Chapter 8
AI and Social Justice: Democratizing Technology for Inclusive Growth

Abhishek Benedict Kumar
https://orcid.org/0000-0002-3998-3150
Symbiosis Law School, Nagpur Campus, Symbiosis International University, Pune, India

Karun Sanjaya
https://orcid.org/0009-0005-1647-1549
Symbiosis Law School, Nagpur Campus, Symbiosis International University, Pune, India

ABSTRACT

Artificial Intelligence has the potential to change the world by facilitating progress in health, education, climate action, and economic development. But, without proper and ethical use, AI seeps into the fabric of societal malice and makes glasshouses in paradise even more glistening. AI can serve both as a butter and gun for the public. It tackles issues including algorithmic bias, digital divide, and data poverty; meanwhile it mentions community-driven AI building as a method to guarantee inclusion and cultural appropriateness. Strategic elements of policy frameworks such as the EU's AI Act are considered essential to responsibly govern AI, while bottom-up approaches and grassroots initiatives show how the nature of participatory design can be powerful. It urges a joint enterprise by governments, corporations and civil society to create a future with AI that serves the most marginalized populations and functions as an engine for equity. An inclusive approach is crucial to ensure that the choices we make today drive the impact of AI in a direction that advances society.

DOI: 10.4018/979-8-3693-9904-0.ch008

1. INTRODUCTION TO ETHICAL CONTENT MODERATION

1.1 The Rise of Digital Platforms and Moderation Needs

The digital landscape has transformed significantly with the rapid growth of social media and digital platforms, where vast amounts of user-generated content are created daily. Platforms like Facebook, YouTube, and Twitter have billions of active users who share diverse forms of content, including text, images, and videos, on an unprecedented scale. This growth has intensified the need for efficient and ethical content moderation strategies to maintain a safe and inclusive online environment (Statista, 2023). Traditionally, content moderation was handled manually by human moderators tasked with reviewing flagged content for appropriateness and compliance with community guidelines. However, as platforms expanded, the sheer volume and complexity of content overwhelmed manual processes, making it difficult to maintain efficiency, consistency, and fairness in moderation decisions (Gillespie, 2020).

The challenges of content moderation include not only the high volume of content but also the psychological toll on human moderators, who are frequently exposed to disturbing and distressing material. Moderators encounter content that spans a spectrum of harmful topics, including graphic violence, abuse, misinformation, and hate speech. This environment presents significant mental health risks, leading to a rise in mental health issues such as anxiety, depression, and even post-traumatic stress disorder (PTSD) among moderators (Roberts, 2019). The need for alternative solutions has spurred the integration of artificial intelligence (AI) into content moderation, which can process large quantities of data and handle repetitive tasks, potentially easing some of the burdens on human moderators.

The integration of AI in content moderation introduces a new paradigm, where machine learning algorithms can quickly identify patterns associated with harmful content, providing a scalable solution to content filtering. However, the role of AI is not to replace human moderators entirely; instead, it is to assist them, creating a synergy where machines handle high-volume tasks, and humans focus on more nuanced cases that require contextual judgment (Crawford, 2020). Despite its potential, AI-based content moderation has raised concerns around ethical issues such as algorithmic bias, transparency, and accountability, which must be addressed to create a responsible and fair moderation system.