Abstract

Title:

AI and Dalit Social Mobility in India: Towards Asian-Informed Theories and Methods for Equitable Digital Development

Authors: [Pranav Dwivedi, Sagar Vishnoi], [Affiliation: Future Shift Labs], [Contact Pranav@futureshiftlabs.com, Sagar@futureshiftlabs.com) +91-8506050873]

Abstract: This research addresses the urgent need for a critical and context-aware approach to the deployment of Artificial Intelligence (AI) in India, particularly concerning its impact on the socio-economic mobility of the Dalit community. While AI presents opportunities for progress, uncritical adoption risks exacerbating existing inequalities rooted in the deeply entrenched Jati system. This paper contributes to a growing body of scholarship that questions the universal applicability of Global North-centric AI models, arguing for the development and implementation of Asian-informed theories, methods, and ethical frameworks to ensure equitable and culturally relevant digital development. Our central focus is the articulation of a framework that centers the lived experiences and historical realities of Dalits in the design, development, and governance of AI systems.

The Jati system, despite being outlawed in India 75 years ago, continues to permeate various aspects of life, limiting opportunities for Dalits in education, employment, and access to resources[2]. Dalits, who constitute 16.6% of India's population, or over 200 million people, face systemic discrimination[1]. A 2021 UN report indicates that one-third of India's Dalits remain in multidimensional poverty. These numbers highlight the persistent and pervasive nature of caste-based inequality in India, making targeted interventions and equitable digital strategies all the more urgent.

The scaling up of AI in India, without careful consideration of these existing inequalities, carries significant risks. As *Shivangi Narayan* argues, *AI systems risk reproducing casteist bias and discriminating against already marginalized groups*[2]. This is further compounded by the fact that Dalits, Muslims, and Indigenous people lag behind higher-caste Indians in smartphone ownership and social media use[1], leading to their underrepresentation in datasets. This underrepresentation can result in AI systems that perpetuate existing biases, for example, mobile safety apps that flag Dalit and Muslim areas as "dodgy," potentially leading to over-policing and unwarranted mass surveillance[1]. *Urvashi Aneja* of Digital Futures Lab points out that the very people excluded from these datasets are then subjected to data-driven

systems that reproduce bias and discrimination[1]. This perpetuates a "*Digital Caste Panopticon*" [1, 2].The potential for bias extends beyond surveillance.

Generative AI applications, such as chatbots, often reflect dominant-caste Hindu surnames when asked for names of Indian doctors and professors, perpetuating skewed representations[1]. Loan approvals, hiring processes, and background checks, increasingly reliant on AI, can further marginalize Dalit communities if these systems are not carefully designed and monitored[1].

As *Nikita Sonavane* of the Criminal Justice and Police Accountability Project notes, AI-assisted predictive policing risks perpetuating legacies of caste discrimination and unjust criminalization[2].To mitigate these risks and harness AI for positive social change, this research proposes a framework grounded in Asian values and methodologies. This includes:

- **Contextualized Ethics:** Moving beyond Western ethical frameworks to incorporate Asian philosophical and cultural values in AI design and deployment. This involves drawing upon the rich philosophical traditions of India, including the concept of *dharma* (duty) and *karma* (consequences), to guide the development of AI systems that are aligned with the values of justice, fairness, and compassion.

- **Participatory Design:** Employing community-based participatory research (CBPR) methods, common in Asian development studies, to ensure that AI solutions are co-created with and for Dalit communities.

- **Data Sovereignty:** Prioritizing data governance frameworks that respect the rights and agency of Dalit communities over their data.

- **Focus on Systemic Issues:** Analyzing the intersection of AI with historical injustices and existing power structures to address root causes of inequality.

This research builds upon the insights of thinkers like Ambedkar, who believed that artificial intelligence can help end social evils like the caste system and reduce inequalities in India[6].

By promoting Asian-led approaches that prioritize equity, social justice, and cultural relevance, this work contributes to a broader effort to decolonize AI research and practice and create a more just and equitable digital future for all. Scholars use the term "AI Dalits" to describe marginalized groups systematically excluded from the benefits of AI advancements[3]. This research seeks to move beyond merely identifying this exclusion to proposing actionable strategies for inclusion and empowerment.