health conditions. In India, it manifests as "Digital Brahminism," where the digital infrastructure is optimized for the values and languages of the upper castes. This paper explores the "machinery" that allows these hierarchies to persist.

Technical Framework of "Biased Machinery"

To understand the failure of modern social/ global media moderation, we must revisit Dr. Babasaheb Ambedkar's 1945 work, *What Congress and Gandhi have done for the Untouchables*. Dr. Babasaheb Ambedkar's central idea was that legal protections, such as Article

17, are ineffective if the people in charge of the system are biased. He argued that a police officer or a judge is not a blank slate; they are members of a specific social community first, "The machinery of Government... is a person in the service of the Hindu community... he shares the sentiments of the Hindu community... He is a Hindu first and a Government servant afterwards." In the digital age, the "Algorithm" has taken over the role of the administrator. Just as Dr. Babasaheb Ambedkar's official "shares the sentiments" of the dominant group, modern AI "shares the sentiments" of the data it is fed. If an AI is trained on a "Savarna-dominated" internet, it will treat Dalit, Bahujan and Adivasi grievances as "outliers" or "spam". The "machinery" of the platform—the code that sorts, hides, and highlights content—acts in collusion with traditional prejudices.

Racial Discrimination in AI: The "New Jim Code"

Ruha Benjamin's "New Jim Code" explains how technology hides racial prejudice behind a veneer of innovation. One of the most dangerous examples of coded hierarchy is in facial recognition. Research by Joy Buolamwini has shown that AI error rates for dark-skinned women are significantly higher than for light-skinned men. When these flawed tools are used by police "machinery," they lead to the wrongful targeting and arrest of Black individuals. This is not a technical glitch; it is the automation of a racialized police state. Black and Grey skinned men and women are rejected on all the grounds of jobs and selection process.

Algorithms used in healthcare, housing, and credit scoring often claim to be "race-blind." However, they use "proxies" like zip codes or purchasing history that are deeply tied to racial history. By targeting these proxies, the AI maintains a racial hierarchy while allowing the companies to claim they are being "objective."

In 2021, whistle-blower Frances Haugen leaked internal Meta (Facebook) documents that provided a "smoking gun" for how platform machinery ignores the Global South. Haugen revealed that Meta spent the overwhelming majority of its safety budget on the U.S. market, leaving India—its largest user base—virtually unpoliced. The "machinery" lacked the ability to understand regional Indian languages. Consequently, casteist slurs and calls for violence against DBA communities were not flagged by the AI. Because the AI couldn't "see" the casteism, reports made by Dalit users were dismissed. This is the modern version of a police officer refusing to file FIR for a marginalized victim. Haugen's disclosures highlighted that "outrage" drives engagement. On Indian social media, caste-based conflict generates high levels of interaction. The algorithm, designed to maximize profit, promotes this content, effectively turning the platform into a tool for Brahminical supremacy.

The critique of digital hierarchies in India has been led by intellectuals like Dr. Vijender Chauhan, who focuses on the "invisible" nature of caste in technology. Dr. Vijender Chauhan has pointed out that Large Language Models (LLMs) like ChatGPT are trained on the "Internet," which is not a neutral archive. In India, the written word has historically been the domain of the upper castes. When AI summarizes "Merit" or "Indian Culture," it reflects a Savarna-centric view. It treats the struggle for reservations (affirmative action) or Dalit history as "subjective" while treating dominant-caste narratives as "objective fact."

Dr. Vijender Chauhan argues that AI hiring tools are the new gatekeepers of the economy. These tools are trained to find "quality candidates" based on metrics like school names, language patterns, and social networks. Since these metrics are proxies for caste privilege, the AI automates the exclusion of DBA candidates, ensuring that the "machinery" of the private sector remains "Savarna-pure."

Findings: The Scaling of Digital Exclusion

The intersection of Ambedkar's philosophy and modern tech reality reveals three critical findings: **Algorithmic Collusion:** Platforms do not need to be "explicitly" racist or casteist to produce inequitable and biased results. The design itself favours those with existing social upper caste and corporate capital.

Linguistic Imperialism: The focus on "standard" English and Hindi obliterates the dialects and expressions of the marginalized, leading to a "Digital Erasure."

The Feedback Loop of Oppression: If a community is historically marginalized, they have less "digital data" to offer training to the AI. The AI then considers them as "less important," which leads to further ostracism. This paper shows that without a social conscience, technology is merely a more efficient and sometimes hidden and sometimes crystal clear way of oppression.

Conclusions

The complex web of caste-based discrimination within the Indian subcontinent, profoundly rooted in historical legacies, continues to infiltrate modern society, making its presence felt across countless facets of life. This malicious societal ill has seamlessly transitioned into the digital landscape, finding a new platform in global corporate media, where hate speech and derogatory references towards marginalized castes and races have become all too common. The alarming prevalence of caste, race, gender, colour, and region-based hate speeches on platforms like Facebook, Instagram, Threads, WhatsApp and Twitter reveal the systemic challenges in moderating and categorizing such content effectively. The incongruences in categorization and identification mechanisms on global media platforms like Facebook and Twitter, despite their public commitments to combat hate speech, remain blatant. These platforms' developing policies, albeit acknowledging caste as a protected category, still lack exactitude and consistency in addressing nuanced forms of class, race, gender and caste-related discrimination, especially within the Indian socio-political context.

The struggle against caste-based discrimination in the digital empire demands determined efforts from social/global media platforms, policymakers, and politicians, government officers who allow and pass the policies, social activists and civil society. Addressing the insufficiencies in content moderation mechanisms, bolstering transparency, and cultivating cultural awareness are crucial steps toward fostering safer and more all-encompassing online spaces, and transcending geographical limitations and boundaries and cultural divides. The authoritative lies not just in distinguishing caste as a protected category but in effectively dismantling the systemic barriers perpetuating caste-based discrimination within the online social/global media world.

In conclusion, the researcher wants to focus on the legal action against the consistent with the terms of service, community guiding principle, and the provisions of the Bhartiya Nyaya Sanhita (BNS), the propagation of hate speech constitutes a criminal and punishable offence. To guarantee compliance and uphold community values and principles, it is imperative that enhanced content moderation practices be implemented, including the deployment of contextual AI technologies to identify and eradicate such content promptly. Additionally, legal measures should be followed against individuals responsible for propagating hate speeches. Given the ongoing election campaigns, it is also advisable for the Election Commission of India to mandate social/global media and video-sharing platforms to meticulously monitor and restrain the spread of harmful content by its creators and distributors, thereby safeguarding the reliability of the electoral process. Despite these official declarations, the actual moderation and administration practices often fall short, allowing caste, creed, race, gender, colour-based hate speeches to proliferate relatively

unchecked across social media networks. This incongruity largely stems from the platforms' limited understanding of the cultural and social nuances specific to caste discrimination, as well as the lack of regionspecific moderation strategies that can effectively recognise and alleviate such content. Furthermore, the dependence on automated systems and algorithmic patterns, which lack the contextual mindfulness necessary to distinguish between harmful content and benevolent communication, exacerbates the issue. Effective implementation of these guidelines necessitates a determined effort from tech companies to invest in culturally informed AI technologies and to expand their teams of local experts and professionals who can provide the necessary comprehension and oversight needed to address this universal issues.

The Ambedkarite Solution

Just as Dr. Babasaheb Ambedkar called for the representation of "Depressed Classes" in every level of government "machinery," we must demand representation in every level of the "algorithmic machinery"—from data labelling to C-suite decision-making. Tech companies must move beyond the "Western" model of moderation and build tools that specifically diagnose the nuances of caste, race, gender, colour, creed and regional discriminations.

Dr. Babasaheb Ambedkar's warning remains our most powerful tool for analyzing the digital age: *"Justice is not a matter of law; it is a matter of the machinery that administers the law."* If the machinery of social media remains under the control of those who harbor traditional prejudices - whether through their upbringing or through the biased data they use—then the digital world will never be free. To achieve digital justice, we must dismantle the "Coded Manu" and replace it with a design philosophy deep-rooted in Constitutional morality and human rights.

References

1. Ambedkar, B. R. What Congress and Gandhi Have Done for the Untouchables (1945) & Annihilation of Caste (1936).
2. Allan, R. (2017, June 27). Hard Questions: Who Should Decide What Hate Speech in an Online Global Community Is? Meta. Retrieved December 12, 2023, from <https://about.fb.com/news/2017/06/hard-questions-hate-speech/>
3. Barrett, P. (n.d.). Who Moderates the Social Media Giants? A Call to End Outsourcing.
4. Caste system in India. (n.d.). Wikipedia. Retrieved December 11, 2023, from https://en.wikipedia.org/wiki/Caste_system_in_India#cite_note-31
5. Countries with most X/Twitter users 2023. (2023, September 13). Statista. Retrieved December 12, 2023, from <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>
6. Dixon, S. J. (2023, August 29). Facebook users by country 2023. Statista. Retrieved December 12, 2023, from <https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/>
7. Equality Labs. (n.d.). Facebook India TOWARDS THE TIPPING POINT OF VIOLENCE: CASTE AND RELIGIOUS HATE SPEECH.
8. Giansiracusa, N. (2021, October 15). How Facebook Hides How Terrible It Is With Hate Speech. WIRED. Retrieved December 12, 2023, from <https://www.wired.com/story/facebooks-deceptive-math-when-it-comes-to-hate-speech/>
9. Giansiracusa, N. (2021, October 15). How Facebook Hides How Terrible It Is With Hate Speech. WIRED. Retrieved December 12, 2023, from <https://www.wired.com/story/facebooks-deceptive-math-when-it-comes-to-hate-speech/>

10. Haugen, F. Facebook Files: India Case Study (Internal Documents, 2021).
11. Kirti, Milind. KrutrimBuddhimattrcheMayajalYug: SahmatichiHukumshahi, Aristocracy to Technocracy, Lokvangmaygruha, Mumbai, March 2025.
12. Kumar, M. (2024). Caste-Based Hate Speech Moderation on Facebook (Meta) & Twitter (X), retrieved on 10th April from [https://escholarship.org/content/qt170845t6/qt170845t6_noSplash_95e6b181f3f92df46ff5dbadc3f238a.pdf? t=s9fn18](https://escholarship.org/content/qt170845t6/qt170845t6_noSplash_95e6b181f3f92df46ff5dbadc3f238a.pdf?t=s9fn18)
13. Parthasarathy, M. (2022), The Supreme court on Hate Speech in Elections, retrieved on 9th April 2024 from <https://www.scobserver.in/journal/the-supreme-court-on-hate-speech-in-elections/>
14. Rajagopal, K. (2022), No specific law against hate speech: Election Commission, The Hindu, retrieved on 9th April 2024 from <https://www.thehindu.com/news/national/hate-speech-took-action-under-provisions-of-ipc-rp-act-due-to-lackof-specific-law-ec-tells-sc/article65889106.ece>