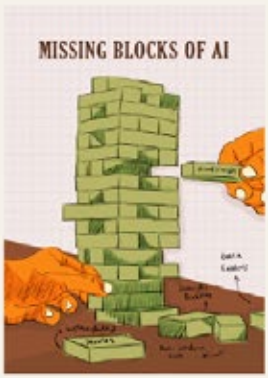


An Anthology

The Platform Question

Power, Accountability and the Global South

Foreword and Introduction by: Amandeep Singh Gill and Osama Manzar






Editors: Dr. Raina Ghosh | Maitri Singh | Dr. Arpita Kanjilal

An Anthology

The Platform Question: Power, Accountability, and the Global South

With Foreword and Introduction by
Amandeep Singh Gill and Osama Manzar

Name of the Publication: The Platform Question:
Power, Accountability, and the Global South

Year of publication: 2025

Edited by: Dr Raina Ghosh, Maitri Singh, Dr Arpita Kanjilal

Design and Illustrated by: Yuvasree Mohan

Published by: Centre for Development Policy and Practice

ISBN: 978-81-993666-4-0-

Supported by: ARISE Community and Digital Empowerment Foundation

Scan QR Code to
Read this Book

About the book

This anthology is the result of a global collaboration among scholars, practitioners, and activists who come together to reflect on the expanding role of platforms in an increasingly digitized world. As platformization reshapes economies, politics, and everyday interactions, it brings with it new forms of regulation, control, and accountability that profoundly affect both marginalized communities and society at large. The collection explores these dynamics through diverse lenses, ranging from debates on platform regulation and Big Tech accountability in content moderation, to feminist critiques of algorithmic bias and safety in online dating spaces. It further examines digital resistance and collective action in contexts such as Latin America and Myanmar, where platform governance intersects with broader struggles for democracy and voice. By engaging with experiences of gig workers, platform-mediated labor, and the emerging ethical questions surrounding artificial intelligence, the anthology situates the global platform economy as a site of tension between freedom and control, innovation and exploitation. Together, these contributions reveal how digital infrastructures are not merely technological systems but deeply political spaces that shape the contours of contemporary life.

Acknowledgements

We would like to express our heartfelt gratitude to everyone who contributed to the creation of this anthology. First and foremost, our sincere thanks to all the writers, illustrators and contributors for bringing this collection to life. We extend our appreciation to ARISE community members, Media Democracy Fund and CREA for supporting research on platform accountability in Global South under Research and Advoacy for Digital Accountability and Rights project. A special thanks to our mentor Mr. Osama Manzar whose guidance helped in shaping this project to its final form. We are also deeply thankful to our readers and supporters, who continue to inspire us to share and celebrate diverse perspectives through literature.

Amandeep Singh Gill

United Nations Under- Secretary-General, Envoy on Technology, Office for Digital and Emerging Technologies

FOREWORD

In today's rapidly evolving digital economy, platforms play a critical role in ensuring a myriad of services in ways that uphold the digital economy, communications and self-expression. I acknowledge their efforts to enhance accountability, and center the citizen perspective in platform governance. A rights-based duty of care approach is essential to ensuring freedom of expression, access to internet, infrastructure and information, while protecting human rights and information integrity as core guiding principles.
In line with the UN's Information Integrity on Digital Platforms policy brief and the UN's Code of Conduct, it is the moral responsibility of platforms to enhance transparency and accountability of their systems through independent audits, content governance, user-centric risk assessments, data protection and data privacy safeguards. With governments playing a pivotal role in setting guardrails consistent with international human rights law, corporations must work simultaneously towards building ethical technologies and platforms, while civil society and academia should continue to play a critical role in designing, testing and monitoring safeguards.

As the UN's Under-Secretary-General and Special Envoy for Digital and Emerging Technologies, it is my honour to support efforts to implement the Global Digital Compact (GDC) and catalyse progress towards cooperative, multistakeholder global digital cooperation mechanisms while anchoring them to local realities. GDC-aligned digital cooperation will help enhance these responsibilities to mitigate harms and co-build a transparent, safe, and trustworthy platform governance, especially for the vulnerable and marginalised populations in the global majority. This multistakeholder dialogue anchored with GDC will strengthen accountability that is open for scrutiny and verifiable, and enable empowerment that is locally informed yet universal.

This anthology is well positioned to reflect on the growing impact of these platforms on citizens and communities in the Global South. I urge platforms, policymakers, regulators, civil society, journalists and researchers to read this book.

PREFACE

This anthology grew out of a vision the ARISE Secretariat had for the ARISE Community to create a space where scholars, practitioners, and activists from across the Global South could think, write, and act together on questions of digital justice. From the outset, our intent was not to produce a conventional publication, but to nurture an experiment a collaborative effort to explore how different vantage points engage with the idea of platform accountability in an unevenly digitised world.

As Big Tech continues to define infrastructures, economies, and everyday life from the Global North, this project seeks to decolonise and decentre those narratives. It brings together contributions that are reflective, critical, and experimental in form -spanning essays, field accounts, and creative reflections.

What makes this anthology special is the collective energy behind it: the voices of ARISE peers and contributors from across regions of the Global South, who have generously shared their experiences, insights, and struggles. Together, they remind us that accountability and justice in the digital age emerge not from a single centre of thought, but through a chorus of perspectives rooted in diverse contexts and lived realities.

– ARISE Secretariat

Osama Manzar

Founder & Director,
Digital Empowerment Foundation

INTRODUCTION

Digital systems and platforms, designed by the current technology regime, are heavily reliant on data, and are increasingly becoming unaccountable, extractive, and exploitative in nature. While the personal, social, cultural, behaviorial, and community data are being collected without informed consent, incentive or guarantee of ownership rights, the platforms and corporations that are conducting drives for data mining, are growing through the concentration of capital and power. This model of 'digital development' has normalised labour commodification through digital-labour based platforms, especially in the Global South where the works of data labeling and data categorisation are advertised as 'creation of jobs'. Moreover, when the criteria or framework for data annotation is not rooted in one's socio-cultural and demographic realities, it resonates with a factory-based model of production run by corporation-led platforms, which entails complete alienation of labour for citizens and communities. Platformisation mandates a 'digital-only' system, which further hinders the communities affected by the digital divide, from accessing and availing their due entitlements and services.

When we address the platform question, it is also crucial to account for the adverse socio-environmental implications of the rising digital infrastructures and data centers that are 'essential' for platforms to thrive - on land, water, power consumption, environment and communities, especially in the global majority countries. For instance, a single Large Language Model (LLM) guzzles as much energy and carbon footprint as one flight, taking 365 rounds around the earth. As we drift towards living in an increasingly platform-driven world, it is our collective responsibility to center human and social intelligence at the heart of these digital systems. In this global digital ecosystem driven largely by the governments and the Big Tech, the Civil Society Organisations (CSOs) must be treated as equal, rightful stakeholders in policymaking as we are best positioned to bring community-informed ground-up perspectives to the table.

However, we see it as a growing trend that CSOs are increasingly been seen as mere means to extract community data at a large scale, and are being approached mostly for seeking endorsement of platforms, models and frameworks that are outsourced, not ethical and community-centered. Aligned with the Global Digital Compact (GDC), this book is a testament to this glocal (global informed by local) dialogue,

"When we address the platform question, it is also crucial to account for the adverse socio-environmental implications of the rising digital infrastructures and data centers that are 'essential' for platforms to thrive - on land, water, power consumption, environment and communities, especially in the global majority countries. For instance, a single Large Language Model (LLM) guzzles as much energy and carbon footprint as one flight, taking 365 rounds around the earth.

About ARISE

A Transregional Community

ARISE (Accountability and Responsibility in South's Ecosystems) is a community space for sharing knowledge, building bonds of solidarity, and a laboratory of ideas from the Global South to counter strategies of global and dominant technology actors that seek to evade platform responsibility in our countries. This is essential in the moment when many technologies exacerbate planetary crises, urgently demanding trans-regional, cooperative action. Comprising 50 member organizations, ARISE represents a diverse range of countries and communities across Latin America, Sub-Saharan Africa, West Asia, North Africa, and Asia. Its membership encompasses expertise across multiple domains, including digital rights, women's and LGBTQI rights, consumer protection, labor and gig economy issues, human rights defense, disinformation, and artificial intelligence governance. ARISE believes in the power of collective reflection and action to create digital environments and standards that meet the human rights and social justice needs of diverse communities in the Global South.

Know the Authors

Syed Mohammad Haroon: Legal researcher and public interest lawyer working at the intersection of technology, constitutional rights, and public policy. He currently serves as Volunteer Legal Counsel at the Software Freedom Law Center, India (SFLC.in), where he is involved in strategic litigation, legal research, and advocacy on digital rights.

Karen Vergara: A journalist with a master's degree in gender studies and Latin American culture. Her work explores the intersections between technology, politics, and sexuality. She is the advocacy director of NGO Amaranta, a decentralized feminist space based in southern Chile that conducts research and provides education to prevent violence.

Tavishi: Tavishi is a Programme Officer at the Centre for Communication Governance at National Law University Delhi. Her research interests include platform regulation, online speech & democracy and global digital cultures. She's interested in investigating the socio-political implications of emerging technologies.

Angelina Dash: Angelina Dash is a Project Officer at the Centre for Communication Governance at the National Law University Delhi, India. Her research focuses on data protection, privacy, and artificial intelligence. She is dedicated to fostering an inclusive and accessible internet, with a strong emphasis on upholding the rights and data autonomy of all, including marginalised and underserved communities.

Nicole Solano Chavarría: Communications professional and audiovisual producer with 9 years of experience transforming ideas into digital content for projects with social, cultural, and environmental impact.

Jamila Venturini: Co-Executive Director at Derechos Digitales. Activist and researcher with over 15 years of experience in civil society organizations. She is a journalist and holds a Master's degree in Social Sciences with a focus on Education from FLACSO Argentina.

Catalina Balla: Journalist with more than 10 years of experience creating strategies and narratives for social

transformation. She holds a Master's degree in Cultural Journalism from Universitat Pompeu Fabra in Barcelona.

Carina Singh, Khush Vachharajani and Rakshita Swamy: The authors are members of the Social Accountability Forum for Action and Research (SAFAR) and are associated with the right to information and right to work campaigns.

Htaike Aung: Htaike Htaike Aung is the Executive Director of Myanmar ICT for Development Organization (MIDO), one of Myanmar's leading ICT-focused NGOs. She has been active in promoting Internet access and digital rights since its public introduction in Myanmar. A co-founder of the Myanmar Blogger Society, she also trains human rights defenders in digital security. Currently, she leads the "Towards an Inclusive Information Society in Myanmar" project to expand ICT adoption nationwide.

Shaik Salauddin: Founder President, Telangana Gig and Platform Workers Union (TGPWU) and Co-founder and National General Secretary, Indian Federation of App-Based Transport Workers (IFAT).

Arpita Kanjilal: Arpita Kanjilal leads the Research & Communications Division at the Digital Empowerment Foundation (DEF), where her work concerns digital rights, equity and sustainable development. With a Ph.D. in Applied Linguistics, she leads several initiatives such as the Just AI Initiative and Digital Swaraj Fellowship, while contributing to platforms like the Digital Just Transition Taskforce and Women for Ethical AI.

Shohini Banerjee: Shohini has spent the last decade working on addressing online and offline gender-based violence through prevention and response programming. She is currently a Knowledge Specialist at Point of View, working to build digital gender justice.

Vaishali Soni: Vaishali Soni is a visual storyteller working at the intersections of gender, sexuality, and technology. She uses art and mixed media as a tool for inquiry, learning, and change across campaigns and creative projects. At Point of View, she works as a Design Specialist and manages the visual universe of the organisation.

The Platform Question: Power, Accountability and Global South

Juliet Nanfuka: A digital rights advocate and researcher with a background in journalism and communication. She has over a decade of experience in technology policy advocacy, and has designed and delivered an assortment of campaigns, research outputs, and multi-stakeholder dialogues that influence policy and public discourse. Her work has been published and cited in regional and international media, policy papers, and academic research.

Akanksha Ahluwalia: Akanksha Ahluwalia leads the Social Inclusion, Media & Information Literacy (MIL), and communication-driven programmes at the Digital Empowerment Foundation (DEF), India. With a background in English Literature and a focus on gender, misinformation, and new media, she brings a nuanced lens to the evolving challenges of digital communication in underserved regions. Her core work revolves around the design and implementation of strategic communication interventions that not only promote digital literacy but also combat the growing threats of misinformation, disinformation, and algorithmic bias, especially in rural and semi-urban India.

Raina Ghosh: Raina Ghosh is a human geographer by training and holds a PhD in Geography from Jawaharlal Nehru University, New Delhi. She currently works with the Digital Empowerment Foundation's Research and Communications Division in New Delhi on digital exclusion, platform accountability, and gendered access to technology.

Osama Manzar: Osama Manzar works at the intersection of Access to Rights and Rights to Access. A Senior Ashoka Fellow and British Chevening Scholar, he is the founder of the Digital Empowerment Foundation (DEF), established in 2002. Under his leadership, DEF has digitally empowered over 35 million people through a network of 2000+ Communication Information Resource Centres across India.
Osama has been a key architect of India's inclusive digital ecosystem, influencing national initiatives such as the Digital Literacy Mission, Common Service Centres, the ban on Free Basics, and the liberalization of ISP licensing through PM-WANI. He has also led pioneering grassroots efforts to counter misinformation, including the creation of a cadre of rural women fact-checkers, frontline digital defenders combating misinformation in underserved communities. A regular columnist for Mint, he has co-authored over 20 publications, including Internet Economy of India and NetChakra.

TABLE OF CONTENTS

The Platform Question: Power, Accountability and Global South

CHAPTER 1

Platform Regulations Across the Globe: Navigating Tensions Between Freedom and Control

By Syed Mohammad Haroon

Introduction

Every day, billions of people interact with digital platforms in different ways, from sharing moments of their personal lives, to expressing opinions, to consuming news, to building businesses, and even to organizing social movements. These interactions have made platforms not just tools of communication, but integral spaces where personal expression, public discourse, and economic activity intersect. Platforms like Meta, X, YouTube, and TikTok now form the invisible infrastructure of modern public life.

Yet, these same systems that empower speech and connection also amplify misinformation, hate speech, sexually explicit content, violence, illegal activities, child sexual abuse material, spam, doxxing, inappropriate imagery, offensive behavior, pervasive surveillance and more. The moderation choices of platforms decide what stories gain attention, which voices are heard, and influence community responses. Platforms over the years have also influenced elections decisions, steered social movements and amplified conflicts.[1] These platforms operate behind closed doors with opaque algorithms, private policies, and commercial priorities and remain unaccountable to the wide-reaching effects they have on society.[2]

In response, governments worldwide are confronting a difficult question: How can platforms be held accountable without undermining free speech? What once thrived as an open and unregulated space is now transforming into a heavily regulated area with constant changes. Legislators are drafting laws that define what platforms can host, how they handle user data,

> “The moderation choices of platforms decide what stories gain attention, which voices are heard, and influence community responses.

and how much power they can exercise over users' online experiences.

These approaches vary significantly across jurisdictions, shaped by each country's constitutional values, political priorities, and interpretation of free expression and state power. Together, they reflect a global effort to find equilibrium between freedom, safety, and accountability in an increasingly digital world.

Evolution of Platform Governance Laws

The early days of the internet in the 1990s and early 2000s were marked by optimism and experimentation. The internet's strength lay in its openness. Anyone could publish, share, or build without needing permission. Regulators were therefore cautious not to impose rules that might slow innovation or discourage investment. The guiding principle of the time was simple i.e. let the internet grow first, regulate later.

This approach led to what came to be known as "light-touch regulation." The idea was to give online intermediaries i.e. internet service providers, web hosts, and later, social media companies, freedom from liability for the content their users created or shared. Governments recognized that holding platforms legally responsible for every user post would make the internet unmanageable and risk stifling free expression. Instead, the focus was on protecting these intermediaries so that speech could flow freely.

> While these platforms helped democratize access to information, they also exposed the risks of mass manipulation, surveillance, and disinformation.

The consequences of platform power soon became evident, particularly during the Arab Spring in 2011, when social media platforms became central to civic mobilization and political change across the Middle East and North Africa.[3] Platforms like Facebook and Twitter were hailed as instruments of empowerment for enabling citizens to organize protests, document state violence, and challenge authoritarian regimes. Yet, while these platforms helped democratize access to information, they also exposed the risks of mass manipulation, surveillance, and disinformation. Governments also felt the need to control platforms from amplifying content critical of the government or state. In Myanmar, Facebook was accused

of facilitating hate speech and incitement that contributed to violence against the Rohingya community.[4] Similarly, in 2019 during the Christchurch mosque attack in New Zealand, the perpetrator broadcasted the attack on Facebook Live, and the video spread rapidly across major platforms before it could be removed, highlighting the failure of content moderation systems and the viral nature of harmful content.[5]

Across these incidents, a common pattern emerged: digital platforms had evolved from being neutral spaces of expression to intermediaries capable of shaping political realities, amplifying extremism, and tested limits of regulatory inaction. What began as a tool of empowerment had also become a medium of manipulation and harm and often amplified existing inequalities. Self regulation approaches by platforms were clearly failing. Governments and civil society began questioning whether the "freedom to innovate" had come at too high a cost. Debates around data privacy, content moderation, algorithmic bias, and corporate accountability grew louder, pushing policymakers to reconsider the adequacy of the early regulatory model. This shift marked the beginning of a new era of platform governance, where platforms were pressured to take responsibility for harmful and illegal content circulating online.

Global Comparison of Governance Laws

1.United States

The United States' framework for regulating digital platforms is deeply rooted in its constitutional commitment to free speech and its long-standing philosophy of limited government intervention in matters of expression.[6] This foundational belief has shaped one of the most influential legal doctrines governing the internet i.e. Section 230 of the Communications Decency Act (CDA), 1996. Section 230, often described as the "First Amendment of the Internet," provides broad immunity to online platforms by treating them as intermediaries rather than publishers of third-party content.[7] This means that platforms are not legally responsible for what users post. Section 230(c)(1) explicitly states that "no provider or user of an interactive computer service shall be treated as the publisher

> Governments and civil society began questioning whether the "freedom to innovate" had come at too high a cost.

or speaker of any information provided by another content provider," and Section 230(c)(2) further protects the platforms' right to moderate content in good faith including the removal of offensive, harmful, or objectionable material without losing immunity.[8]

This law was designed to foster innovation and free expression during the early years of the internet, ensuring that new platforms could emerge without the fear of constant litigation. The First Amendment further strengthens this approach by prohibiting government censorship, thereby protecting both individual users and private companies from state interference. This principle was further reaffirmed in the landmark Supreme Court case, Moody v. NetChoice (2024)[9], where the Court struck down state laws attempting to restrict platforms' content moderation practices, the court recognised social media platforms discretion similar to editorial discretion of traditional publishers. This ruling also underscored the U.S.'s commitment to treating online platforms as private entities with constitutional protections of free speech, rather than public utilities subject to strict regulation.

The disinformation campaigns that circulated widely during the 2020 and 2024 U.S. elections highlighted the harmful impact of algorithm-driven amplification and underscored how difficult it is to strike a fair balance between protecting free expression and ensuring public accountability in the digital space.

However, Section 230's sweeping protections are not absolute. They do not extend to violations of federal criminal law such as the distribution of child sexual abuse material or to intellectual property infringements, which are instead governed by the Digital Millennium Copyright Act (DMCA). Over time, the once-celebrated immunity provision has become one of the most contested areas of internet law, as the digital ecosystem has evolved and online harms have multiplied. Many critics believe that the broad immunity granted under Section 230 has allowed online platforms to avoid taking responsibility for the spread of false information, hate speech, harassment, and deepfake content, all of which carry serious social and political repercussions.[10] The disinformation campaigns that circulated widely during the 2020 and 2024 U.S. elections highlighted the harmful impact of algorithm-driven amplification and underscored how difficult it is to strike a fair balance between protecting free expression and ensuring public accountability in the digital space.[11]

2. European Union

The European Union (EU) has long taken a proactive and rights-based approach to regulating platforms, aiming to create a safer, fairer, and more transparent online environment. Unlike the United States, where platform liability is limited by Section 230, the EU's framework has evolved around the principles of accountability, user protection, and market fairness. The E-Commerce Directive (2000) was the EU's foundational law for online intermediaries. It introduced the concept of "safe harbour", protecting platforms from liability for illegal content uploaded by users as long as they acted "expeditiously" to remove it upon gaining actual knowledge. This notice-and-takedown model reflected an early attempt to balance innovation with responsibility, ensuring platforms could grow without constant litigation while still being compelled to respond to unlawful material.[12]

However, the rapid expansion of social media and online marketplaces revealed the limitations of this model. Platforms had become not just hosts but active curators of online content, shaping public discourse, amplifying misinformation, and influencing elections. The EU began recognizing that the earlier legal framework was insufficient to handle all emerging challenges such as hate speech, disinformation, targeted advertising, algorithmic bias, and monopoly power among tech giants.[13] This led to a major shift in the EU's digital regulatory landscape with the introduction of the Digital Services Act (DSA) and the Digital Markets Act (DMA) in 2022. The DSA focuses on transparency, accountability, and systemic risk management, imposing obligations on platforms to assess and mitigate harms related to disinformation, illegal content, and threats to fundamental rights. It mandates clearer content moderation procedures, transparency in algorithmic decision-making, and independent auditing for Very Large Online Platforms (VLOPs). Platforms must provide users with clear and easily accessible information about their terms of service, redressal mechanisms, and remedies in machine-readable language. When an order is issued by a judicial or administrative authority for the removal of illegal content or to disclose user information, the platform must inform the user about the available channels for appeal and redress. In cases of content removal or account suspension, users must

Platforms must provide users with clear and easily accessible information about their terms of service, redressal mechanisms, and remedies in machine-readable language.

have access to an internal complaint-handling system, which is free, electronic, and effective, ensuring fair recourse against moderation decisions. The DMA, on the other hand, addresses competition and market dominance, targeting "gatekeeper" platforms to prevent anti-competitive behaviour, ensure fair business practices, and open digital markets to smaller innovators.

Together, the DSA and DMA represent a paradigm shift from reactive to proactive platform governance approach to balance freedom of expression with responsibility by trying to prioritize both user safety and corporate transparency. As enforcement begins, the EU has positioned itself as a global standard-setter for digital regulation, influencing similar legislative efforts across regions.

3. India

India's journey toward platform governance has evolved through three distinct phases from an initial light-touch framework designed to facilitate innovation, moving toward regulatory intervention, and finally entering a phase of assertive state oversight.

"

India's journey toward platform governance has evolved through three distinct phases from an initial light-touch framework designed to facilitate innovation, moving toward regulatory intervention, and finally entering a phase of assertive state oversight.

The Information Technology Act, 2000 (IT Act) was enacted to address the needs of a rapidly digitizing economy. Initially, the Act's focus was on enabling e-commerce, recognizing digital signatures, and defining cyber offences, rather than regulating intermediaries. The question of intermediary liability emerged prominently in Avnish Bajaj v. State[14], in this case the CEO of the e-commerce platform Bazee.com, was held liable for obscene material uploaded by a user. The case exposed serious gaps in the law particularly around what constitutes "knowledge" and how far a platform is responsible for third-party content. In response, the Information Technology (Amendment) Act, 2008 introduced a redefined Section 79, establishing a "safe harbour" provision. This shielded intermediaries from liability for user-generated content, provided they exercised due diligence and acted upon receiving "actual knowledge" of illegality. However, the interpretation of "actual knowledge" soon became contentious. Many intermediaries began removing content preemptively upon receiving private complaints, often without judicial oversight, leading to widespread over-censorship and

a chilling effect on free speech. This issue was addressed in Shreya Singhal v. Union of India[15] wherein the court held that intermediaries are required to take down content only upon receiving a court order or a notification from a government agency authorized under Section 69A, rather than by private complaints.

A new era began with the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, which expanded the regulatory scope significantly. These rules brought social media platforms, OTT services, and digital news publishers under one framework and imposed stringent due diligence obligations. Platforms must acknowledge complaints within 24 hours and comply with takedown orders within 36 hours of notification. An intermediary or any person who fails to assist the agency shall also be punishable with 7 years or given fines. These provisions can subject individual employees to personal and criminal liability.

"

A new era began with the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, which expanded the regulatory scope significantly. These rules brought social media platforms, OTT services, and digital news publishers under one framework and imposed stringent due diligence obligations.

Another noteworthy issue is Rule 16 of the Information Technology (Procedure and Safeguards for Blocking for Access of Information by Public) Rules, 2009 which has emerged as a major point of concern for transparency and due process. It empowers the government to issue confidential takedown and blocking orders, prohibiting intermediaries from disclosing any details about such directives, including the identity of the authority issuing them or the nature of the content taken down. This absolute confidentiality requirement effectively overrides the principles of transparency, accountability, and natural justice. As a result, users whose content has been blocked are often not informed or given an opportunity to contest the decision.

In July 2025, the Karnataka government proposed the Misinformation and Fake News (Prohibition) Bill, 2025 to curb online misinformation and hate speech. While the aim is valid, the Bill's unclear and overly broad wording is problematic. It allows platforms to be punished for "knowingly or unknowingly" aiding, abetting, or assisting in the commission of a hate crime. Merely allowing use of their platform can hold them liable, thus, running counter to Shreya Singhal v. Union of India, which held that platforms are liable only when they have actual knowledge of illegal content through a court or government order. Ignoring this principle could erode the safe harbour protections under

Section 79 of the IT Act and lead to arbitrary censorship. If passed, the Bill could set a troubling precedent by expanding government power over online speech and increasing the liability of digital platforms.

4. Brazil

In Brazil, platforms are governed under the Marco Civil da Internet, also known as Brazilian Civil Rights Framework for the Internet ("BCRFI"). As per the BCRFI, platforms will incur civil liability if they do not remove unlawful content (generated by users) after a specific order from the courts. Under exceptional circumstances, platforms will be required to remove content that invades a person's privacy by disseminating non-consensual intimate images ("NCII") once the user reports the same to the platform. In cases involving child pornography, platforms might face criminal liability if they fail to remove such content upon the receipt of notice. The executive arm does not have powers to issue content takedown orders to social media platforms. [16]

> "Platforms may bear a greater burden to prove they acted diligently in removing infringing or harmful content. Platforms must establish self-regulation systems with notice and appeal mechanisms, publish transparency reports, create user-friendly complaint channels, and appoint local legal representatives.

In June 2025, Brazil's Supreme Federal Court (STF) significantly changed the platform governance framework under the Marco Civil da Internet.[17] The court partially struck down Article 19, which had previously shielded platforms from civil liability for user-generated content unless they failed to comply with a specific court order. The Court held that platforms can be held liable for serious offences such as hate speech, terrorism, child sexual abuse, and gender-based violence even without a prior judicial order.[18]

The ruling emphasized that platforms will have to bear presumption of liability in cases involving paid promotion or automated dissemination (e.g., bots). In such situations, platforms may bear a greater burden to prove they acted diligently in removing infringing or harmful content. Platforms must establish self-regulation systems with notice and appeal mechanisms, publish transparency reports, create user-friendly complaint channels, and appoint local legal representatives in Brazil with authority to respond to judicial and administrative proceedings.

In essence, the STF's ruling marks a turning point in Brazil's digital governance. It moves the country away from a passive, court-order-dependent model toward one of shared responsibility, where platforms must act with greater care and transparency.

5. China

Analysts often describe three major approaches to internet governance — the U.S. model of minimal regulation, the European model of balanced regulation, and the Chinese model of strong state control. While this may oversimplify global differences, China does stand out for its unique combination of domestic tech power and government-driven content regulation. The Chinese government takes a proactive role in regulating both platforms themselves and the content hosted on them. Regulations such as the 2020 Provisions on Ecological Governance of Network Information Content require platforms to remove harmful material, undergo compliance inspections, and regularly report on content moderation practices.[19] China has long maintained tight control over its digital ecosystem, with strict censorship laws that demand platforms comply with the Great Firewall—a vast system of surveillance and content regulation.[20] Global platforms like Google and Facebook have exited China due to the impossibility of operating within China's strict censorship regime while adhering to international norms on free speech.

Platforms that operate within China, like WeChat, supposedly censor and surveil texts shared within the platform. Platforms also need to train personnel to conduct human reviews of uploaded content. If they do not fulfill their monitoring responsibilities, they risk consequences like warnings, fines, service suspension, and the cancellation of permits or licenses for conducting business.[21]

> "Global platforms like Google and Facebook have exited China due to the impossibility of operating within China's strict censorship regime while adhering to international norms on free speech.

Conclusion

Across jurisdictions, the governance of digital platforms reveals one central truth, the struggle to balance freedom, accountability, and safety in the digital age is far from settled. Each country's approach is a reflection of its political philosophy, social priorities, and constitutional values. The United States continues to champion free expression and innovation,

often at the cost of weaker accountability mechanisms. The European Union's rights-based framework, through the Digital Services Act and Digital Markets Act, prioritizes transparency, systemic risk assessment, and user protection. India, navigating a complex digital and political landscape, is steadily moving toward a model of state-led oversight and personal liability that raises critical concerns around censorship and due process. Brazil's evolving jurisprudence demonstrates a shift toward shared responsibility between platforms and the state, while China's centralized model illustrates the consequences of absolute state control where speech is tightly monitored and dissent systematically curtailed.

Another growing trend in global platform regulation is the rise of global or extraterritorial takedown orders, where courts or governments in one country direct platforms to remove certain content not only within their borders but worldwide. While intended to prevent harmful or unlawful material from resurfacing in other regions, these orders often raise complex questions of jurisdiction, sovereignty, and free expression. These orders also reflect a growing assertion of national sovereignty in the digital space and the state's intent to control how information circulates across platforms that operate globally. Such orders raise complex questions about jurisdiction, free expression, and the balance between local laws and global norms. When one nation's standards are applied globally, it risks overreach, conflicting legal obligations for platforms, and the gradual fragmentation of the open internet.[22]

"

When one nation's standards are applied globally, it risks overreach, conflicting legal obligations for platforms, and the gradual fragmentation of the open internet.

What emerges from these comparative frameworks is a global dynamic of the relationship between platforms, governments, and users. Platforms are no longer neutral intermediaries, they are active contributors of information, influence, and power, which is why they should not be allowed to operate without transparency or accountability. However, excessive regulation and criminal liability risks can equally undermine digital civic space and suppress legitimate expression. The challenge, therefore, lies in designing governance systems that preserve the internet's openness while ensuring that those who design and profit from it remain answerable to the society.

In the coming years, the global debate will not merely be about moderating content, but about defining who moderates the moderators, how to safeguard human rights, democratic discourse, and digital autonomy in an era where platforms function as both the marketplace and the medium of modern life. The goal must be to create frameworks that are rights-respecting, transparent, and adaptable, so that the internet remains a safe public space.

The Platform Question: Power, Accountability and Global South

Endnotes:

[1] Olaniran B and Williams I, "Social Media Effects: Hijacking Democracy and Civility in Civic Engagement," Rhetoric, Politics and Society (Springer International Publishing 2020)

[2] Sylvia Lu, Algorithmic Opacity, Private Accountability, and Corporate Social Disclosure in the Age of Artificial Intelligence, 23 Vanderbilt Journal of Entertainment and Technology Law 99 (2020)

[3] Khondker, H. H. (2011). Role of the new media in the Arab Spring. Globalizations, 8(5), 675–679

[4] "Facebook Admits It Was Used to 'incite Offline Violence' in Myanmar" (BBC News, November 6, 2018)

[5] Youn S, "Facebook Admits Its AI Failed to Flag the New Zealand Terror Attack Livestream" ABC News (March 21, 2019)

[3] Khondker, H. H. (2011). Role of the new media in the Arab Spring. Globalizations, 8(5), 675–679

[4] "Facebook Admits It Was Used to 'incite Offline Violence' in Myanmar" (BBC News, November 6, 2018)

[5] Youn S, "Facebook Admits Its AI Failed to Flag the New Zealand Terror Attack Livestream" ABC News (March 21, 2019)

[6] Volokh, E. (2023). Free Speech in the Digital Age. UCLA Law Review, 70(4), 987–1023.

"

In the coming years, the global debate will not merely be about moderating content, but about defining who moderates the moderators, how to safeguard human rights, democratic discourse, and digital autonomy in an era where platforms function as both the marketplace and the medium of modern life.

[7] Kosseff, J. (2019). The Twenty-Six Words That Created the Internet. Cornell University Press

[8] Goldman, E. (2023). In Defense of Section 230. Stanford Technology Law Review, 26(3), 301-325.

[9] Moody v. NetChoice is 594 U.S. 900 (2024)

[10] Johnson A and Castro D, "Fact-Checking the Critiques of Section 230: What Are the Real Problems?" Information Technology and Innovation Foundation | ITIF (February 222021)

[11] Wendling M, "Voter Fraud Claims Flood Social Media before US Election" (BBC News, November 3, 2024)

[12] López J, Phd R and Richart JL, "A New Legal Framework for Online Platforms in the European Union (and Beyond)" (Katolicki Uniwersytet Lubelski Jana Pawla II, December 20, 2024)

[13] European Commission, Commission Staff Working Document: Impact Assessment Accompanying The Document Proposal For A Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and Amending Directive 2000/31/Ec, Page 21.

[14] Avnish Bajaj v. State (NCT of Delhi), (2005) 116 DLT 427

[15] Shreya Singhal v Union of India, (2015) 5 SCC 1.

[16] X was suspended for its failure to nominate a legal representative within Brazil. See Lisandra Paraguassu, Luana Maria Benedito and Ricardo Brito, "Brazil watchdog moves to block access to Elon Musk's X after court order" (Reuters, 31 August 2024)

[17] Extraordinary appeal (re) no. 1,037,396 (theme 987) and Extraordinary appeal (re) no. 1,057,258 (theme 533)

[18] Action D, "A Turning Point for Platform Responsibility in Brazil" (Digital Action, July 17, 2025)

The Platform Question: Power, Accountability and Global South

[19] Gorwa, Robert, 'Platform Regulation and the Majority World', The Politics of Platform Regulation: How Governments Shape Online Content Moderation, Oxford Studies in Digital Politics (New York, 2024; online edn, Oxford Academic, 23 May 2024)

[20] House F, "China" (Freedom House), Freedom on the Net 2021 <https://freedomhouse.org/country/china/freedom-net/2021#footnote2_m4g-24py>

[21] Xiao B, "Making the Private Public: Regulating Content Moderation under Chinese Law" (2023) 51 Computer Law & Security Review 105893

[22] Chang S, "Global Takedown Orders in the GDPR Era: New Internet Governance on Transnational Cooperation" (April 15, 2025)

CHAPTER 2

A Big Tech Accountability in Content Moderation with Feminist Perspective

By Karen Vergara

On June 12, 2018, the meeting "Online Safety for Women" was held in Buenos Aires, Argentina, organized by Facebook[1] and managed through the company's Security Policy and Programs Management for Latin America. The event brought together more than thirty Latin American organizations, mostly feminist, linked to the defense of human rights and the prevention of gender-based violence facilitated by technology. The one-day event was presented as a space for knowledge exchange and network strengthening, along with the presentation of Facebook's female team in the area of ***Trust & Safety.*** The event passed quickly, with endless presentations by the company and little space for civil society organizations to express their concerns and showcase their work. This highlighted the tensions inherent in the relationship between technology corporations and feminist organizations: while the former appealed to grandiloquent discourses of social responsibility and security at this meeting, the latter demanded structural changes to the platforms that would effectively address the multiple forms of violence faced by women and dissidents in the digital space, as well as greater transparency in the functioning of their reporting and complaint models.

As an NGO, we documented some of the main needs of civil society organizations and activists at that meeting, which we summarize briefly in the following points:

- Real content moderators: a specific team to address the main complaints related to digital sexual violence, gender-based violence facilitated by technology, and other types of events with a high emotional and digital impact.
- Reporting, complaint, blocking, and usage tutorial channels that incorporate the Spanish language, as well

> "The event "Online Safety for Women" highlighted the tensions inherent in the relationship between technology corporations and feminist organizations.

as the particularities of each dialectal variant by country and region.
- Incorporate a gender and human rights perspective in the development of each stage of the technologies.

After that first approach in 2018, a second meeting was organized by the company in Chile in 2019, as our country was being targeted to consolidate a growth hub for Facebook and Instagram. This time, the general announcements pointed to the possibility of developing a collective alliance to more quickly address cases of technology-facilitated gender-based violence.

The possibility for civil society organizations was reduced to having a direct contact at Facebook to address complex cases not resolved by the platform's security policies in the first instance (appeals, blocks, deactivated accounts, usurpation, among others), and which required extra verification of trust. This direct contact lasted about a year and a half. Far from specifically addressing the needs of organizations, it meant more extra (and unpaid) work for civil society organizations, specifically those that ran feminist helplines. Our organization, for example, operated with two colleagues checking email and social media daily, providing support to victims of digital violence, and documenting cases that had not received a response from the platforms in order to contact their Trust & Safety teams.

"

Far from specifically addressing the needs of organizations, it meant more extra (and unpaid) work for civil society organizations, specifically those that ran feminist helplines.

The main cases of technology-facilitated gender-based violence that we observed during that period were directly linked to the social context in Chile. As the feminist protests that began in universities in 2018 spread to schools, workplaces, and public discourse, we began to accompany increasingly frequent cases of feminist groups losing access to their accounts, coordinated digital attacks by hordes of individuals, the exposure of their personal and intimate data, and the sending of sexually aggressive content to their emails and messaging services. If 2018 was the year of the feminist movement, 2019 was the year of widespread social unrest. This new scenario amplified and transformed the digital violence that activists were already suffering. The social uprising that began on October 18 in Chile, in response to the high cost of living, not only filled the squares, but also social media

with evidence of brutal repression: beatings, illegal coercion by state agents, and more than 460 eye injuries caused by police gunfire. It was in this environment of high visibility and online confrontation that platforms also consolidated themselves as a battlefield and a source of identity. Violence was no longer limited to harassment for being a feminist, but was compounded by persecution for participating in the uprising. While social movements used platforms to organize protests and aid, their opponents used them to coordinate attacks, engage in doxing, and make threats. The impunity that characterized the repression in the streets was mirrored in the platforms' inaction in the face of digital violence, creating a risky ecosystem where any woman in the public eye became a potential target on all fronts. The Observatory Against Street Harassment (OCAC), a Chilean organization that addresses violence and harassment in public spaces, noted at the time:

"Social media and our website are the fastest way we have to reach the women who follow us. That's where we publish the activities we're doing, and we consider it a channel through which people know they can write to us. For this reason, we have to deal with a lot of misogyny from men who try to take down our site or insistently write on every post we publish to saturate our accounts with hate messages and thus prevent the message from reaching the women and teenagers who need it."

As an NGO, Amaranta developed a research project during that period called "Aurora," which allowed us to conduct a series of educational workshops for feminist organizations and the LGBTIQ+ community, focusing on digital literacy, how to correctly report content that violates the rules and policies of large platforms, and how to use the same tools to their advantage for these reports (for example, it is easier to report humiliating content for copyright infringement than for misogyny). As we approached 2020, the constant communication channels we had with Facebook began to fail. High staff turnover made it impossible to reconnect with the platform. Added to its merger and name change, contacts and communication channels were lost, and cases of gender-based violence facilitated by technologies that were not addressed in their reporting instances began to increase. Their teams remained deployed in Brazil and Mexico, but the

"

While social movements used platforms to organize protests and aid, their opponents used them to coordinate attacks, engage in doxing, and make threats.

Andean region and the Southern Cone were left adrift. There was no longer anyone human to refer cases involving their platforms to. Eva Cruells, founder and coordinator of Fembloc, a feminist initiative dedicated to combating digital gender-based violence, told the Spanish media outlet Público:

"The thing is, the reporting mechanisms are very opaque and often unresponsive. We have our own data: of the total number of cases we have reported, only 4% have received any affirmative response from the platform. Because often they are bots or automated emails. Meta, for example, has greatly reduced its content moderation staff."

We revealed a similar situation as NGO Amaranta in 2020, with the study Chile and gender-based violence on the internet,[4] where we revealed that 11% of victims of doxing, digital harassment, cyberflashing, defamation, and loss of access to their account or unauthorized access to it, tried to report it to the platforms without success. When victims were asked what actions they took with the platforms, the vast majority said that the reporting sites were in English, did not respond to them, or had bugs that made them return to the same page to fill in their details over and over again.[5]

Who takes care of what we don't want to see on the internet?

The history of digital moderation reveals not only the technical complexity of the process, but also its profound human inequalities.

The history of digital moderation reveals not only the technical complexity of the process, but also its profound human inequalities. Meta began filtering content in 2007, after New York Attorney General Andrew Cuomo demanded that sexual images proliferating on Facebook be removed within 24 hours. Since then, the company has outsourced the task to subcontractors, some of which, such as oDesk, paid moderators in Morocco as little as $1 an hour, a situation that is repeated in several countries. Behind the algorithms that protect the audience, there are people exposed daily to traumatic content without psychological support or decent working conditions. A 2019 article in The Verge reveals how workers at a Cognizant subsidiary in Tampa faced precarious working conditions, nine-minute breaks, the death of a co-worker from a heart attack at his desk, and even how they had to explicitly approve certain violent content simply because it did not comply with Facebook's moderation policies.[6] Many

The Platform Question: Power, Accountability and Global South

of these workers joined the platform under the promise of job stability and believing that they were performing a protective role for social media users. However, one animal rights activist worker recounts in the article that he had to watch a video showing a case of extreme animal torture by teenagers repeatedly. The platform's policies did not allow him to take it down, and his superiors consoled him by pointing out that keeping it posted could prompt law enforcement authorities to launch an investigation and find the culprits. That never happened. The case highlights how moderation policies, rather than protecting, end up reproducing other forms of violence and trauma. For example, it is not possible to remove a sexually aggressive video without a valid argument as to why it cannot be shown. Spanish lawyer Francesc Feliu argues that this precision is intended to train AI. "It cannot be trained if the content it feeds on is violent, pedophilic, or murderous," he adds. Feliu legally represents more than forty workers dedicated to moderating META content in Spain, who have jointly sought compensation for the physical and mental damage to which they are exposed in this work. The lawyer points out that most of his clients suffer from anxiety disorders, depression, and post-traumatic stress, and that several are even admitted to psychiatric centers. All of them live with the pressure of not being able to talk openly about the problem, out of shame or fear of reprisals from the company, due to the confidentiality clauses they signed.[7]

This reality is repeated on other platforms such as TikTok. In 2022, it was revealed that some of the staff responsible for moderating content must follow accelerated review protocols (sometimes even seconds per video) to keep up with the massive volumes of posts they are asked to analyze every day[8] in categories such as "explicit violence," "hate speech," and "sensitive content."[9] Automated language models are no exception to this reality. The work of content moderators for ChatGPT follows this same pattern of bad practices and job insecurity. In Kenya, workers have revealed that they review up to 700 passages of text per day, many of them laden with explicit sexual violence, suicide, murder, child abuse, and bestiality. They report having lost their families in the process and not having any psychological support to cope with the situation. While large companies advertise technological advances that promise to free society from arduous tasks, that same development depends on people who bear the emotional burden of working with the dark side of the internet.

While large companies advertise technological advances that promise to free society from arduous tasks, that same development depends on people who bear the emotional burden of working with the dark side of the internet.

When moderation gets involved with politics

In 2024, The Guardian revealed testimonies from Meta workers who denounced unfair moderation policies toward the Palestinian community.[10] According to an anonymous employee and a letter signed by more than 200 people, internal rules known as Community Engagement Expectations (CEE) are used to remove discussions about Gaza, even when they express condolences or question product errors that affect the visibility of Palestinian voices. The document contrasts the institutional silence in the face of more than 35,000 civilian deaths in Palestine with the public support the company expressed toward its Israeli employees after the attacks of October 7, 2023.

According to an anonymous employee and a letter signed by more than 200 people, internal rules known as Community Engagement Expectations (CEE) are used to remove discussions about Gaza, even when they express condolences or question product errors that affect the visibility of Palestinian voices.

The signatories accuse a long history of biased moderation against Palestinian content, despite calls from Human Rights Watch and Senator Elizabeth Warren to ensure greater transparency. In their letter, they demand an end to internal censorship, recognition of Palestinian suffering, and a public statement calling for a ceasefire. The report also exposes the structural fragility of the moderation system: human decisions must align with Meta's policies, but the application of those rules varies depending on language and geographic context. Moderation, therefore, is not just a technical or automated exercise, as it also faces tension between corporate standards, international treaties, and specific local contexts.

And in Latin America?

Returning to our experience, from 2023 to the present, we have noticed a steady increase in requests for help sent to the Amaranta NGO's email address, not only from Chile but from various Spanish-speaking countries. Seeking to systematize the cases and trying to find the source of these requests, we discovered that an organization that had been providing services to Meta for some time recommended us as partners for cases of technology-facilitated gender-based violence. No one consulted us or informed us of this. This replicates what other civil society organizations denounce, which is how a multimillion-dollar corporation shifts responsibility to small organizations in the global south. This fact accurately sums up the extractive and unequal logic of today's digital ecosystem.

Week after week, we continue to receive cases of digital sexual violence, non-consensual dissemination of intimate images, and identity theft, including cases from Arab countries. We try to respond to each message with humanity, knowing that behind every automated form that did not work for the victims, there is a story that the platforms do not want to address. Ivana Feldfeber points out the following in the prologue to the report Helplines for addressing cases of gender-based violence in digital environments[11] :

"It is essential to consider the role of technology companies, which facilitate the use of these platforms. And at this point, it is necessary to emphasize that companies do not give us free products altruistically; they are not mere service providers to society. Rather, behind these social networks are large economic interests, with millions of dollars in contracts with companies, governments, and other groups to offer advertising and traffic their users' data (...) These companies, when asked to block users, remove content disseminated without consent, or recover hacked accounts, they often do not comply with these requests. The same companies that censor images of female nipples on their platforms ignore calls for help from their users who are victims of digital gender-based violence."

Meta, like other companies in the sector, outsources digital violence to precarious and subcontracted workers, making their suffering invisible and normalizing technological progress based on the emotional precariousness of those who moderate content and also train artificial intelligence to appear more human for the same price.

Meta, like other companies in the sector, outsources digital violence to precarious and subcontracted workers, making their suffering invisible and normalizing technological progress based on the emotional precariousness of those who moderate content and also train artificial intelligence to appear more human for the same price, without forgetting that they end up delegating their responsibility for digital violence to civil

Conclusion

The discussion about content moderators is becoming political and requires immediate attention. Who decides what is acceptable to see on the internet? And at what human cost is that decision sustained? Can we talk about safe digital spaces if this task is left in the hands of precarious workers with low wages and little psychological protection? How can we demand accountability from platforms that includes all these aspects if they are currently invisible due to outsourcing?

To quote Sonia Reverter:

"We feminists have learned that the struggle to destabilize the patriarchal system requires large, but also small proposals; collective projects, but also personal ones. Because patriarchy is a system that organizes and acts on both the macro and micro levels, we must respond with feminist actions on all fronts. The reorganization of the symbolic is perhaps the heaviest of the tasks to be carried out, because the burden of the symbolic is massive for the human constitution, both individually and collectively." (Reverter, 2013, p.458)

Adopting a feminist stance in the face of the crisis of content moderation on platforms involves analyzing the issue in a situated way, recognizing the impact that large platforms have on people's lives, and how precariousness, inequality, and class intersect with outsourcing, to outsource opaque work for which we must demand the highest ethical standards of operation. To do this, we need open data, comprehensive reports, and external audits carried out by civil society organizations, international human rights organizations, and academia, which study the variables involved in content moderation in an intersectional manner, as well as its implications for the lives of its workers. It is not possible for technological advances to serve only to increase capital, while preying not only on the earth, but also on the bodies of those who face the worst of

It is not possible for technological advances to serve only to increase capital, while preying not only on the earth, but also on the bodies of those who face the worst of the internet.

Endnotes:

[1] In 2018, the company operated under the name Facebook, Inc. The name change to Meta Platforms, Inc. took place in October 2021 as a way to expand the brand into the metaverse, including the WhatsApp, Instagram, and Oculus platforms, among others.

[2] Interview conducted in 2018 with Melissa Gutiérrez, OCAC communications director, for the report "Lazos activistas en red, el ciberfeminismo chileno" (Activist Network Ties, Chilean Cyberfeminism), authored by me.

[3] Falcó, O. (September 13, 2025). Eva Cruells: "Digital gender-based violence is normalized as if it were bar talk, but it can have an impact as serious as physical assault." Diario Público. https://www.publico.es/mujer/violencia-machista/violencias-machistas-digitales-normalizan-conversacion-bar-pero-pueden-tener-impacto-grave-agresion-fisica.html

The Platform Question: Power, Accountability and Global South

[4] Chile and gender-based violence on the internet, experiences of cis and trans women in digital spaces https://amarantas.org/wp-content/uploads/2020/08/informe-proyecto-aurora.pdf

[5] Ananías, C. Vergara, K. et al (2023). Digital gender-based violence in Chile: a study during the COVID-19 pandemic. Sexuality, Health, and Society, (39). https://doi.org/10.1590/1984-6487.sess.2023.39.e22306.a.es

[6] Newton, C. (June 19, 2019). Bodies in seats. The Verge. https://www.theverge.com/2019/6/19/18681845/facebook-moderator-interviews-video-trauma-ptsd-cognizant-tampa

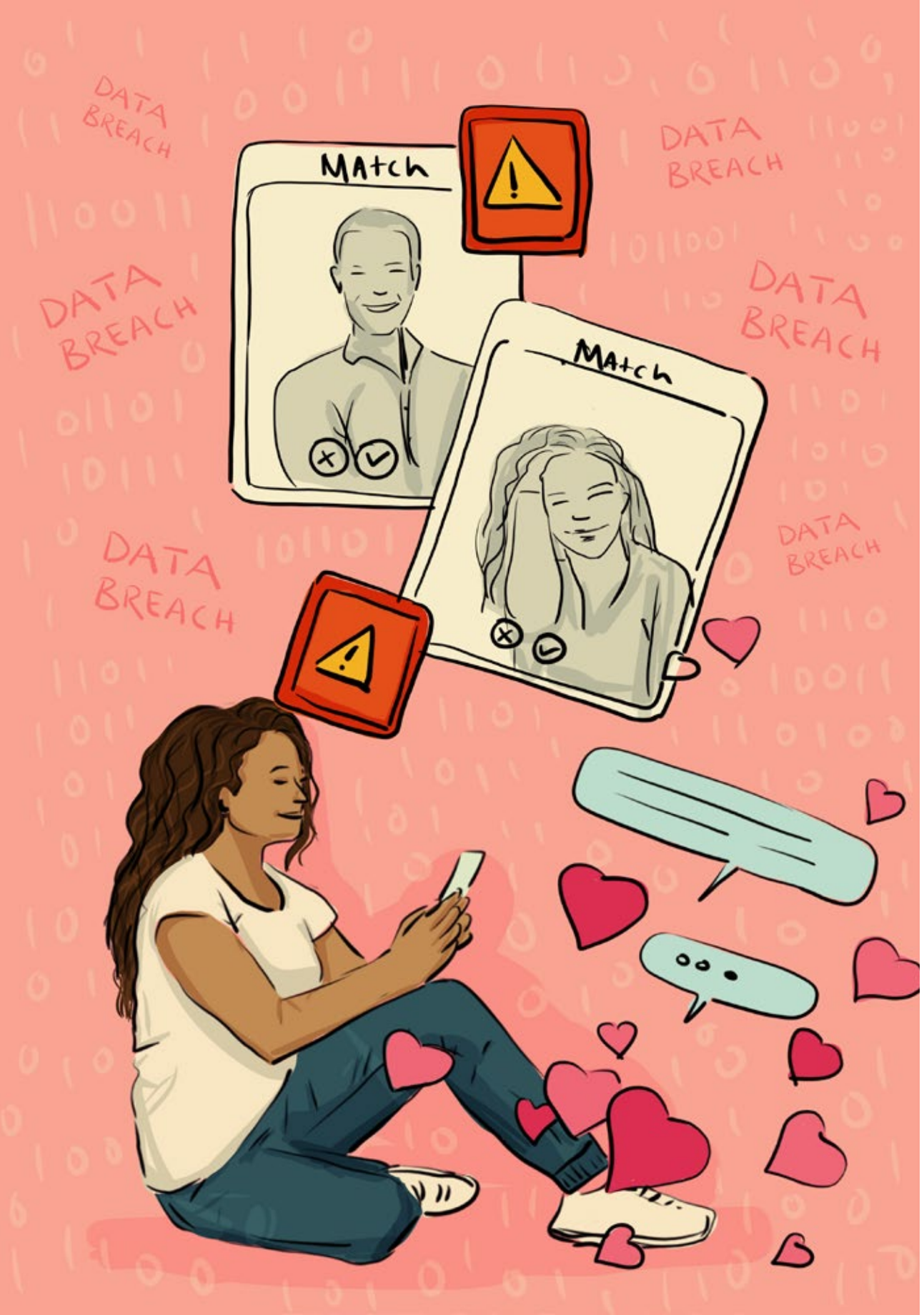
[7] Dalfó, L. P. (May 6, 2025). The lawyer for Meta's content moderators: "They have anxiety, post-traumatic stress, and some are in psychiatric hospitals." Ediciones EL PAÍS S.L. https://elpais.com/espana/catalunya/2025-05-06/el-abogado-de-los-moderadores-de-contenidos-de-meta-tienen-ansiedad-estres-postraumatico-y-alguno-esta-ingresado-en-un-psiquiatrico.html

[8] Farah, H. (2023, December 21). Diary of a TikTok moderator: 'We are the people who sweep up the mess'. The Guardian. https://www.theguardian.com/technology/2023/dec/21/diary-of-a-tiktok-moderator-we-are-the-people-who-sweep-up-the-mess

[9] 'It's destroyed me completely': Kenyan moderators decry toll of training of AI models https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai

[10] Paul, K. (2024, August 15). Meta struggles with moderation in Hebrew, according to ex-employee and internal documents. The Guardian. https://www.theguardian.com/technology/article/2024/aug/15/meta-content-moderation-hebrew

[11] Derechos Digitales. (2024). Helplines to address cases of gender-based violence in digital environments: Monitoring and trends in Bolivia, Brazil, and Ecuador. https://www.derechosdigitales.org/wp-content/uploads/LineasAyuda-ESP.pdf



CHAPTER 3

Platform Accountability in Online Dating: A Critical Analysis of Privacy, Discrimination and Safety Harms in India

By Tavishi and Angelina Dash

Introduction

The expansion of online dating platforms in the Global South brought with it the promise of greater agency for users in societies where romantic desire and sexual autonomy are strictly policed.[1] The last decade alone has witnessed a significant increase in the presence of dating platforms in India, including both global dating platforms like Tinder and Bumble,[2] as well as homegrown dating platforms like Aisle and Quack Quack.[3] Unfortunately, this expansion has been accompanied by user harms such as financial fraud,[4] casteist abuse,[5] and sexual harassment,[6] especially targeting women and members of marginalised communities.[7] It thus becomes imperative for dating platforms to provide meaningful accountability and effective design interventions towards creating safe, inclusive and equal spaces for online dating. In this essay, we focus on three interlinked categories of user risks: privacy violations, discriminatory practices, and safety threats; and we contextualise these risks to India's unique socio-political and cultural context.[8] We analyse these risks primarily through secondary literature, supplementing our findings with a case study on two understudied popular homegrown dating platforms,[9] Aisle[10] and Quack Quack.[11] We conclude the essay by reiterating the need for greater accountability for dating platforms, briefly outlining recommendations for transparency in content moderation practices and algorithmic curation.

“The expansion of online dating platforms in the Global South brought with it the promise of greater agency for users in societies where romantic desire and sexual autonomy are strictly policed.

Data Privacy

Globally, online dating platforms collect and process highly personal and sensitive user data, which can include personal identifiers and interests, device metadata, precise location data, demographic information like religious and political beliefs, and sometimes biometric and financial data.[12] They often collect additional data through social media integration[13] and user behaviour like swiping habits and frequency of app use.[14]

While access to some of this data can support user experience, many dating platforms share data with third parties for targeted advertising.[15] A Mozilla Foundation study of 25 dating applications popular across North America and Europe found that 80% of them may share or sell user data to third parties for advertising.[16] A 2020 report by the Norwegian Consumer Council found that popular dating platforms like Grindr, Tinder and OKCupid share sensitive personal information, including their users' exact location, sexual orientation, political beliefs, etc., to third parties.[17]

A 2020 report by the Norwegian Consumer Council found that popular dating platforms like Grindr, Tinder and OKCupid share sensitive personal information, including their users' exact location, sexual orientation, political beliefs, etc., to third parties.

Often, users are unaware of the nature and extent of such data sharing. Many platforms' privacy policies do not clearly disclose data-sharing practices,[18] and these clauses often remain hidden within obscure and dense terms of service. Often, the privacy policies do not mention details of these third parties, and even in the few cases where company names are explicitly provided, users must read the privacy policies of these third parties to truly comprehend how they process their personal data.[19] Studies on privacy policies across platforms and regions have found that users pay little attention to them, and that these policies fail to provide meaningful consent, often on account of complex legalese and information overload.[20]

In the EU, much of the data collected by dating platforms falls within what the General Data Protection Regulation (GDPR) recognises under a "special category", requiring explicit consent for collection and processing, as it may heighten risks of discrimination and infringe on fundamental rights of citizens.[21] In the Indian context, the personal data of users on online dating platforms has been particularly vulnerable in the absence of a data protection law, which, while enacted two years ago,[22] has still not been enforced. Further, users in India are less likely to benefit from additional safeguards accorded to sensitive personal data even after the Digital Personal Data Protection Act 2023[23] comes into force, given the absence of a separate category of protection for sensitive personal data.[24, 25]

Moreover, India's data protection framework does not specify mechanisms for obtaining granular consent, unlike other major data protection frameworks across the globe,[26] forcing users to either share all requested data or entirely forgo services. This becomes all the more relevant as several dating platforms are integrating AI, in addition to complex matching algorithms, within their services in the form of chatbots,[27] picture selection tools, and profile or messaging feedback.[28] However, the use of personal data for AI features often leads to the sharing of data with third parties. For instance, recently, the European nonprofit noyb filed a complaint with the Austrian Data Protection Authority against Bumble's processing of personal data for its AI Icebreakers feature, which relies on OpenAI's ChatGPT to aid users in starting conversations.[29]

In the past, the US Federal Trade Commission (FTC) filed a petition for information about a data-sharing deal between the Match Group's OkCupid and Clarifai AI that enabled the training of facial recognition software without users' knowledge or consent.[30] Without adequate data security safeguards and consent mechanisms, privacy violations through data breaches and illicit data sharing can expose users to risks like identity leaks, surveillance, and reputational harm.[31] Women and marginalised communities face disproportionate risks from data breaches.

A 2023 study by Cybernews revealed that it was possible[32] for hackers to triangulate users' almost real-time locations by retrieving the last known location ID of any OkCupid user.[33] This vulnerability could have exposed women to stalking and sexual violence, and endangered members of the LGBTQIA+ community.[34] Dating platforms can also facilitate surveillance of sexuality in jurisdictions where homosexuality is criminalised. For instance, in Cairo, police officers detained a man simply for having downloaded same-sex dating apps on his phone.[35] In India, despite the decriminalisation of homosexuality,[36] queer individuals continue to face stigma and violence, with queer men reporting assaults, robberies, and blackmail through threats of being outed on platforms like Grindr.[37] Further,

India's data protection framework does not specify mechanisms for obtaining granular consent, unlike other major data protection frameworks across the globe, forcing users to either share all requested data or entirely forgo services.

women from marginalised communities can be particularly vulnerable to privacy breaches. In the recent past, pictures of Indian Muslim women, obtained from social media, have been mock auctioned on open-source apps derogatorily named Bulli Bai[38] and Sulli Deals.[39] These incidents have heightened the apprehension regarding the use of dating platforms amongst Muslim female users, especially given the opacity surrounding their privacy policies and redressal mechanisms.[40]

Case Study: Privacy Practices of Aisle and Quack Quack[41]

The following analysis examines the privacy practices of Aisle and Quack Quack on the basis of publicly available information in their privacy policies, terms of use and user interface.

“

In the recent past, pictures of Indian Muslim women, obtained from social media, have been mock auctioned on open-source apps derogatorily named Bulli Bai and Sulli Deals. These incidents have heightened the apprehension regarding the use of dating platforms amongst Muslim female users, especially given the opacity surrounding their privacy policies and redressal mechanisms.

	Aisle	Quack Quack
Collects sensitive personal data	Yes[42]	Yes[43]
Explicit user consent for processing sensitive personal data for specific purpose	No[44]	No[45]
Explicit consent to collect geolocation data	Yes	Yes[46]
Allows users to withdraw consent easily	Not Clear[47]	Not Clear[48]
Shares data with third parties for advertising	No[49]	Yes[50]
Allows users to opt-out of third-party data sharing	No[51]	No[52]
Specifies data retention duration	Yes[53]	No[54]
Users can request platform to delete their personal data	Yes[55]	Not Clear[56]
Data breach notification policy	No	No
Private Mode Feature allowing the profile to be hidden	Yes, your profile will only be visible to users you like/comment. However, this is a paid premium feature.	No, only certain details like age, location and profession can be hidden, and this option is only available to Quack Quack Plus subscribers as a paid feature.
Features allowing users to monitor how many people their matches have interacted with in the past few days	Yes, the Exclusivity Feature is included as a premium feature to Aisle users.[57]	

Discrimination

Romantic and sexual desire, although deeply personal, is entrenched in existing power structures and shaped by histories of oppression and exclusion.[58] With the emergence of online dating platforms, many saw the possibility to transgress existing racial and ethnic boundaries by dating outside of conventional friends and family networks.[59] However, romantic desire online continues to be mediated by gender, caste, class, race, ethnicity, religion, language, and so on.[60] In fact, these choices are often normalised and legitimised as personal preferences,[61] and further reinforced by opaque matching algorithms, platform policies and design.

Numerous studies in the Global North have highlighted the prevalence and normalisation of sexual racism[62] in online dating[63] and the resultant hierarchy of desirability[64] which privileges Whiteness.[65] This reinforces discrimination and the invisibilisation of racial and ethnic minorities, increasing their vulnerability to racial abuse, and even sexual harassment and physical violence.[66] Consequently, many global platforms have faced criticism for enabling users to express racial preferences in bios or through ethnicity filters.[67] Even when users do not engage in explicit filtering, matching algorithms of many prominent dating platforms use collaborative filtering,[68] which can potentially learn discriminatory racial preferences of users in dating through the feedback loop of user interactions, and amplify existing social norms of homophily.[69] This can result in the homogenisation of recommendations even for users who do not wish to be restricted by the dominant preferences of their racial group.[70] Dating platforms often create an illusion of choice, since users are often unaware of the extent to which the profiles presented to them are curated by the hidden logics of the opaque matching algorithms, especially when the recommendations conform to dominant societal expectations. In comparison to the research on discrimination based on racism in online dating in the Global North, there are relatively fewer studies on how discrimination manifests in dating platforms in India. This is especially relevant given that endogamy has been the defining feature of upholding caste-based segregation and discrimination.[71] This is reflected in the persistently low percentages of intercaste[72] and interreligious marriages.[73] Any transgression of caste or religious lines can result in violence, often directed at Dalit and Muslim

“

Dating platforms often create an illusion of choice, since users are often unaware of the extent to which the profiles presented to them are curated by the hidden logics of the opaque matching algorithms, especially when the recommendations conform to dominant societal expectations.

partners,[74] who are vilified as a threat to the honour of Savarna communities.[75]

While a large percentage of young users have used dating platforms to explore romantic relationships, often in secrecy,[76] there is limited research on how considerations around caste, class and religion shape their user behaviour. Although there is often an uncomfortable silence surrounding the question of swiping based on caste and religion reported across multiple studies in India,[77] participants in some studies confided that they would ultimately have to marry according to their parents' will.[78, 79] Some dating platforms have also come under criticism for providing caste filtering.[80] Even when dating platforms do not provide exclusive caste filtering, users may present their dominant caste identity or preferences in their profile or swipe based on markers that are often proxies for caste and class status in a highly unequal society.[81] The commodification of intimacy on dating platforms means that the construction of user profiles becomes an activity in self-branding using limited avenues for user information like profile pictures, bio, and text messages.[82] This scarce information often forms the basis for judging the desirability of potential matches. Dhanaraj notes how users may judge caste through markers like "surnames, localities, dialects, jobs of parents, religion, economic status, political and pop culture idols, food choices, ideology, complexion, and others."[83] In their study of Grindr in India and South Africa, Philip found that, while these platforms provide important sites for expressing sexual desire in societies where heteronormative patriarchy prevents such expression freely offline, they also result in the commodification of gay identities and a hierarchy between "classy gays" and "poor gays" through class, race and caste markers.[84] Users both perform affluence, and judge desirability through markers like Western brands in clothing, and trendy locations in profile pictures.[85]

"

The commodification of intimacy on dating platforms means that the construction of user profiles becomes an activity in self-branding using limited avenues for user information like profile pictures, bio, and text messages.This scarce information often forms the basis for judging the desirability of potential matches.

Similarly, studies note how terms like "creeps" and "weirdos" are used for those who use lower-quality images, whose English is unrefined, and who lack aesthetics and premium branding.[86] Kisana has also noted how fluency and command over the English language and access to Western pop culture are used to police caste boundaries in a society where such cultural capital often signifies intergenerational literacy and access to English medium schooling.[87] The geolocation feature in many dating platforms further consolidates segregation in a country where Dalit and Muslim families live in ghettoised margins across urban cities.[88]

Even when Dalit and Bahujan users match with Savarnas on dating platforms, they often have to face casteist abuse and humiliation,[89] and face heightened vulnerability to sexual harassment or physical violence. Many report being blocked or unmatched when dominant caste matches discover their identity.[90] Dalit women, who are often stereotyped as angry, unfeminine, and promiscuous, face heightened safety risks on dating platforms.[91] Paik has drawn parallels between the experiences of marginalisation faced by Black women in the US and Dalit women in India.[92] She notes that both White and Brahminical systems protected White and upper caste women's honour by restraining them from the public sphere, while sexual access to the bodies of black and lower caste women has been legitimised and institutionalised.[93]

While users may look at dating platforms as opportunities to exercise autonomy in romantic partnership through "modern ways", they often still rely on prevalent social norms to make dating decisions.[94] This user behaviour on platforms, which also feeds into algorithmic curation, together creates a socio-technical system that upholds caste, class, and religious endogamy. Dattani highlights how "endogamous social intimacies" are co-constructed by user behaviour and the algorithmic infrastructure of the platform.[95]

Case Study: Privacy Practices of Aisle and Quack Quack[41]

Community Guidelines and Prohibition of caste-based discrimination and casteist hate speech. It is telling that both Aisle and Quack Quack, designed specifically for Indian audiences, do not explicitly prohibit caste-based discrimination and casteist hate speech in their Community Guidelines.

Aisle prohibits, among other things, "threatening, harassing, racially offensive, or illegal material, or any material that infringes or violates another party's rights"

"

While users may look at dating platforms as opportunities to exercise autonomy in romantic partnership through "modern ways", they often still rely on prevalent social norms to make dating decisions.

in its Community Guidelines and Safety section nestled within the Terms of Service.

Quack Quack goes a little further and prohibits racism, religious discrimination and bigotry.[96] While broad categories of "discriminatory" and "bigoted" content should include casteist abuse, the lack of explicit prohibition is likely to discourage users from reporting harmful content.

From an analysis of the community guidelines of these two major dating platforms, it appears that although Dalit, Bahujan and Adivasi constitute the majority of the country's population, platforms are designed by and targeted towards only the upper-caste elite.

Safety

Dating platforms pose complex safety risks, including those that go beyond in-app interactions, due to factors like conversations between users transitioning to other platforms or offline interactions. Recently, systemic failures in Match Group's[102] management of sexual assault reports on its dating platforms in the US have come to light.[103] The investigation revealed that repeat offenders, including those who were reported for rape and violence, continued to operate on the group's platforms, easily creating new accounts when older profiles got banned. Another study cautioned against groomers and pedophiles using dating platforms to target single mothers to carry out child sexual abuse.[104]

> "Women often self-censor their online behaviour, and may eventually withdraw from online dating after repeated experiences of sexual abuse and intrusive messages.

In India, too, concerns about the safety of women and gender minorities on dating platforms are increasingly coming into focus. Women have reported experiencing cyberstalking, doxxing, gendered hate speech, and the unsolicited sharing of explicit images on dating platforms.[105] Users have expressed concerns regarding the practice of 'catfishing',[106] and being misled into engaging in casual encounters while intending to pursue a serious relationship on the platform.[107] Women participants in a study reported that men who faced rejection on dating platforms often tracked them down on other social media platforms like Facebook, and harassed them with

repeated friend requests.[108] Reports of sexual harassment and violence during in-person dates arranged through dating platforms have also come to light.[109] It is likely that many incidents of online and physical sexual harassment go unreported. Studies have reported that women in India often choose not to report instances of sexual violence.[110] This is exacerbated by social taboos against dating in Indian society, which can further preclude women and gender minorities from reporting sexual violence, specifically stemming from interactions on dating platforms, to law enforcement.[111] Women often self-censor their online behaviour, and may eventually withdraw from online dating after repeated experiences of sexual abuse and intrusive messages.[112]

While gender-based online harms stem from broader social contexts, it is important to investigate how the design and operation of dating platforms are often inadequate to both prevent and respond to safety incidents. Many platforms provide safety guides, and some, more than others, place the onus of safety on the users.[113] Some platforms have started crisis text lines for users in select jurisdictions,[114] and partnered with counselling and support organisations.[115] However, these are not uniformly available across all platforms and jurisdictions. Further, reporting mechanisms and content moderation practices of platforms have proven to be insufficient to address safety concerns. Platforms need to streamline reporting mechanisms, reduce response time, and provide more transparency to users on the action taken in response to user complaints.[116] Further platforms must be more transparent on how they design their community guidelines and safety guides, and create spaces for more consultation and feedback from users and civil society.

Case Study: Reporting Mechanisms on Aisle and Quack Quack

Quack Quack allows users to report a profile or a conversation on multiple grounds, including inappropriate profile photos or profile content, indecent behaviour via personalised messages, false information (fake age, profile), using multiple accounts or scamming and spamming. Apart from this, Quack

> "Platforms must be more transparent on how they design their community guidelines and safety guides, and create spaces for more consultation and feedback from users and civil society.

Quack also provides a blank form for users to submit a ticket on any concern/issue/feedback that directs users to its website.[117]

Aisle allows users to report a profile on grounds including "Not interested", "inappropriate photos", "inappropriate messages", "feels like spam" or "others." Aisle also provides an in-app support option, which grants relatively more flexibility in reporting. Here, users can report a safety concern under impersonation, harassment or hacked accounts. Users must enter their registered phone number and other personal details, and have an option to provide relevant information in a textbox and upload attachments. Neither platforms provide users an option to report under categories like "sexual harassment/ assault", "violence", "child abuse", "hate speech", "identity-based abuse", "stalking", etc.[118]

It does provide the residual "other" category where users can type in a textbox. However, this reporting interface can discourage reporting of sexual harassment, physical abuse or racist, religious, casteist or ethnic abuse, either on the platform or during their offline meetings. This reporting interface falls short of how other dating platforms provide reporting options,[119] and the platforms' own Code of Conduct and Community Guidelines, which prohibit users from threatening or harassing others users, or disseminating material that is discriminatory on the basis of race[120] or religion.[121] The reporting mechanisms also fall short of the grievance redressal mechanisms mandated under the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 (Intermediary Guidelines).[122] As per the intermediary guidelines, all intermediaries need to implement a grievance redressal mechanism for users to report NCII[123] (Non-consensual intimate image abuse)[124] and "take all reasonable and practicable measures" to remove/disable access to such content within 24 hours.[125] Both Quack Quack and Aisle do not provide separate reporting mechanisms or an option within the common reporting mechanisms to specifically report NCII content[126] and ensure expedited redressal of these complaints.[127]

"

Dating platforms increasingly rely on automated content moderation, which often fall short in interpreting context-heavy speech, particularly in "low-resourced" languages.

The Platform Question: Power, Accountability and Global South

The inadequacy of content moderation practices of dating platforms in accounting for safety concerns is even more pertinent for multicultural jurisdictions like India. This is due to multiple axes of oppression that must be considered when annotating, identifying or actioning harmful content, including caste, religion and gender.[128] For instance, male gig workers recruited for content moderation may not infer harmful comments that Indian women may face online.[129] Dating platforms increasingly rely on automated content moderation,[130] which often fall short in interpreting context-heavy speech, particularly in "low-resourced" languages.[131] [132] Moreover, proactive moderation tools like Bumble's 'Are You Sure?' and Tinder's 'Does This Bother You?'[133] raise concerns of social surveillance and over-policing of sexual behaviour.[134] Such features are likely to significantly impact users' sexual agency in countries like India, where any form of sexual expression outside heteronormative patriarchal morality is strictly policed and sanctioned. Thus, it is important to recognise that the design of safety features, including proactive content moderation, user verification, and collaboration with third parties to detect harm, can also create surveillance and introduce additional safety risks for marginalised communities.[135]

Towards Meaningful Platform Accountability

As global and Indian dating platforms expand their presence, it is important to institute meaningful accountability. Platform design and operation are often divorced from the lived realities of users and inadvertently reinforce heteronormative patriarchal morality and strict adherence to caste endogamy. It is thus important that global dating platforms hire and consult with a diverse set of experts with experience rooted in the local context, while Indian platforms should structure hiring policies to ensure that marginalised castes and minority communities are adequately represented in their leadership and trust and safety teams.

Transparency in content moderation

In spite of repeated calls for more transparency and accountability on platforms' safety systems, including complaint-handling and content moderation, there has been very little progress.[136]

"

Platform design and operation are often divorced from the lived realities of users and inadvertently reinforce heteronormative patriarchal morality and strict adherence to caste endogamy.

Disclosure of Community Guidelines

Major dating platforms publish community guidelines to inform users what speech and behaviour is impermissible on their platforms.[137] However, these are often buried inside long and incomprehensible Terms of Service documents (see Aisle). Platforms should provide these guidelines in an accessible and understandable format, with an option for users to get more detailed information through examples.[138] These must be available in regional languages. The Intermediary Guidelines, 2021 and the Digital Personal Data Protection Act, 2023 mandate that platforms make their policies available in the preferred languages of their users (English or any language in the Eighth Schedule of the Constitution).[139] We found that neither Aisle nor Quack Quack provided their Community Guidelines, Codes of Conduct or Safety tips in local languages at the time of writing. It is also important that platforms disclose information on how they enforce these community guidelines to ensure user safety, including processes followed internally on receiving user complaints, the use of automated tools to flag certain categories of harmful content, and the linguistic expertise of human content moderators.

> Dating platforms must provide easily accessible reporting mechanisms for users[140] who want to report harmful content or behaviour by other users they met through the platform, either on the platform or outside of it.

Accessible Reporting Mechanisms for users

Dating platforms must provide easily accessible reporting mechanisms for users[140] who want to report harmful content or behaviour by other users they met through the platform, either on the platform or outside of it. They must inform complainants of the progress in their report and the action taken. In case they decide not to act, they must provide the complainant with information on the grounds on which such a decision was taken.

It must be noted that, as per the Intermediary Guidelines, all intermediaries, including dating platforms, must "prominently publish" the name and contact details of the Grievance Officer.[141] We could not locate this information on both Aisle and Quack Quack.

Periodic Transparency Reporting

Dating Platforms have been the worst offenders in terms of releasing baseline aggregate information on content moderation publicly.[142] While social media platforms and even some ride-hailing platforms (in limited jurisdictions) have released aggregate transparency reports, there is no public information on the number of complaints received by dating platforms, even for serious crimes like sexual harassment and rapes.[143] In February 2025, Bumble released its first-ever transparency report for the European Union[144] under Digital Services Act obligations.[145] Prominent dating platforms from Match Group (like Hinge, Tinder), and Bumble started sharing redacted versions of their statement of reasons for adverse action on user accounts or content with the publicly available DSA Transparency Database.[146] Australia's Code of Practice for dating services[147] lays down annual transparency reporting obligations for dating platforms and the first transparency reports will be available in the coming months.[148]

It is, however, unlikely that global platforms will voluntarily extend transparency reporting to other jurisdictions in the Global South, thus periodic reporting of platform's content moderation actions disaggregated by local languages must be prescribed through legislation. Efficiency of automated tools in different contexts and languages, as well as the language proficiency, qualifications and diversity in the human moderation team, must be disclosed. It is also important that transparency reports include aggregate information on state requests for user data.

> While social media platforms and even some ride-hailing platforms (in limited jurisdictions) have released aggregate transparency reports, there is no public information on the number of complaints received by dating platforms, even for serious crimes like sexual harassment and rapes.

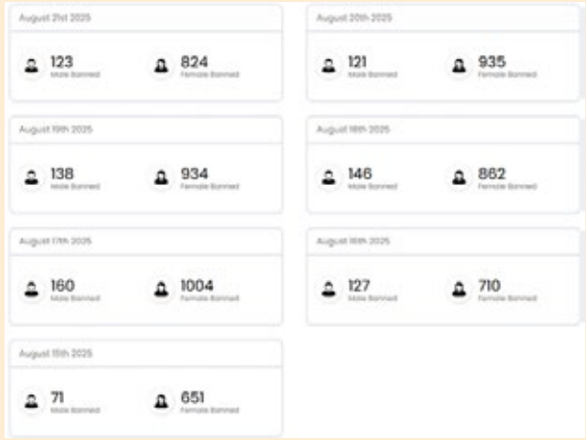


Fig 2: Screenshot of Banned Users on Quack Quack <https://www.quackquack.in/securitytips/>

Quack Quack attempts to provide some aggregate statistics on the number of accounts blocked in the past week based on their code of conduct violations, categorised by gender. However, such information disclosure is not meaningful because it does not categorise account bans on the grounds of violation (for instance, impersonation or hate speech). It also does not provide information on whether actions were taken pursuant to user reports or through self-moderation, including automated means. It is thus difficult to explain why female accounts are blocked around 5 to 10 times more often than male accounts, despite men significantly outnumbering women on dating platforms in India, including Quack Quack.[149] These numbers could be interpreted in many ways: automated moderation tools disproportionately targeting female users, a high prevalence of catfishing accounts reported by male users or even the content moderation prioritising certain harms/reports ***(like fake accounts) over other harms like sexual harassment or hate speech. All in all, no conclusions can be drawn from this limited information, and hence, it is necessary to disclose detailed, meaningful data, even from the platform's own perspective of gaining user trust.)***

> Many users in India are registering on a dating platform for the first time, and it would be useful if this basic information is provided in local languages in an accessible and comprehensible manner.

Transparency in matching algorithms

Often, users consider profile recommendations to be neutral and objective; it is thus important, at the very least, to disclose to users basic qualitative information on the matching algorithms deployed by the dating platform and the kind of user data and behavioural signals that feed into the system. Many users in India are registering on a dating platform for the first time, and it would be useful if this basic information is provided in local languages in an accessible and comprehensible manner. However, there is a need for meaningful algorithmic accountability and user empowerment beyond baseline disclosures. Algorithms are complex socio-technical assemblages that are co-constructed by code, training data, and user inputs and interaction.[150] This means that causal explanations for matching recommendations cannot be deciphered conclusively, even by developers and experts.151 One way dating platforms can make algorithmic accountability more meaningful is through collaboration with independent researchers.[152] Another way dating platforms could potentially understand the impact of matching algorithms on diverse users is through submission to independent third-party audits.[153]

These audits would be conducted by expert groups working with victims of sexual harassment and violence, especially those belonging to LGBTQIA+, Dalit, Bahujan, Adivasi and religious minority communities. Platforms and academic, civil society and technical researchers must find ways to conduct research and audits without compromising the privacy of users and the security of the data. It is also important that both civil society and dating platforms reflect on who gets access to funding and data to conduct audits and independent studies. Civil society groups in tech policy must reflect on the lack of diversity and prioritise the voice of marginalised communities. Accountability measures must go beyond performative transparency, and users should be empowered to gain more control over the dating choices they make in a safe, secure and inclusive space. It is important that those who experience the most egregious harms from dating platforms lead the way in holding them accountable and designing alternative models to current platform design and economics.

> Accountability measures must go beyond performative transparency, and users should be empowered to gain more control over the dating choices they make in a safe, secure and inclusive space.

Endnotes:

1. Benson Rajan, 'Harassment and abuse of Indian women on dating apps: a narrative review of literature on technology-facilitated violence against women and dating app use' (2025) 12 Humanities and Social Sciences Communications 55 <https://doi.org/10.1057/s41599-024-04286-6> accessed 11 October 2025; Sunaina Arya and Aakash Singh Rathore, 'Introduction: Theorising Dalit Feminism', Dalit feminist theory (Routledge India 2019).

2. See 'Tinder Launches India Operations, Appoints Taru Kapoor as India Head' The Times of India (6 January 2016). https://timesofindia.indiatimes.com/technews/tinder-launches-india-operations-appoints-taru-kapoor-as-india-head/articleshow/50465904.cms> accessed 22 August 2025; Lata Jha, 'Bumble Heats up India's Dating Scene' (mint, 6 December 2018) <https://www.livemint.com/Consumer/IKWgIRTlYX4QIPqnomWShI/BumbleheatsupIndiasdatingscene.html> accessed 22 August 2025.

3. According to Grand View Research, India's dating industry generated $547.9 million in revenue in 2023 and is expected to grow to $1 billion by 2030. See Rwit Ghosh, 'Where the Heart Is: India's Dating Apps Find Love Outside Metros' (mint, 7 August 2025) <https://www.livemint.com/companies/start-ups/indian-dating-apps-dating-apps-in-india-best-dating-apps-for-tier-2-cities-quackquack-dating-app-aisle-dating-app-11754288856407.html> accessed 26 August 2025; Uma Kannan, 'Gen-z: Dating Apps Swipe Right' (The New Indian Express, 14 May 2023) <https://www.newindianexpress.com/business/2023/May/13/gen-z-dating-apps-swipe-right-2574864.html> accessed 26 August 2025.

4. Ankita Deshkar, 'Love, Lies, and Loss: How Scammers on Dating Apps Lure Lonely Hearts' The Indian Express (2 May 2025) <https://indianexpress.com/article/technology/tech-news-technology/love-lies-and-loss-how-scammers-on-dating-apps-lure-lonely-hearts-9747168/> accessed 26 August 2025.

5. T. N. M. Staff, 'Dalit Journalist Subjected to Horrific Harassment for Writing about Casteism in Dating Apps' The News Minute (21 May 2024) <https://www.thenewsminute.com/news/dalit-journalist-subjected-to-horrific-harassment-for-writing-about-casteism-in-dating-apps> accessed 26 August 2025.

6. Nandita Singh, Simrin Sirur, 'Indian Women Are Getting Assaulted on Tinder Dates and No One Knows How to Stop It' ThePrint (18 November 2018) <https://theprint.in/feature/indian-women-are-getting-assaulted-on-tinder-dates-and-no-one-knows-how-to-stop-it/151022/> accessed 26 August 2025.

7. Rajan (n 1).

8. Although many of the harms discussed in this essay may be experienced by users of other platforms, like peer-to-peer messaging, the focus of this essay is online dating platforms.

9. Dia Rekhi, 'Desi Dating Apps Go All Out to Court Users in Small Cities - The Economic Times' ET Prime (11 February 2023) <https://economictimes.indiatimes.com/tech/technology/dating-apps-go-desi-woo-townsfolk/articleshow/97807114.cms?from=mdr> accessed 3 September 2025.

10. Aisle is a "high-intent dating app for Indians." It was launched in 2014 (acquired by InfoEdge in 2022) and was positioned as a dating platform for long-term relationships leading to marriage. It positioned itself between matrimonial apps like shaadi.com and casual dating apps like Tinder. It has launched several vernacular dating platforms under its brand, including Arike, Anbe, Neetho and Neene. However, for the purpose of this essay, we focus exclusively on the flagship platform Aisle. According to a report, Aisle has over a million monthly active users (with 50% of the total users on its flagship platform Aisle), and the company claims an overall user base of 16 million. See Pooja Yadav, 'The Info Edge Effect: How Dating App Aisle's Revenue Soared 146% After Acquisition' (Inc42, 10 April 2025) <https://inc42.com/startups/the-info-edge-effect-how-dating-app-aisles-revenue-soared-146-after-acquisition/> accessed 27 August 2025.

11. Quack Quack is a dating platform for singles in India that allows users to match anywhere in India (across cities), enables chatting without matching and offers the services of a matchmaker on a premium subscription. The platform has been gaining traction beyond metro cities and claims to have over 35 million active users. See DC Correspondent, 'QuackQuack Hits 35 Million Users, Unveils Key Dating Trends' Deccan Chronicle (18 January 2025) <https://www.deccanchronicle.com/lifestyle/relationship/quackquack-reaches-35-million-users-revealing-key-trends-in-indias-dating-scene-1854903> accessed 27 August 2025; Ghosh (n 3).

12. Jen Caltrider, Misha Rykov and Zoë MacDonald, 'Data-Hungry Dating Apps Are Worse Than Ever for Your Privacy' (Mozilla Foundation, 23 April 2024) <https://foundation.mozilla.org/en/privacynotincluded/articles/data-hungry-dating-apps-are-worse-than-ever-for-your-privacy/> accessed 7 April 2025.

13. ibid.
14. ibid.
15. ibid.
16. ibid.

17. Bobby Allyn, 'Study: Tinder, Grindr And Other Apps Share Sensitive Personal Data With Advertisers' NPR (14 January 2020) <https://www.npr.org/2020/01/14/796427696/study-grindr-tindr-and-other-apps-share-sensitive-personal-data-with-advertisers> accessed 20 August 2025.

18. When a 2018 study by Norwegian nonprofit research group Sintef uncovered Grindr's data sharing with two companies Apptimize and Localytics, the dating app's privacy policy did not explicitly stipulate third-party data sharing. Instead, the privacy policy stated, "...if you choose to include information in your profile,

and make your profile public, that information will also become public." See Azeen Ghorayshi and Sri Ray, 'Grindr Is Letting Other Companies See User HIV Status And Location Data' (3 April 2018) <https://www.buzzfeednews.com/article/azeenghorayshi/grindr-hiv-status-privacy#.yp0J48W0N> accessed 20 August 2025; 'SINTEF-9012/Grindr-Privacy-Leaks' <https://github.com/SINTEF-9012/grindr-privacy-leaks> accessed 20 August 2025.

19. ibid.

20. See Jonathan A Obar and Anne Oeldorf-Hirsch, 'The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services' (2020) 23 Information, Communication & Society 128 <https://www.tandfonline.com/doi/full/10.1080/1369118X.2018.1486870> accessed 1 September 2025; Rishab Bailey and others, 'Disclosures in Privacy Policies: Does "Notice and Consent" Work?' (Social Science Research Network, 11 December 2018) <https://papers.ssrn.com/abstract=3328289> accessed 1 September 2025.

21. This includes, for instance, information pertaining to sexual orientation, political opinions, religious or philosophical beliefs, and is recognised under Article 9 of the GDPR. Consent under GDPR must be "free, specific, informed and unambiguous." Explicit consent presents a higher threshold and as per the European Data Protection Board (EDPB) guidelines, the term explicit refers to the manner in which the data subject grants consent as an "express statement of consent." See Guidelines (EDPB) 05/2020 on Consent Under Regulation 2016/679 [2020], para 93.

22. 'Digital Personal Data Protection Bill Gets President's Assent' The Economic Times (12 August 2023) <https://economictimes.indiatimes.com/news/india/digital-personal-data-protection-bill-gets-nod-from-president/articleshow/102660125.cms?from=mdr>.

23. The Digital Personal Data Protection Act 2023 <https://www.meity.gov.in/static/uploads/2024/06/2bf1f0e9f04e6fb4f8fef35e82c42aa5.pdf>.

24. See Siddharth Sonkar, 'How Dating Apps Exploit India's Loosely Formed Definition of "Personal Information"' (ThePrint, 27 March 2022) <https://theprint.in/pageturner/excerpt/how-dating-apps-exploit-indias-loosely-formed-definition-of-personal-information/889370/> accessed 27 August 2025.

25. Although there are legitimate criticisms highlighting the limitation of the sensitive personal data approach in privacy law, recognising additional safeguards for such personal data can perhaps ensure a higher degree of user awareness, especially when no alternative frameworks exist. See Daniel J Solove, 'Data Is What Data Does: Regulating Use, Harm, and Risk Instead of Sensitive Data' [2023] SSRN Electronic Journal <https://www.ssrn.com/abstract=4322198> accessed 1 September 2025; Centre for Communication Governance, 'Comments on the Digital Personal Data Protection Bill, 2022'(2022) <https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/ccg-nlu-comments-to-meity-on-the-draft-digital-personal-data-protection-bill-2022-334.pdf>

26. Centre for Communication Governance, 'Comments on the Draft Digital Personal Data Protection Rules' (2025) <https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/ccg-nlud-comments-on-the-digital-personal-data-protection-rules-2025-1-742.pdf); CCG, 'UNDP Guide- Drafting Data Protection Legislation: A study of regional frameworks' (UNDP 2023), p 46. <https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/undp-drafting-data-protection-legislation-march-2023-443.pdf>.

27. For instance, Tinder recently rolled out a game where users can interact with AI personas and receive scores and feedback to improve their dating skills. See Lauren Forristal, 'Tinder's New AI-Powered Game Assesses Your Flirting Skills' (TechCrunch, 1 April 2025) <https://techcrunch.com/2025/04/01/tinders-new-ai-powered-game-assesses-your-flirting-skills/> accessed 1 September 2025. Grindr plans to launch an AI wingman. See Reece Rogers, 'I Took Grindr's AI Wingman for a Spin. Here's a Glimpse of Your Dating Future' Wired <https://www.wired.com/story/hands-on-with-grindr-ai-wingman/> accessed 1 September 2025.

28. See Paige Collings, 'Dating Apps Need to Learn How Consent Works' (Electronic Frontier Foundation, 21 July 2025) <https://www.eff.org/deeplinks/2025/07/dating-apps-need-learn-how-consent-works> accessed 27 August 2025; Lauren Forristal, 'Hinge's New AI Feature Determines If Your Prompt Response Is Too Basic' (TechCrunch, 15 January 2025) <https://techcrunch.com/2025/01/15/hinge-new-ai-feature-determines-if-your-prompt-response-is-too-basic/> accessed 27 August 2025; Xavier Harding, 'Dating Apps And Your User Privacy — What To Keep In Mind' (Mozilla Foundation, 16 May 2024) <https://www.mozillafoundation.org/en/blog/dating-app-user-privacy/> accessed 10 August 2025.

29. Collings (n 28); 'Bumble's AI Icebreakers Are Mainly Breaking EU Law' (noyb, 26 June 2025) <https://noyb.eu/en/bumbles-ai-icebreakers-are-mainly-breaking-eu-law> accessed 27 August 2025.

30. Alison Frankel, 'Pssst! Match.Com Does Not Want You to Know about This FTC Case' Reuters (6 July 2022) <https://www.reuters.com/legal/litigation/pssst-matchcom-does-not-want-you-know-about-this-ftc-case-2022-07-06/> accessed 27 August 2025.

31. Joe Tidy, 'Kink and LGBT Dating Apps Exposed 1.5m Private User Images Online' (BBC News, 30 March 2025) <https://www.bbc.com/news/articles/c05m5m5v327o> accessed 7 April 2025.

32. The study reported that this vulnerability has now been addressed after it was reported to OkCupid.

The Platform Question: Power, Accountability and Global South

33. Edvardas Mikalauskas, 'Popular Dating App Leak Puts Millions of Women at Risk' (Cybernews, 6 March 2020) <https://cybernews.com/security/popular-dating-app-leak-puts-millions-of-women-at-risk/> accessed 17 August 2025.

34. ibid.

35. The police used messages expressing "I like you" as evidence of queerness. See Matt Burgess, 'How Police Abuse Phone Data to Persecute LGBTQ People' Wired (7 March 2022) <https://www.wired.com/story/lgbtq-phone-data-police/> accessed 11 October 2025.

36. Krishnadas Rajagopal, 'SC Decriminalises Homosexuality, Says History Owes LGBTQ Community an Apology' The Hindu (6 September 2018) <https://www.thehindu.com/news/national/sc-de-criminalises-homosexuality-says-history-owes-lgbtq-community-an-apology/article61535787.ece> accessed 3 September 2025.

37. 'Dating App Scam Targets Queer Men' The Times of India (Pune, 2 March 2025) <https://timesofindia.indiatimes.com/city/pune/dating-app-scam-targets-queer-men/articleshow/118666385.cms> accessed 3 September 2025.

38. Ankita Garg, 'What Is Bulli Bai App, What Is Its Link to Sulli Deals, and How GitHub Is Involved: Story in 10 Points' India Today (10 January 2022) <https://www.indiatoday.in/technology/features/story/what-is-bulli-bai-app-what-is-its-link-to-sulli-deals-and-how-github-is-involved-story-in-10-points-1898365-2022-01-10> accessed 27 August 2025.

39. Quratulain Rehbar and Pallavi Pundir, 'Muslim Women Were "Auctioned Like Cattle" on a Hate Site' [2021] VICE <https://www.vice.com/en/article/muslim-women-islamophobia-hate-app-india/> accessed 7 August 2025.

40. Arbab Ali and Nadeem Sarwar, 'Muzz, the World's Largest Muslim Dating App, Is Struggling in India' (Rest of World, 24 April 2023) <https://restofworld.org/2023/muzz-dating-app-muzmatch-growth-slows/> accessed 27 August 2025.

41. It must be noted that both Aisle and Quack Quack will have to alter their privacy policies as the Digital Personal Data Protection Act, 2023 comes into force, including ensuring free, specific, informed, unconditional and unambiguous consent from the data principal. They will have to establish mechanisms enabling ease of withdrawing consent for processing of personal data, deletion and erasure rights to data principals, and notification protocols on breach of personal data.

42. Aisle collects sensitive personal information, including "interests, philosophy, age, height, religion, ethnicity, place of residence." See Privacy Policy <https://app.aisle.co/mobile/privacy_policy> accessed 11 October 2025.

43. Quack Quack's Privacy Policy notes that "some of the information you choose to provide us may be considered 'special' or 'sensitive' in certain jurisdictions, for example your racial or ethnic origins, sexual orientation and religious beliefs." It also collects,"information on device sensors such as accelerometers, gyroscopes and compasses." See QuackQuack.in Privacy Policy <https://www.quackquack.in/privacypolicy/> accessed 11 October 2025.

44. Aisle takes user consent at the time of registration to process profile information, partner preferences, photographs, work and interests, geolocation, among others. However, it doesn't ask for explicit informed consent for processing specific information (other than geolocation) after registration. It appears that users have to mandatorily grant consent to processing all such information (including sensitive personal information) to be able to create an account on the platform.

45. Quack Quack's Privacy Policy notes that "Some of the information you choose to provide us may be considered 'special' or 'sensitive' in certain jurisdictions... by choosing to provide this information, you consent to our processing of that information." It neither asks for explicit user content to process sensitive personal information, nor limits the purpose of such processing.

46. It is worth noting that once you grant Quack Quack permission to collect geolocation information, it may continue to collect the information in the background even when you are not using these services (through various means including GPS, Bluetooth and Wifi Connections).

47. The Privacy Policy states that, "you have the right to withdraw your consent at any time for processing your location." This can presumably be done through device access controls. However, withdrawing consent will hamper the platform's functionality as sharing geolocation information is "necessary to see which other users of Aisle are nearby at any given moment." Aisle's Privacy Policy does not elaborate on the process to withdraw consent for processing of other kinds of personal or sensitive personal information.

48. Quack Quack's Privacy Policy mentions three legal bases for processing user data including, "providing services to you", "legitimate interest" and "consent." It appears to cover most data processing under the first two categories, and does not clarify what kind of user data is only processed and collected upon user consent. It is also not clear whether users can withdraw consent for processing of specific pieces of information that the platform may consider necessary to provide services. It just states, "From time to time, we may ask for your consent to use your information for certain specific reasons. You may withdraw your consent at any time by contacting us at the address provided at the end of this Privacy Policy." Further, like Aisle, Quack Quack's privacy policy states that users can deny permission to collect user geolocation (presumably through device access controls).

49. Aisle's Privacy Policy explicitly states, "We do not sell your Personal Data to third parties."

The Platform Question: Power, Accountability and Global South

50. Quack Quack's Privacy Policy states that it may share non-personal information as well as personal information (in hashed, non-human readable format) with "other group products and third parties (notably advertisers) to develop and deliver targeted advertising on our services and on websites or applications of third parties, and to analyze and report on advertising you see."

51. Aisle, like many other platforms, does not provide users granular control to determine how their content is shared and processed by third parties. For instance, the Privacy Policy states that when users upload their pictures, Aisle "may use an external face recognition tool and other techniques to select the best pictures for Your Account." It does not provide more information on these external tools and it appears from both the Privacy Policy and the user interface of the platform, that there is no option for users to deny consent for such processing.

52. Quack Quack shares user information with service providers and partners, other group products and legal authorities. It does not grant users an option to opt out of such data sharing except under very limited (and unclear) circumstances. For instance, as per the Privacy Policy, when sharing user information with group products for the purpose of cross functionality or visibility on new services developed by the group, "we will of course comply with applicable law and, where relevant, notify you of any such opportunity and allow you to agree or to refuse."

53. As per the Privacy Policy, data is retained as long as strictly necessary or allowed by law (whichever is shorter). Personal data may be erased if the user has not logged in for more than five years. On account deletion, personal data is deleted after a three-year safety retention window.

54. The platform retains "personal information only as long as we need it for legitimate business purposes and as permitted by applicable law." It does not specify the duration of such retention.

55. As per the platform's Privacy Policy, users can contact their support team to delete data permanently from their records. It is unclear whether data is still retained for "safety retention window of upto three years following account deletion."

56. On Quack Quack, users can only deactivate their account from the user interface and they need to contact the platform to Suspend/Delete their account. Even on deletion, a user's profile information is only removed from other users' view and their information may continue to appear in certain search results. The platform will also continue to retain the information for "record keeping." It is unclear how users can permanently delete their information from the platform.

57. The Exclusivity feature tracks and allows premium Aisle members to find out how many people their match interacted with in the last 3 days (without the knowledge of the latter). This is categorised into 'fewer than 5', 'more than 5', or 10, or 20 interactions.

58. See Kenneth James Lay, 'Sexual Racism: A Legacy of Slavery' (1993) 13 National Black Law Journal 165 <https://heinonline.org/HOL/Page?handle=hein.journals/natblj13&id=169&div=&collection=>; Manoj Mitta, 'The Outrage of Marrying Up', Caste pride: Battles for equality in Hindu India (Context 2023).

59. Josue Ortega and Philipp Hergovich, 'The Strength of Absent Ties: Social Integration via Online Dating' (arXiv, 14 September 2018) <http://arxiv.org/abs/1709.10478> accessed 10 August 2025.

60. Apryl Williams, Not My Type: Automating Sexual Racism in Online Dating (Stanford University Press 2024).

61. See Mitchell and Wells for the argument that while it might generally be justifiable to date or not date someone based on attractiveness, maintaining racially exclusive dating pools is not morally defensible. Megan Mitchell and Mark Wells, 'Race, Romantic Attraction, and Dating' (2018) 21 Ethical Theory and Moral Practice 945, <http://link.springer.com/10.1007/s10677-018-9936-0> accessed 17 August 2025.

62. Williams defines sexual racism as, "personal racialised reasoning in sexual, intimate, and or romantic partner choice of interest." See Williams (n 60).

63. A study highlights how a majority of male respondents considered racism on dating platforms as problematic but 70% did not consider indicating racial preferences online as a form of racism. See Denton Callander, Christy E Newman and Martin Holt, 'Is Sexual Racism Really Racism? Distinguishing Attitudes Toward Sexual Racism and Generic Racism Among Gay and Bisexual Men' (2015) 44 Archives of Sexual Behavior 1991 <http://link.springer.com/10.1007/s10508-015-0487-3> accessed 17 August 2025.

64. In 2014, OKCupid released data on how men and women across different groups rate attractiveness. It found that Asian, Latino and White men rate Black women as 20%, 18%, and 17% less attractive than other women on their platforms, respectively. By contrast, white men were rated as 18%, 12% and 19% more attractive by Asian, Latina and White women as compared to other men on the platforms. See Blog <http://blog.okcupid.com/index.php/race-attraction-2009-2014/> in Williams (n 60).

65. Williams (n 60). Also see Wei-Chin Hwang, 'Who Are People Willing to Date? Ethnic and Gender Patterns in Online Dating' (2013) 5 Race and Social Problems 28 <https://doi.org/10.1007/s12552-012-9082-6> accessed 17 August 2025; Gina Potârcă and Melinda Mills, 'Racial Preferences in Online Dating across European Countries' (2015) 31 European Sociological Review 326 <https://academic.oup.com/esr/article-lookup/doi/10.1093/esr/jcu093> accessed 17 August 2025.

The Platform Question: Power, Accountability and Global South

66. For instance, Carlson highlights the experiences of Indigenous Australians in navigating online dating, where they routinely encountered racial abuse and violence. See Bronwyn Carlson, 'Love and Hate at the Cultural Interface: Indigenous Australians and Dating Apps' (2020) 56 Journal of Sociology 133 <https://journals.sagepub.com/doi/10.1177/1440783319833181> accessed 17 August 2025.

67. See, for instance, Chris Stokel-Walker, 'Why Is It OK for Online Daters to Block Whole Ethnic Groups?' The Observer (29 September 2018) <https://www.theguardian.com/technology/2018/sep/29/wltm-colour-blind-dating-app-racial-discrimination-grindr-tinder-algorithm-racism> accessed 10 July 2024; 'Grindr Removes "Ethnicity Filter" after Complaints' (1 June 2020) <https://www.bbc.com/news/technology-52886167> accessed 18 August 2025; Amy Thomson, Olivia Carville and Nate Lanxon, 'Match Opts to Keep Race Filter for Dating as Other Sites Drop It' Bloomberg.com (8 June 2020) <https://www.bloomberg.com/news/articles/2020-06-08/dating-apps-debate-race-filters-as-empowering-or-discriminating> accessed 18 August 2025.

68. Collaborative Filtering is used for personalised recommendations where users are recommended items (in this case, potential partners) based on the opinion or behaviour of a community of users who share similar preferences. See J Ben Schafer and others, 'Collaborative Filtering Recommender Systems' in Peter Brusilovsky, Alfred Kobsa and Wolfgang Nejdl (eds), The Adaptive Web: Methods and Strategies of Web Personalization (Springer 2007) <https://doi.org/10.1007/978-3-540-72079-9_9> .

69. Liesel L Sharabi, 'Finding Love on a First Data: Matching Algorithms in Online Dating' [2022] Harvard Data Science Review <https://hdsr.mitpress.mit.edu/pub/i4eb4e8b> accessed 31 July 2025; Karim Nader, 'Dating through the Filters ' (2020) 37 Social Philosophy and Policy 237 <https://www.cambridge.org/core/product/identifier/S0265052521000133/type/journal_article> accessed 7 August 2025; Williams (n 60); 'Dating App Algorithms.' (MonsterMatch) <https://monstermatch.hiddenswitch.com/algorithms> accessed 18 August 2025.

70. Nader (n 69).

71. B.R. Ambedkar, 'Castes in India: Their Mechanism, Genesis and Development' (1917).

72. A study examining inter-caste marriages from 1951-2012, using data from the India Human Development Survey 2011-12 (IHDS) notes an inconsequential rise in inter-caste marriage during the time period, with only 4.5% of the women surveyed in the sample having married outside their caste during 2011-12. See Pralip Kumar Narzary and Laishram Ladusingh, 'Discovering the Saga of Inter-Caste Marriage in India' (2019) 54 Journal of Asian and African Studies 588 <https://doi.org/10.1177/0021909619829896> accessed 22 April 2025.

73. A study examining the data from the National Family Health Survey (NHFS III 2005-2006), found that only 2.1% of marriages were interreligious. See Kumudin Das and others, 'Dynamics of Inter-Religious and Inter-Caste Marriages in India' [2011] Population Association of America, Washington DC, USA ; As per a survey by Pew Research Centre, nearly all (99%) of respondents stated that they shared the same religion as their spouse. See Pew Research Center, 'Religion in India: Tolerance and Segregation' (June 29, 2021) <https://www.pewresearch.org/wp-content/uploads/sites/20/2021/06/PF_06.29.21_India.full_.report.pdf>.

74. See Shruti Tomar, 'Dalit Man Beaten to Death for Inter-Caste Marriage in Madhya Pradesh Village' (Hindustan Times, 25 August 2025) <https://www.hindustantimes.com/cities/noida-news/dalit-man-beaten-to-death-for-inter-caste-marriage-in-madhya-pradesh-village-101756125601181.html>; The Hindu Bureau, 'Tirunelveli "Honour" Killing Case: Victim Kavin's Mother, Teachers Express Shock' The Hindu (28 July 2025) <https://www.thehindu.com/news/national/tamil-nadu/sc-man-murder-in-tirunelveli-activists-demand-law-to-prevent-honour-killing/article69864758.ece> accessed 11 October 2025; Hannah Ellis-Petersen and Ahmer Khan, '"They Cut Him into Pieces": India's "Love Jihad" Conspiracy Theory Turns Lethal' The Guardian (21 January 2022) <https://www.theguardian.com/world/2022/jan/21/they-cut-him-into-pieces-indias-love-jihad-conspiracy-theory-turns-lethal> accessed 11 October 2025.

75. See for instance, Charu Gupta, Sexuality, Obscenity, Community: Women, Muslims, and the Hindu Public in Colonial India (Orient Blackswan 2001); David James Strohl, 'Love Jihad in India's Moral Imaginaries: Religion, Kinship, and Citizenship in Late Liberalism' (2019) 27 Contemporary South Asia 27; Shewli Kumar and Iswarya Subbiah, 'Crimes in the Name of Honour: A National Shame' (DHRDNet 2022) <https://www.dhrdnet.org/honour-crimes-research-report/>.

76. See for instance, Nitya Kuthiala and Keegan McBride, 'How Silicon Valley Developed Digital Dating Platforms Are Transforming Love and Relationship Culture in India' (Social Science Research Network, 15 January 2025) <https://papers.ssrn.com/abstract=5097768> accessed 22 January 2025.

77. Kuthiala and McBride (n 76); Kavita Dattani, 'Data-Bility: Endogamous Social Intimacies on Dating Apps in Mumbai' (2025) 50 Transactions of the Institute of British Geographers e12687 <https://onlinelibrary.wiley.com/doi/abs/10.1111/tran.12687> accessed 9 July 2025 ; Benson Rajan, '"It Follows You Home": Emotional and Psychological Impacts of Dating-App Harassment on Indian Women' (2025) 112 Women's Studies International Forum 103129 <https://www.sciencedirect.com/science/article/pii/S0277539525000780> accessed 11 August 2025.

78. The study interviewed 18 participants using Tinder, Bumble or Hinge to understand how Indian users navigate Silicon Valley built digital dating platforms given the conflict between the platform's embedded values and those of their society. See Kuthulia and McBride (n 76).

79. It is thus pertinent to pay attention to Kang's problematisation of intercaste love as a site of power, where intercaste love is seen as a "temporary

The Platform Question: Power, Accountability and Global South

exploration for Savarnas until they settle down with more serious partners." See Akhil Kang, 'Savarna Citations of Desire: Queer Impossibilities of Inter-Caste Love' (2023) 133 Feminist Review 63 <https://doi.org/10.1177/01417789221146514> accessed 8 April 2025.

80. It has been reported that an Australian dating platform for South Asian diaspora called Dil Mil allows filters to match within 'dominant' castes but has no options for 'lower' caste groups. See 'Caste Discrimination Continues Its Journey – Now within the Australian Diaspora' (Dalit Solidarity Network, 11 February 2021) <https://www.dsnuk.org/2021/02/11/caste-discrimination-continues-its-journey-now-within-the-australian-diaspora/> accessed 2 September 2025.

81. See Christina Thomas Dhanaraj, 'Swipe me left, I'm Dalit' (GenderIT.org, 14 April 2018) <https://www.genderit.org/articles/swipe-me-left-im-dalit> ; Deep Mukherjee, 'Looking for Love and Finding Caste on Dating Apps' The Indian Express (28 October 2023) <https://indianexpress.com/article/opinion/columns/dating-apps-love-caste-9003207/> accessed 11 October 2025.

82. Carolina Bandinelli and Alessandro Gandini, 'Dating Apps: The Uncertainty of Marketised Love' (2022) 16 Cultural Sociology 423 <https://doi.org/10.1177/17499755211051559> accessed 12 August 2025.

83. Dhanaraj (n 81).

84. Shannon Philip, 'Grindr Wars: Race, Caste, and Class Inequalities on Dating Apps in India and South Africa' (2024) 26 Ethnoscripts <https://journals.sub.uni-hamburg.de/ethnoscripts/article/view/2329> accessed 2 April 2025.

85. ibid.

86. See Kuthiala and McBride (n 76); Benson Rajan, 'Fearing the "Known Unknown" Men: A Study on "Red Flags" and Safety Work on Dating Apps in India' [2025] Women's Studies in Communication 1.

87. Ravikant Kisana, 'Dating Like a Savarna' (The Swaddle, 29 April 2023) <https://www.theswaddle.com/dating-like-a-savarna> accessed 9 July 2025.

88. Laurent Gayer and Christophe Jaffrelot, 'Conclusion - "In Their Place?" The Trajectories of Marginalisation of India's Urban Muslims' (Columbia University Press ; Hurst Publishers 2012) <https://sciencespo.hal.science/hal-03415405> accessed 2 September 2020.

89. See Kisana (n 87); Anonymous, 'Being Right-Swiped as a Dalit Woman on Dating Apps' (LiveWire 19 July 2019) <https://livewire.thewire.in/livewire/dating-dalit-woman-casteism/>; TNM Staff, 'Dalit Journalist Subjected to Horrific Harassment for Writing about Casteism in Dating Apps' (The News Minute, 21 May 2024) <https://www.thenewsminute.com/news/dalit-journalist-subjected-to-horrific-harassment-for-writing-about-casteism-in-dating-apps> accessed 9 July 2025.

90. Manisha Mondal, 'Being Dalit on a Dating App. Upper Caste Men Only Want to Argue over Reservation, EWS' (ThePrint, 16 May 2024) <https://theprint.in/opinion/pov/being-dalit-on-a-dating-app-upper-caste-men-only-want-to-argue-over-reservation-ews/2087897/> accessed 9 July 2025; Dhanaraj (n 81); Kisana (n 87).

91. Dhanaraj (n 81); Noel Mariam George, 'Beauty, Femininity and the Politics of "Desire"' (12 February 2020) <https://www.roundtableindia.co.in/beauty-femininity-and-the-politics-of-desire/> accessed 17 August 2025.

92. Shailaja Paik, 'Building Bridges: Articulating Dalit and African American Women's Solidarity' (2014) 42 WSQ: Women's Studies Quarterly 74 <https://muse.jhu.edu/article/572234> accessed 12 August 2025.

93. ibid.

94. Kuthiala and McBride (n 76).

95. Dattani (n 77).

96. As per its Code of Conduct, "No racist remark or religious persecution; any direct or indirect remarks will trigger the immediate termination of your account." The list of prohibited content also mentions content that "is abusive, insulting or threatening, discriminatory or that promotes or encourages racism, sexism, hatred or bigotry."

97. Users must choose between Vegetarian, Non-vegetarian or Eggetarian.

98. Dolly Kikon, 'Dirty Food: Racism and Casteism in India' (2022) 45 Ethnic and Racial Studies 278 <https://doi.org/10.1080/01419870.2021.1964558> accessed 29 August 2023.

99. Users must indicate whether they are Non-vegetarian, vegetarian, eggetarian, vegan or pescatarian.

100. Aisle requires users to select from a list containing: Hindu, Spiritual, Muslim, Christian, Atheist, Agnostic, Buddhist, Jewish, Parsi, Sikh, Jain, Bahau or Other.

101. See Yashica Dutt, '"Indian Matchmaking" Exposes the Easy Acceptance of Caste' (The Atlantic, 1 August 2020) <https://www.theatlantic.com/culture/archive/2020/08/netflix-indian-matchmaking-and-the-shadow-of-caste/614863/> accessed 2 September 2025.

102. Match Group owns some of the most popular dating platforms, including Tinder, OkCupid, Hinge, Plenty of Fish, etc.

The Platform Question: Power, Accountability and Global South

103. This followed an eighteen-month investigation by the Pulitzer Center's AI Accountability Network and the Markup. See Emily Elena Dugdale and Hanisha Harjani, 'Rape under Wraps: How Tinder, Hinge and Their Corporate Owner Chose Profits over Safety' The Guardian (13 February 2025) <https://www.theguardian.com/us-news/2025/feb/13/tinder-hinge-match-investigation> accessed 7 August 2025.

104. Salter, M., Tyson, D., Woodlock, D.,'Swipe wrong: How sex offenders target single parents on dating apps to exploit their children' in Searchlight 2025 – Childlight's Flagship Report (Childlight – Global Child Safety Institute 2025) <https://www.childlight.org/searchlight/study-d-swipe-wrong-how-sex-offenders-target-single-parents-on-dating-apps-to-exploit-their-children> accessed 11 October 2025.

105. Singh and Sirur (n 6); Singh Rahul Sunilkumar, 'Majority of Indian Adults Ignorant about Consent in Relationship: Tinder Survey' Hindustan Times (5 September 2022) <https://www.hindustantimes.com/technology/indian-adults-ignorant-about-consent-in-relationship-says-tinder-survey-101662369660962.html> accessed 8 July 2025.

106. Women navigating dating apps in India often express concerns about the authenticity of information men provide, including age, income, and educational background See Lingutla A, Kumar V (2022) Evolution of Online Dating: Analysis of Dating Preferences, User Psychology and Pain Points in Context to Indian Market. Int Res J Modernization Eng Technol Sci 4(11):784–795< https://doi.org/10.56726/IRJMETS31248>.

107. Lingutla and Kumar (n 106).

108. Anita Gurumurthy, Amrita Vasudevan and Nandini Chami, 'Born Digital, Born Free? A Socio-Legal Study on Young Women's Experiences of Online Violence in South India' (IT for Change 2019).

109. See Arvind Ojha, 'Delhi Woman Accuses Bumble Date of Rape, Case Filed' India Today (New Delhi, 26 October 2023) <https://www.indiatoday.in/cities/delhi/story/woman-alleges-rape-by-bumble-date-delhi-police-register-case-2453908-2023-10-26> accessed 2 August 2025; Aakash Ghosh, 'Swipe Right for Trouble: Dating Apps New Playground for Criminals?' Hindustan Times (12 April 2025) <https://www.hindustantimes.com/cities/lucknow-news/swipe-right-for-trouble-dating-apps-new-playground-for-criminals-101744398505973.html> accessed 3 September 2025.

110. Vignesh Radhakrishnan and Rebecca Rose Varghese, 'No sexual violence survivor contacted a lawyer, only 4.7% took police help in 2019-21' The Hindu (16 May 2022) <https://www.thehindu.com/data/data-no-sexual-violence-survivor-contacted-a-lawyer-only-47-took-police-help-in-2019-21/article65419734.ece> accessed 11 October 2025.

111. Rajan (n 1).

112. Rajan (n 1).

113. See, for instance, Aisle's safety section nestled in its Terms of Use, which makes it clear that users bear the "sole responsibility for taking all appropriate safety precautions." The section then outlines user responsibility to not post "any defamatory, inaccurate, abusive, obscene, profane, offensive, sexually oriented, threatening, harassing, racially offensive, or illegal material." The safety section notably does not outline the platform's own safety policies. It does not elaborate on reporting/complaint mechanisms available to the user and the process followed by the platform on receiving such complaints. It also does not provide information on the platform's own content moderation initiatives.

114. See for instance, 'Crisis Text Line' <https://policies.tinder.com/safety-center/tools/crisis-text-line/us/en/#:~:text=Text%20TINDER%20to%20741741%20from,about%20any%20type%20of%20crisis.> .

115. See for instance, Bloom Trauma Support Program <https://bumble.com/en-in/help/bloom-for-sexual-assault-survivors--online-trauma-support-program-now-available> accessed 11 October 2025.

116. See Rajan (n 1).

117. See Contact Us <https://www.quackquack.in/v2/help/contactus.php> accessed 11 October 2025.

118. See Catherine RK O'Brien and others, 'Online Dating Platform Safeguards and Self-Protection: How Dating Platforms Characterise, Respond to, and Safeguard Against Harms', Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (ACM 2025) <https://dl.acm.org/doi/10.1145/3706599.3719825> accessed 27 August 2025.

119. For instance, Bumble allows users to report multiple facets of interaction on their dating platform. This includes content on a user's profile (both images and profile text) which may be hateful, abusive or sexually explicit. This also includes user behaviour through messaging or offline interactions which involve sexual harassment, physical violence, sexual assault and stalking.

120. See Terms of Use <https://www.aisle.co/termsofuse.html> accessed 11 October 2025.

121. See Code of Conduct <https://www.quackquack.in/codeofconduct/> accessed 11 October 2025.

122. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 (Intermediary Guidelines 2021) <https://www.meity.gov.in/static/uploads/2024/02/Information-Technology-Intermediary-Guide-

lines-and-Digital-Media-Ethics-Code-Rules-2021-updated-06.04.2023-.pdf> .

123. Intermediary Guidelines 2021, rule 3(2)(c).

124. As per rule 3(2)(b) of the Intermediary Guidelines 2021, all intermediaries must take reasonable measures to remove content that is "prima facie in the nature of any material which exposes the private area of such individual, shows such individual in full or partial nudity or shows or depicts such individual in any sexual act or conduct, or is in the nature of impersonation in an electronic form, including artificially morphed images of such individual" within 24 hours of receipt of the complaint. Although this provision under the Intermediary Guidelines has been rightly criticised for not taking into account the consent of the individual, we highlight this provision in this discussion to bring to light the absence of adequate mechanisms for reporting NCII content (even when mandated under law).

125. Intermediary Guidelines 2021, rule 3(2)(b).

126. While both platforms explicitly prohibit sexual content on their platform, users of dating platforms are still vulnerable to their intimate images being disseminated without their consent by matches on the platform or even through other means. Users should be able to report such content being shared via personal messaging and profiles that have engaged in such conduct outside the platform.

127. While the reporting interface of Quack Quack notes that its content moderation teams will look into all complaints of users within 24 hours, it does not specify a timeline for their resolution. Aisle does not specify a timeline for acknowledgement or resolution on its reporting interface, and instead notes that a member of their support staff will respond "as soon as possible."

128. Farhana Shahid, Mona Elswah and Aditya Vashistha, 'Think Outside the Data: Colonial Biases and Systemic Issues in Automated Moderation Pipelines for Low-Resource Languages' [2025] arXiv preprint arXiv:2501.13836.

129. ibid.

130. See, for instance, 'Safety And Policy Center' (Tinder) <https://policies.tinder.com/safety-and-policy/intl/en/>

131. For instance, innocuous terms could be used to target marginalised communities. The term 'shuttlecock' could be used in a derogatory manner against burka-wearing Muslim women. See Shahid et al. (n 128).

132. Rowe, J. 2022.Marginalised languages and the content moderation challenge; Nigatu, H. H.; and Raji, I.
D. 2024. "I Searched for a Religious Song in Amharic and Got Sexual Content Instead": Investigating Online Harm in Low-Resourced Languages on YouTube.

In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, FAccT '24, 141–160. New York, NY, USA: ACM; Gabriel Nicholas and Aliya Bhatia, 'Toward Better Automated Content Moderation in Low-Resource Languages' (2023) 2 Journal of Online Trust and Safety <https://www.tsjournal.org/index.php/jots/article/view/150> accessed 2 September 2024.
133. Zahra Stardust, Rosalie Gillett and Kath Albury, 'Surveillance Does Not Equal Safety: Police, Data and Consent on Dating Apps' (2023) 19 Crime, Media, Culture: An International Journal 274 <https://journals.sagepub.com/doi/10.1177/17416590221111827> accessed 27 August 2025.

134. ibid.

135. Stardust Et al. (n 133).

136. Spandana Singh, 'Dating Apps Are Even Less Transparent Than Facebook and Google' (New America) <http://newamerica.org/oti/articles/dating-apps-are-even-less-transparent-than-facebook-and-google/> accessed 16 August 2025.

137. This is also mandated under rule 3(1)(a) of the Intermediary Guidelines 2021. Further, as per rule 3(1)(c), intermediaries must periodically (at least once a year), inform users that they can remove content or terminate access to accounts upon violation of their rules and regulations, privacy policy or user agreement.

138. For instance, Bumble's Community Guidelines gives an option to users to examine different rules in more detail. See 'Bumble's Community Guidelines' (Bumble) <https://bumble.com/en-in/guidelines> On the other hand, Aisle's Community Guidelines club a wide range of harms together without providing adequate explanation. See Aisle, 'Terms of Use' <https://www.aisle.co/termsofuse.html>.

139. As per rule 3(1)(a) of Intermediary Guidelines, 2021, platforms must prominently publish rules and regulations, privacy policies and user agreements in English or any language specified in the Eighth Schedule of the Constitution for access or usage of the services by any person in the language of their choice. Also see section 5(3) of the Digital Personal Data Protection Act, 2023, which mandates data fiduciaries to give data subjects the option to access notice to process personal data in either English or any language specified in the Eighth Schedule of the Constitution

140. As per rule 3(2)(a) of the Intermediary Guidelines, 2021, all intermediaries must "prominently publish" mechanisms for user complaints. Here, "prominently publish" means clearly visible and accessible on the homescreen or through a link on the homescreen.

141. Intermediary Guidelines 2021, rule 3(2)(a).

142. Singh (n 136).

143. ibid; Emily Elena Dugdale and Hanisha Harjani, 'Dating App Cover-Up: How Tinder, Hinge, and Their Corporate Owner Keep Rape Under Wraps - The Markup' (13 February 2025) <https://themarkup.org/investigations/2025/02/13/dating-app-tinder-hinge-cover-up> accessed 19 August 2025.

144. 'Digital Services Act: Transparency report'< https://bumble.com/en-us/help/transparency-report>

145. Article 15 of the Digital Services Act also lays down mandatory transparency reporting requirements for all intermediaries (except MSMEs that are not VLOPS/VLOSEs).

146. Article 24(5) of the Digital Services Act mandates online platforms to submit "statements of reasons" for their content moderation decisions to the EC to be included in a publicly accessible machine-readable database. See EC, 'Digital Services Act Transparency Database' <https://transparency.dsa.ec.europa.eu/> accessed 16 December 2023.

147. Code of Practice 2024 <https://www.australianonlinedatingcode.com.au/wp-content/uploads/2025/07/Australian-Voluntary-Code-for-Online-Dating-Services-Code-of-Practice.pdf> .

148. Similar to the reporting obligations under the DSA, dating platforms must provide aggregate statistics information on the number of accounts terminated, classified by the policy violation. The reports must also provide disaggregated information on the content moderation undertaken by platforms classified by mechanism of detection and enforcement action. They must provide information on the platform's own content moderation initiatives, including the type of measures that impact the availability, visibility and accessibility of user-generated information. The report must also contain information on the use of automated technology (including qualitative description, purpose, accuracy and error rates). Dating platforms must also disclose the training and assistance provided to human content moderators. See Code of Practice 2024, para 8.4.

149. 'Dating Apps Move to Friend Zone in Search of Profits' The Economic Times (14 November 2024). <https://economictimes.indiatimes.com/news/international/business/dating-apps-move-to-friend-zone-in-search-of-profits/articleshow/115287860.cms?from=mdr> accessed 22 August 2025;My Story of Building India's Leading Dating App, Quack Quack : Ravi Mittal (Directed by Wealth Lessons Club, 2021) <https://www.youtube.com/watch?v=x-gJthgOslg> accessed 2 September 2025.

150. See Mike Ananny and Kate Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' (2018) 20 New Media & Society 973 <https://doi.org/10.1177/1461444816676645> accessed 28 February 2023

151. See ibid; Arvind Narayanan, 'Twitter Showed Us Its Algorithm. What Does It Tell Us?' (Knight First Amendment Institute, 10 April 2023) <http://knightcolumbia.org/blog/twitter-showed-us-its-algorithm-what-does-it-tell-us> accessed 29 August 2024; Paddy Leerssen, 'The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems' (24 February 2020) <https://papers.ssrn.com/abstract=3544009> accessed 14 June 2023.

152. Data access for researchers is one of the most significant accountability mechanisms for social media platforms. See Naomi Shiffman and Brandon Silverman, 'The Case for Transparency: How Social Media Platform Data Access Leads to Real-World Change' (Social Science Research Network, 7 May 2025) <https://papers.ssrn.com/abstract=5245757> accessed 2 September 2025. The Digital Services Act also mandates Very Large Online Search Engines and Very Large Online Platforms to grant data access for vetted researchers. See Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065>.

153. Historically, social media platforms have also voluntarily submitted themselves to audits like the GNI Company Assessments See, Global Network Initiative, 'GNI Assessment Toolkit' (October 2021) <https://globalnetworkinitiative.org/wp-content/uploads/2021/11/AT2021.pdf> accessed 30 May 2024. For more on Algorithmic audits, see Christian Sandvig and others, 'Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms' (2014) 22 Data and discrimination: converting critical concerns into productive inquiry 4349;'Why We Need to Audit Algorithms and AI from End to End' (AlgorithmWatch, 1 October 2024) <https://algorithmwatch.org/en/auditing-algorithms-and-ai-from-end-to-end/> accessed 29 January 2025.

CHAPTER 4

Digital Resistance in the Age of Algorithmic Governance: Insights from the Latin American Experience

By Nicole Solano, Jamila Venturini and Catalina Balla

Social media platforms were initially welcomed by civil society organizations, grassroots movements, and human rights defenders across Latin America as powerful tools for visibility, mobilization, and counter-narrative building[1]. They created new opportunities to expose injustice, call for collective action, share knowledge, and amplify perspectives traditionally excluded from mainstream media. Where democratic space was shrinking and public participation was increasingly restricted, these platforms became one of the few remaining avenues for exercising public voice.

While governments have attempted to "surveil and silence online expression, which continue to be bravely resisted by Latin American civil society"[2], today the promise of social media is being challenged by another growing threat: automated decision-making systems that determine what content is visible, what remains hidden, and what is removed entirely. Algorithm-powered content moderation not only affects those who publish but also shapes what information the public can access, engage with, and discuss. What was once a space to elevate marginalized voices has increasingly become a mechanism for filtering, distorting, or erasing non-dominant discourse.

In Latin America, this shift has been particularly acute. Movements that once found in social media a rare avenue to document repression and demand accountability now confront additional barriers to express their views.

> "What was once a space to elevate marginalized voices has increasingly become a mechanism for filtering, distorting, or erasing non-dominant discourse.

Navigating algorithmic ambiguity

The exercise of freedom of expression and the protection of online communities should not be mutually exclusive goals. Yet automated moderation often turns this tension into silencing or forces people to speak in code so their struggles are not erased.

Algorithmic governance does not merely limit what can be expressed in the present, it narrows the space for envisioning and articulating more just futures. By conditioning which narratives are allowed to circulate, these systems help shape the limits of political imagination. This impact is especially pronounced when speech, denouncing any form of structural violence is labeled "inappropriate" and eliminated.

Speeches which call out systemic injustice, inequality, or state violence are often censored under ambiguous and inconsistently applied community standards. Human rights defenders, activists, and journalists are particularly affected, as their content is flagged as "sensitive" or "dangerous" by systems that cannot grasp political context or intent. This creates a hierarchy of discourse that privileges dominant narratives, while dissenting content is penalized.

"The automated enforcement of moderation policies is inconsistent when dealing with analogous cases and lacks contextual understanding, leading to the silencing and self-censorship of voices, particularly those aiming to expose systemic and structural violence through sharp, uncomfortable, or denunciatory language and content", points out Paloma Lara-Castro, Public Policy Director at Derechos Digitales[3], who adds that "this is especially pronounced in countries of the Global Majority, where it is further intensified in critical social contexts such as protests". Her analysis highlights how algorithmic decisions not only fail to apply equitable standards, but also reproduce and reinforce structural biases that disproportionately affect those exercising their right to speak from positions of resistance or denunciation. In doing so, they contribute to the shrinking of civic space online in regions already facing systemic marginalization.

> By conditioning which narratives are allowed to circulate, these systems help shape the limits of political imagination.

Systemic silencing of human rights violations

The Friedrich Ebert Foundation[4], a German organization with decades of experience promoting democracy and human rights around the world, including through partnerships with social actors in Latin America, experienced this dynamic firsthand. For years, the Foundation used Instagram to share information about social programs and selection processes across Latin America. One day, the platform blocked its account for alleged violations of its community standards, offering no specific explanation. Although the account was later restored, it was permanently suspended in March 2024, without warning or a meaningful opportunity to appeal. "We had to start from scratch. We lost all our reach and followers," the team explained.

The alleged violation concerned "account integrity", a vague category emblematic of opacity that characterizes enforcement. "We weren't even notified," they added. "We only realized it when the system stopped recognizing our email." The foundation now operates with extreme caution, fully aware that lack of clarity and inconsistency forces them into self-censorship. They have called for early-warning mechanisms that allow organizations to amend posts before sanctions are applied, as well as greater transparency about which rules are being enforced and why.

The Palestinian case is illustrative of platforms discretion when applying their standards to silence certain discourses. Meta's content moderation policies have long demonstrated a troubling and consistent pattern of suppressing Palestinian voices while allowing harmful and inflammatory content targeting Palestinians who remain online. This practice has been particularly visible during moments of acute conflict. In 2018, 7amleh, the Arab Center for the Advancement of Social Media, documented the company's inequitable moderation practices in "a landmark report"[5]. The situation escalated in 2021, when a spike in digital rights violations against Palestinian content was widely reported. Posts were mass-deleted, accounts suspended, and entire media outlets shadow-banned, systematically silencing Palestinian narratives.

> The Palestinian case is illustrative of platforms discretion when applying their standards to silence certain discourses.

This resonates with Latin American experiences, where content exposing corruption, police abuse or gender inequality is frequently flagged or hidden. A similar pattern of silencing was identified in Chile during the social protests of 2019 when platform community standards were applied to remove content and accounts documenting and reporting the State violence against protesters, as documented by Derechos Digitales[6] and Datos Protegidos Foundation[7]. The latter was able to identify 169 cases of algorithmic silencing including: difficulties to upload content, deletion of accounts and removal of publications in a period of only nine days. Similar reports were registered during the massive protests in Colombia in 2021[8], when Karisma Foundation called social media platforms to be sensitive in their content moderation strategies. This double standard, that punishes whistleblowing while allowing abuse, is not a technical error. It is a direct outcome of systems built without contextual understanding, human rights frameworks, or accountability. Evidence is censored, but violence remains in place. Complaints are penalized, but not the structures that provoke them. By protecting certain narratives and silencing others, algorithms uphold the status quo and obstruct the possibilities of social change.

This double standard, that punishes whistleblowing while allowing abuse, is not a technical error. It is a direct outcome of systems built without contextual understanding, human rights frameworks, or accountability.

Gendered & Linguistic Exclusions

Experiences shared by organizations, journalists, and activists reveal that platform standards are not applied uniformly. In January 2024, Instagram disabled the account of a nonbinary creator which had spent over seven years creating and disseminating content on sex education, harm reduction, and LGBTQ+ rights. The account was disabled without clear explanation, but coincided with posting about Palestine. According to Meta, the suspension was due to violations related to "drugs and weapons," despite no such content being published for years. "It was clearly because I spoke about Palestine", they wrote when reaching out to Derechos Digitales to seek support to restore the account, describing a progressive shadowban of their content and a steep decline in visibility since October 2023.

The pattern is telling: political expression in support of certain causes is often flagged as dangerous by automated systems or hidden moderation practices. As the owner of the Instagram account explains, "social media has become a space of struggle and resistance, but also of injustice and impunity". Their testimony, echoed by others who have been censored for discussing Palestine[9], demonstrates how freedom of expression online is increasingly shaped by geopolitical interests and dominant narratives.

In Latin America, this asymmetry has a significant impact on feminist activism and journalism as well. In another case closely followed by Derechos Digitales, a developer reported having her sexual and reproductive rights application removed from Google Play Store[10], showing that the impacts of algorithmic management go way beyond social media. Developed to "reclaim sovereignty of our own bodies [...] and develop new meanings to the concepts of menstruation and menstrual health", the app challenged mainstream discourses about menstruation and was deemed in violation to Google policies regarding sexual content. The same policy does not seem to apply, for instance, to the multiple deep fakes applications widely used to generate non-consensual sexualized images of women.

Argentine journalist Luciana Peker puts it bluntly: "women are on the digital frontlines, and we've always had to put our bodies on the line"[11]. In an increasingly hostile digital environment, journalists who cover gender, reproductive rights, or feminism face not only coordinated online abuse but also algorithmic systems that amplify hate and restrict the visibility of their work. The data is alarming: 80% of women journalists who have experienced violence online, report limiting their online participation[12], and one in three have changed jobs or been fired as a result. This is not just an individual issue, it is structural where, self-censorship becomes a survival strategy[13] in an environment where exposure can lead to personal and professional harm.

Luciana Peker further warns that digital platforms, in failing to implement protection protocols for women, are undermining journalism as a public service. And the concern is broader: "It's no longer just a gender issue, it's an attack on freedom of expression." The International Journalists' Network[14] notes that content moderation follows hidden rules and shifting standards. As Forced Migration Review[15] has pointed out, when marginalized communities reclaim terminology or engage in denunciatory speech, algorithms are often quicker to remove

In an increasingly hostile digital environment, journalists who cover gender, reproductive rights, or feminism face not only coordinated online abuse but also algorithmic systems that amplify hate and restrict the visibility of their work.

The Platform Question: Power, Accountability and Global South

their content than to moderate hate speech directed at them. This imbalance is particularly severe for refugee and racialized communities, whose voices are further pushed to the margins. The opacity not only undermines visibility but also make it harder to track reports of abuse and erases key information from activists, organizations, and social movements.

Algorithmic failure to recognize the context of communities in the Global South, deepens digital exclusion and exacerbates inequalities in access to public speech, evidencing the urgent need for a digital public sphere that reflects the plurality of voices. Research by the Center for Democracy & Technology (CDT) on content moderation in Quechua[16] illustrates how these failures are amplified in non-English and Indigenous languages in Latin America. Despite Quechua being one of the most widely spoken Indigenous languages in South America, the report found that automated moderation tools, particularly large language models (LLMs), are not trained to understand it, leading to unjust removals, inconsistent enforcement, and the spread of abuse. The study highlights that these challenges are deeply gendered: women who identify as Quechua online are disproportionately targeted with harassment, while moderation teams lack the linguistic and cultural expertise to intervene effectively.

Structural weaponization of platform rules

> “Intellectual property rules have for long time been used to silence uncomfortable content, particularly when it exposes abuse of power or human rights violations

Intellectual property rules have for long time been used to silence uncomfortable content, particularly when it exposes abuse of power or human rights violations and that is also true for Latin America. This tactic is also evident in the case of Ponte Jornalismo[17], a Brazilian independent outlet investigating police violence and structural racism. In 2023, they published reports revealing that instructors at the a police academy were teaching torture and execution techniques. Videos documenting these practices were removed from YouTube following copyright claims filed by the academy. Ironically, their original videos, glorifying these practices, remained online.

Ponte appealed, but now faces the risk of having their entire channel suspended if further claims are submitted. The Brazilian Association of Investigative Journalism (Abraji) has raised concern over the dangerous precedent this sets: when

intellectual property is systematically prioritized over public interest and access to information, press freedom is severely compromised. Still in Brazil, Intervozes[18], a collective that defends the right to communication reported the abusive removal of a video denouncing misrepresentation of women in TV shows due to copyright infringement. No consideration regarding fair use was taken into account for the removal of productions that advanced critical reflection about human rights. Similar cases multiply affecting activism with automated removal ignoring local and international copyright norms.

In the hands of the powerful, platforms' rules to enforce copyright become the law, and their tools are used to erase evidence, silence dissent, and reinforce silence. It is increasingly evident that platform moderation policies are far from neutral. In the absence of clear protections for journalists and civil society, copyright enforcement can become a convenient excuse to suppress content. The situation is so alarming that law enforcement agencies have even learned to weaponize copyright protections themselves.

A striking example was documented in California, United States, where police reportedly used copyrighted music to block the dissemination of videos documenting abuse.[19] As a response, Latin American organizations are experimenting with collective archives to preserve censored content, building local platforms that operate outside corporate control, and engaging in litigation to demand transparency and accountability from global companies. They claim for algorithmic transparency for the Global South on par with the standards already applied in Europe by the Digital Services Act, or DSA.[20] The message is clear: platforms need to explain how they moderate content, under what criteria, and with what biases. This demand is not technical, it is political. It's about defending freedom of expression, and also protecting the political, cultural, and social issues that are being erased by automated systems.

Users' voices against algorithmic silencing

> “Language is where the power of social struggle resides: to name violence, demand rights, and preserve collective memory. But that same power is read as a threat by automated systems governed by inconsistently applied rules that ignore political,

Language is where the power of social struggle resides: to name violence, demand rights, and preserve collective memory. But that same power is read as a threat by automated systems governed by inconsistently applied rules that ignore political,

educational, or cultural intent. In response, individuals and collectives have developed strategies to circumvent censorship by distorting words or using symbols and euphemisms. On platforms like TikTok, YouTube, or Instagram, it is common to see expressions like “4bu$e,” “su1c1de,” or “unalive.” This tactic, known as ‘algospeak’[21], is not just a creative strategy, it is a form of resistance within a system that routinely penalizes the discussion of uncomfortable truths. Algospeak is a result of creativity and resistance in an environment that redefines what can be said, imposes automated controls, and reorganizes what is considered socially relevant. The consequences are not only individual: they affect what can be remembered, what is collectively understood, and how communities tell their stories and demands.

Automated systems that penalize words without context force activists to reshape how they speak. Many of the terms flagged by algorithms have taken years to be defined and legitimize within social movements. Being unable to name things as they are weakens the message and fragments collective memory. It threatens the heart of what Latin American democracies had to build to overcome a history of authoritarianism and abuse “the very right to truth and memory”[22].

> Individuals and collectives have developed strategies to circumvent censorship by distorting words or using symbols and euphemisms. On platforms like TikTok, YouTube, or Instagram, it is common to see expressions like “4bu$e,” “su1c1de,” or “unalive.” This tactic, known as ‘algospeak’.

The underlying issue is that algorithms centralized in few giant corporations fail to understand context. As one TikTok user explained: “AI systems struggle to grasp the intent behind a word.” This technical limitation becomes a mechanism of structural silencing, especially for complex issues like health, sexual education, or human rights. Another user added: “If a video includes sensitive language, a human should evaluate its context.” Although human moderation is key, it is also far from perfect. In fact, the different lawsuits regarding the “precarious working conditions”[23] of content moderators also highlight the lack of proper training needed to enforce platform policies consistently, often leaving the process up to the individual judgment of each person human moderating. As a result of moderation being automated, and human moderation conducted in precarious conditions, educational content on sexual health has been removed, and posts using terms like “blood” have been penalized. “Prohibiting certain words creates a taboo around crucial topics” one user noted, emphasizing that algorithmic censorship doesn’t merely delete content, it reshapes the boundaries of public discourse.

The Platform Question: Power, Accountability and Global South

Such an environment may result in a reality where those who understand the codes may avoid censorship while those who don’t, remain excluded[24]. Digital Future Society warns that activists, content creators, and journalists are being forced to constantly reinvent their language to remain visible in a system dominated by algorithmic opacity[25].

When language must be constantly altered to avoid censorship, it becomes harder to build shared meaning and ensure accessibility. While adapting is necessary, it has limits. How much can we reshape language before it loses its meaning? What happens to those who can’t decode algospeak? Whose voices are systematically excluded from the digital conversation?

Naming things clearly is a political act. Calling realities by their names, without euphemisms, is itself a form of resistance. Reclaiming algorithmic language does not mean normalizing it. It means making visible a system that silences. Some organizations have begun building collective resources to decode this language, such as the ‘Algospeak Dictionary’[26] developed by the Digital Rights Collective for Sexual and Reproductive Health. Democratizing these tools is essential so that more people can participate in resistance.

Building pathways of resistance

Resistance is about creating alternatives that embody different values. Latin American collectives are also experimenting with decentralized technologies and feminist servers that place care and community above profit. Communities are migrating to hybrid models that blend traditional media with fediverse platforms, ensuring that local struggles are not algorithmically buried. Cooperatives of journalists are joining forces to publish in distributed networks where content cannot be unilaterally removed by corporations. These are not futuristic dreams but concrete practices already happening across the region. Practices that point toward an internet where freedom of expression is not contingent on opaque standards but on collective governance. These practices are deeply rooted in the region’s history of grassroots organizing, echoing past struggles where communities built their own media and communication channels to bypass censorship and ensure their stories were heard.

> When language must be constantly altered to avoid censorship, it becomes harder to build shared meaning and ensure accessibility. While adapting is necessary, it has limits.

It is urgent to build and support networks that are not subject to corporate interests. Decentralized platforms like the fediverse, community media, digital cooperatives, and alliances between civil society organizations offer alternatives to preserve our stories and sustain collective memory. Resistance also means protecting spaces where our words are not silenced by opaque algorithms and reclaiming our right to develop alternative technologies. Words matter. And when we write them in full, even when they're uncomfortable or challenging, we affirm that some truths cannot be contained, not even by an algorithm. Recovering, protecting, and defending language is an urgent task for those working toward a more just internet. In an era of automation and silent censorship, reclaiming the right to name is also reclaiming the right to exist. Adaptation is not enough. We must transform the digital environment so our words never have to hide. Content moderation must move beyond punitive approaches and be reimagined as a form of participatory governance, one that is accountable to the communities it affects and that centers freedom of expression, dignity, and equity as fundamental principles, not as optional considerations.

We must transform the digital environment so our words never have to hide.

Endnotes

[1] Outside the algorithm: How to communicate beyond social media. (2025, July 23). Derechos Digitales. https://www.derechosdigitales.org/en/recursos/outside-the-algorithm-how-to-communicate-beyond-social-media/

[2] Derechos Digitales, & Camacho, L. (2024). Perfilamiento en redes sociales y ciberpatrullaje como nuevas modalidades de la vigilancia masiva desplegada por los Estados: casos relevantes en América Latina. In Derechos Digitales [Report]. https://www.derechosdigitales.org/wp-content/uploads/Informe-RELE-vigilancia-masiva_cerrado.pdf

[3] Derechos Digitales is a Latin American-based non-profit organization whose mission is to defend, promote, and develop human rights in the digital environment through research, information dissemination, and advocacy in public policy and private practice, fostering social change grounded in respect for the rights and dignity of individuals. For more information visit https://www.derechosdigitales.org/.

[4] For more details about Friedrich-Ebert-Stiftung, See https://www.fes.de/en/

[5] 7amleh – Arab Center for Social Media Advancement. (n.d.). Hamleh - 7AMLEH releases new report: "Erased and Suppressed: Palestinian Testimonies of Meta's Censorship." https://7amleh.org/post/erased-and-suppressed-palestinian-testimonies-of-meta-s-censorship-en.

[6] Situación de derechos humanos y el uso de tecnología en el contexto de la protesta social en Chile (Octubre-Noviembre de 2019). (2025, July 23). Derechos Digitales. https://www.derechosdigitales.org/recursos/situacion-de-derechos-humanos-y-el-uso-de-tecnologia-en-el-contexto-de-la-protesta-social-en-chile-octubre-noviembre-de-2019/

[7] Fundación Datos Protegidos & Observatorio del Derecho a la Comunicación. (2019). Libertad de expresión en el contexto de las protestas y movilizaciones sociales en Chile durante el estado de emergencia entre el 18 y el 27 de octubre 2019. https://datosprotegidos.org/wp-content/uploads/2019/11/Informe_CIDH_Preliminar_DP_ODC-1.pdf

[8] Press Release (2021, May 14). Fallas de internet, bloqueos de redes y censura de contenidos en protestas: Realidades y retos para el ejercicio de los derechos humanos en los contextos digitales. Fundación Karisma. https://web.karisma.org.co/paronacionalcolombia-fallas-de-internet-bloqueos-de-redes-censura-de-contenidos-realidades-y-retos-para-el-ejercicio-de-los-derechos-humanos-en-los-contextos-digitales/

[9] Demoted, deleted, and denied: there's more than just shadowbanning on Instagram – the markup. (2024, February 25). https://themarkup.org/automated-censorship/2024/02/25/demoted-deleted-and-denied-theres-more-than-just-shadowbanning-on-instagram

[10] Venturini, J. (n.d.). A feminist lead towards an alternative digital future for Latin America. Bot Populi. https://botpopuli.net/a-feminist-lead-towards-an-alternative-digital-future-for-latin-america/

[11] El periodismo feminista como resistencia en tiempos de conservadurismo extremo y violencia digital. (2025, March 11). Fundación Gabo. https://fundaciongabo.org/es/etica-periodistica/entrevistas/el-periodismo-feminista-como-resistencia-en-tiempos-de?fbclid=IwY2xjawLQ-fBleHRuA2FlbQIxMABicmlkETFidTZuejB2ZlJUMFFFc29pAR4PDKnuUhfdg5IIY5aeaDYXv12crzCCDennH7_-TmJ8h-aCiVkHrvbF8cw1TQ_aem_e5ni7n1YB5vpSzpE9mYgQg

[12] Beck, I., Alcaraz, F., Rodríguez, P., Alianza Regional por la Libre Expresión e Información, & ONU Mujeres. (2022). Violencia de género en línea hacia mujeres con voz pública. Impacto en la libertad de expresión (D. Urribarri & A. Arias, Eds.). https://lac.unwomen.org/sites/default/files/2023-03/Informe_ViolenciaEnLinea-16Mar23.pdf

[13] Violencia digital: nuevos formatos, viejas formas de censura. (2025, July 31). Derechos Digitales. https://www.derechosdigitales.org/recursos/violencia-digital-nuevos-formatos-viejas-formas-de-censura/

[14] Pase, I. (n.d.). Por qué es importante para el periodismo descifrar el "algo-

The Platform Question: Power, Accountability and Global South

speak" | Red internacional de periodistas. Red Internacional De Periodistas. https://ijnet.org/es/story/por-que-es-importante-para-el-periodismo-descifrar-el-algospeak

[15] Wells, A. (2024, August 20). Resistencia, poder, representación y censura algorítmica digitales de las comunidades refugiadas - Forced Migration Review. Forced Migration Review. https://www.fmreview.org/disrupcion-digital/wells/

[16] Thakur, D. (2025). Moderating Quechua Content on Social Media. https://cdt.org/wp-content/uploads/2025/06/2025-06-25-Quechua-Report-English-final.pdf

[17] Vídeos da Ponte são removidos do YouTube após reportagens sobre exaltação a tortura em curso para aspirantes a PMs. (n.d.).https://abraji.org.br/videos-da-ponte-sao-removidos-do-youtube-apos-reportagens-sobre-exaltacao-a-tortura-em-curso-para-aspirantes-a-pms

[18] Intervozes notificou YouTube pela remoção de vídeos críticos à programação das emissoras de TV – Intervozes. (May 12, 2018.). https://intervozes.org.br/publication/intervozes-notificou-o-google-brasil-empresa-controladora-do-youtube-pela-remocao-de-videos-criticos-a-programacao-das-emissoras-de-tv/

[19] BBC News. (2021, July 2). US officer plays Taylor Swift song to try to block video. https://www.bbc.com/news/technology-57698858

[20] Home - DSA Transparency Database. (n.d.). https://transparency.dsa.ec.europa.eu/

[21] Aleksic, A. (2025). Algospeak: How social media is transforming the future of language. Knopf.

[22] United Nations. (n.d.). Right to Truth Day - EN | United Nations. https://www.un.org/en/observances/right-to-truth-day

[23] Dias, T., & Schurig, S. (2024, December 9). Moderators received 7 cents per task to comb through violence, pornography, and extreme content on X. Intercept Brasil. https://www.intercept.com.br/2024/12/09/moderators-received-7-cents-per-task-to-comb-through-violence-pornography-and-extreme-content-on-x/

[24] Berniga, L. G., & Pavlicich, J. Q. (2023). Desinformación y discursos de odio: Amenazas digitales a la participación política de las mujeres en elecciones. https://doi.org/10.31752/idea.2023.101

[25] Amenazan los algoritmos la libertad de expresión? | Digital Future Society. (2021, April 21). Digital Future Society. https://digitalfuturesociety.com/es/qanda/amenazan-los-algoritmos-la-libertad-de-expresion/

[26] https://www.algospeak.net/dictionary

The Platform Question: Power, Accountability and Global South

CHAPTER 5

Platformed Lives: Technology, Accountability, and the Reshaping of Everyday Work

By Carina Singh, Khush Vachharajani and Rakshita Swamy

Over 90% of India's workforce remains informal, characterized by low wages, insecure employment, and a near-total absence of social security protections. While precarity has intensified and real wages have stagnated over the past few years, a complex architecture of digital technologies has also been introduced into the everyday lives of workers. These technologies demand not only empirical scrutiny but also critical reflection grounded in the principles of natural justice, social accountability, and workers' rights. Initiatives such as Aadhaar, the Unified Payments Interface (UPI), Direct Benefit Transfer (DBT) systems, facial recognition technologies, and the broader imagination of IndiaStack have been central to this digital reform project.

While presented as instruments of modernization, transparency, and inclusion, their emergence cannot be understood purely as a domestic policy shift. Rather, India's digitization push reflects a transnational convergence of state ambition, international financial influence, corporate and philanthropic interests. International financial institutions such as the World Bank, along with global philanthropies including the Gates and Ford Foundations, have promoted digital identity and interoperable data systems as efficient, "leak-proof" mechanisms for welfare delivery and financial inclusion. Simultaneously, domestic industry networks and policy entrepreneurs, most notably the India Software Product Industry Roundtable (iSPIRT) designed the technical architectures and policy frameworks that legitimize the integration of private actors into public service delivery.

This coalition has rapidly produced near-universal biometric identification, centralised welfare transfers through the

"India's digitization push reflects a transnational convergence of state ambition, international financial influence, corporate and philanthropic interests.

DBT framework, and an expanding ecosystem of API-driven services under the IndiaStack. Both state and private actors interchangeably drop loaded terms like “transparency”, “efficiency”, “accountability”, “freedom”, “flexibility” and “innovation” in the name of public welfare and social good. Yet it has also normalized a new political economy of welfare, one in which public goods are reimagined as digital infrastructure, employer-employee relationships are mediated through digital platforms, and the boundaries between justifications for inclusion, surveillance, and exclusion become increasingly blurred.

Positioning technology as both the means and the end of governance reform, the state and its private partners have advanced a set of recurring justifications for these interventions. Six stated intentions, in particular, have framed the deployment of digital systems across labour and welfare domains:

- Legitimacy and Targeting – to ensure that entitlements reach the “correct” person and eliminate fraud
- Efficiency and Convenience – to reduce bureaucratic friction and deliver services directly “at the door step”
- Transparency and Accountability – to use data to track performance and curb corruption
- Flexibility and Autonomy – to empower individuals through choice
- Inclusion and Empowerment – to integrate the poor into formal digital and financial systems
- Innovation and Modernization – to establish India as a global model of digital public infrastructure and “unlock” value and potential.

The following sections interrogate these narratives by examining how digital infrastructures, across MGNREGA, platform-based gig work, and street vending reshape access to livelihoods and entitlements, redefine accountability, and reconfigure the social contract between labour, state, and society at large. It fundamentally reconstitutes how workers access livelihoods, how entitlements flow, and where accountability resides when systems fail.

> “Positioning technology as both the means and the end of governance reform, the state and its private partners have advanced a set of recurring justifications for these interventions.

The Platform Question: Power, Accountability and Global South

MGNREGA: The State’s Laboratory for Plugging Technology at Scale

MGNREGA, the world’s largest public employment guarantee programme spending about 12 billion USD annually is implemented through an integrated management information system (MIS) called NREGASoft. Intended as a tool for transparency and accountability, its pathbreaking potential lay in the public investment that was channeled towards building a real-time, transaction-based digital platform that would not only help administer all aspects of the programme but also disclose crucial information relating to job card registration, muster roll issuance, wage payments and so on. In the process, however, the State has also managed to selectively shape it as a tool to manipulate information and deny workers their rights guaranteed by law.

For instance, the MGNREG Act provides for compensation to be paid to workers if wages are delayed beyond the statutory period of 15 days, wherein Stage 1 (8 days) involves generation of the Fund Transfer Order (FTO) by the State Government and Stage 2 (7 days) involves processing of the FTO and payment of wages by the Central Government. The MIS, though, has been coded in a manner that the completion of Stage 1 i.e., FTO generation, is misleadingly shown as payment of wages to the worker. Only Stage 1 delays, which is the responsibility of the State Government, are publicly visible on the MIS. Whereas delays in Stage 2 , which involves the actual crediting of wages into the worker’s account and is the responsibility of the Central Government, are neither calculated nor displayed. The misrepresentation is deliberately built into the system to hide the liability of the Central Government from the public.

Similarly for unemployment allowance, which is statutorily mandated to be paid if work is not provided within 15 days of a worker demanding it. Yet again, the MIS has been selectively engineered to calculate and show unemployment allowance in respect of only those workers whom it has actually been paid to, despite there being other workers who may be eligible but have not been paid. Given both work demand and muster

> “The MIS has been selectively engineered to calculate and show unemployment allowance in respect of only those workers whom it has actually been paid to, despite there being other workers who may be eligible but have not been paid.

rolls are administered through the MIS, this could have been easily automated and computed to ensure workers are paid their statutory dues, but the omission is by design to escape accountability. In addition to the denial of wages, the MIS has also been employed to deny the right to work itself, as evidenced by the arbitrary cap of 20 open works per Panchayat introduced on NREGASoft from July 2022. A work is considered "open" until all processes relating to it have been completed, including wage payment, material costs etc. This technical intervention leads to workers being denied work under NREGA, given the MIS does not allow for new muster rolls to be issued if there are already twenty open works in the Panchayat. Moreover, it goes against the principles of decentralisation and democratic participation fundamental to MGNREGA, wherein works are to be planned and executed at the Panchayat level in consultation with the Gram Sabha based on local needs. But the imposition of the 20-works ceiling undermines the authority of the Gram Sabha, diminishes the role of the people in providing for their infrastructural needs, and makes NREGA a top-down scheme at the behest of the government in power.

Two other technologies introduced in the name of "citizen oversight" and "preventing fraud" in MGNREGA are digital attendance through the National Mobile Monitoring System (NMMS) and wage payment using the Aadhaar-Based Payment System (ABPS). NMMS requires time-stamped, geo-tagged photos to be uploaded from the worksite twice a day (minimum four hours apart) as evidence of workers' attendance, which can be a herculean task in remote rural areas having poor internet infrastructure and low digital literacy. Previously, MGNREGA attendance was captured by Mates on physical muster rolls which enabled workers to inspect the muster rolls before signing or putting their thumbprint on it. But with attendance now being recorded solely through the NMMS app, accessible only through the Mates' login, it has effectively done away with workers' oversight altogether.

> "The absence of physical muster rolls also means workers do not have any proof of the actual number of workdays completed by them, or worse, they themselves are not aware of discrepancies in their attendance.

The absence of physical muster rolls also means workers do not have any proof of the actual number of workdays completed by them, or worse, they themselves are not aware of discrepancies in their attendance. ABPS, on the other hand, involves a cumbersome process of seeding the worker's job card and bank account with Aadhaar as well as connecting their bank account with the NPCI mapper. This involves meeting

stringent KYC requirements where even minor discrepancies in spelling of names, addresses etc. lead to failure. Thousands of workers have not been paid for work done by them either due to their attendance not being marked on NMMS or for not being ABPS-compliant. Moreover, ABPS has resulted in widespread deletion of job cards, with 9 crore workers' names deleted since 2022.

Street Vendors and Digital Dispossession in the Informal Economy

Street vending sustains the urban economy and provides livelihoods to millions who remain excluded from formal employment. Recognizing this, the Street Vendors (Protection of Livelihood and Regulation of Street Vending) Act, 2014 was a landmark victory, one that codified the right to vending and institutionalized participatory governance through Town Vending Committees (TVCs). The law's spirit is rooted in inclusion, decentralization, and democratic oversight. Yet, across India, its implementation remains mired in bureaucratic opacity and administrative inertia.

In Meghalaya, the Greater Shillong Progressive Hawkers and Street Vendors Association (GSPHSVA) has been at the forefront of efforts to ensure that this progressive legislation is realized in both letter and spirit. Representing vendors in the Town Vending Committee, the association has long advocated for transparent and accountable systems-especially for conducting the in-situ survey, the first and most crucial step in identifying existing vendors and recognizing their legal right to livelihood. Historically, this process has been fraught with manipulation, exclusion, and administrative chaos. To address these challenges, the Association supported the idea of a digital survey that could ensure accuracy, legitimacy, and real-time accountability through features such as acknowledgement receipts, unique tracking numbers, and public disclosure of survey results.

However, the promise of this digital intervention quickly unraveled. Despite a public expenditure of nearly Rs 10 lakh, the survey system failed to generate even the most basic transparency tools. No real-time receipts were issued, no tracking numbers were assigned, and most critically, the full

> "Despite a public expenditure of nearly Rs 10 lakh, the survey system failed to generate even the most basic transparency tools. No real-time receipts were issued, no tracking numbers were assigned, and most critically, the full survey data was never disclosed.

survey data was never disclosed. Instead, local authorities uploaded only a partial list of "eligible vendors", without explaining the criteria for eligibility or the grounds for exclusion. This violated the core provision of the Street Vendors Act, which mandates that all existing vendors identified in the survey are to be issued certificates of vending and accommodated in vending zones. Out of 1,400 vendors surveyed in-situ across Shillong, only 760 were arbitrarily deemed eligible. This was an act of digital erasure that disenfranchised nearly 50% of those who were officially documented.

The consequences have been immediate and devastating. In Police Bazaar, one of Shillong's oldest and most vibrant heritage markets, nearly 100 hawkers, about a quarter of those surveyed have already been evicted for not possessing vending certificates, despite being recorded in the digital survey. Similar patterns are emerging elsewhere, in Laitumkhrah, the projections suggest that as many as 82% of surveyed vendors risk eviction if this exclusionary model continues.

The administration has consistently invoked the "digital sanctity" of the system to justify these exclusions, claiming that errors are impossible because the survey was conducted online. This faith in technological infallibility has enabled a dangerous abdication of responsibility. There are no parallel non-digital processes for claims, objections, or grievance redressal effectively denying vendors the right to be heard, a fundamental principle of natural justice. Instead of empowering hawkers with a verifiable digital footprint, the system has produced an opaque database that dictates their eligibility to exist in public space. The outcome is a stark inversion of the Act's intent.

> "Street vendors, already among the most precarious urban workers, now face new threats to their survival, mediated not through the visible apparatus of state coercion but through the silent authority of a database.

What was meant to protect livelihoods has instead become an instrument of digital dispossession. Street vendors, already among the most precarious urban workers, now face new threats to their survival, mediated not through the visible apparatus of state coercion but through the silent authority of a database. The digital survey in Shillong exemplifies how the language of transparency and efficiency can be weaponized to obscure injustice, displace accountability, and erase entire communities from the right to work with dignity.

Platform-Based Gig Work: The Illusion of Freedom

While the above examples illustrate the use of publicly-funded IT systems to violate the letter of the law, digital platforms are also being projected as enablers of autonomy and flexibility as in the case of platform-based gig work. This is a digitally-mediated work relationship touted to be a blessing for the unemployed as it enables freedom, flexibility and choice to work on one's own terms. Sign in and take up a "gig" whenever you like, and 'get off' the platform when you don't wish to work. But the truth could not be further. Platform aggregators frequently perpetuate traditional mechanisms of labor exploitation once the domain of contractors and factory owners, be it arbitrary wage-cuts, withholding critical information or penalising collective action. A vast, historically marginalized workforce now encounters these familiar challenges through modern veneers like application-based interfaces and performance ratings.

Algorithms designed by platform aggregators enable them to control and discipline workers, to ensure that they follow the companies' rules and cooperate with customers. They do so in two key ways, "nudge" workers to perform actions that are profitable for the company and use threats of suspension or deactivation of workers' accounts and penalties to prevent workers from acting against the platform-companies' protocols and policies. Documented evidence and protests against algorithms designed to surveil, manipulate, inhibit worker autonomy and control, build opaqueness, build in bias and discrimination have started to emerge. In response, the movement for gig worker rights has rightly focused on demanding transparency and unveiling the realities concealed by digital systems.

Conclusion: The Reconfiguration of Power

The above examples demonstrate how digital technologies are being used to redraw lines of access and dignity between the poor and the powerful. In each of these instances, databases, algorithms, and digital platforms play a central role

> "Platform aggregators frequently perpetuate traditional mechanisms of labor exploitation once the domain of contractors and factory owners, be it arbitrary wage-cuts, withholding critical information or penalising collective action.

in recalibrating work relations, only to further entrench power in the hands of those who already possess it. Beyond specific instances of harm, these systems reshape societal relations, redistribute economic resources, and erode the democratic fabric of governance.

Databases that monitor and control people's engagement with public services, from the cradle to the grave, confer immense power upon those who design and administer them. The digitization of birth and death registration, social security pensions, public hospitals, and welfare programs has centralized authority in the state, while adequate legal safeguards remain absent. Digital technologies amplify exploitation by depersonalizing public service delivery and obscuring accountability. When confronted with systemic failures, public officials routinely displace blame onto algorithms or data systems, suggesting that the technology itself is at fault. This inversion of responsibility places the burden of proof on marginalized citizens, who must demonstrate the harms inflicted upon them by systems they neither designed nor consented to. The architecture thus encodes top-down, technocratic visions of governance, efficiency, transparency, financial inclusion, while enabling unprecedented surveillance, new pathways of exclusion, and a diffusion of accountability that undermines worker agency.

Beyond specific instances of harm, these systems reshape societal relations, redistribute economic resources, and erode the democratic fabric of governance.

Yet this trajectory is neither inevitable nor irreversible. Across India, workers and collectives are reclaiming digital tools for democratic accountability and social justice, through initiatives such as the Jan Soochna Portal in Rajasthan, the Sand Mining Watch archives, the Coastal Resource Centre's geospatial monitoring, and emerging "Early Warning Systems" against land grabbing in the Western Ghats. Legislative developments across states like Rajasthan, Karnataka, Bihar, and Jharkhand, which seek to secure the rights of platform-based gig workers, similarly demonstrate the transformative potential of technology when conceptualized and governed by those it is meant to serve. Provisions such as auto-registration, transparent calculation of social security contributions, real-time disclosure of algorithmic decisions, and worker-led governance models illustrate how digital infrastructures can be reoriented toward collective empowerment.

To ensure that such transformations become the norm rather than the exception, the design, development, and deployment of digital technologies must be guided by a set of principles that place people, not data at the center of digital governance.

1. Rights Before Efficiency: Every digital system must uphold constitutional values of dignity, equality, and justice over the pursuit of administrative or technological efficiency.
2. Lawful and Transparent Mandate: No digital infrastructure should function without clear legislative authorization, transparent purpose, and public oversight.
3. Participation and Co-Governance: Workers, citizens, and collectives must be active participants in the design, testing, and governance of technologies that affect their rights and livelihoods.
4. Accountability by Design: Each digital system must define who is responsible for errors and harms, provide accessible grievance redressal, and guarantee reparations when rights are violated.
5. Open and Auditable Systems: Algorithms, datasets, and processes used in welfare and labour governance must be subject to independent social and technical audits.
6. Separation of Public and Commercial Interests: Public welfare technologies must remain free from private profit motives; data collected for public good cannot be used for commercial exploitation.
7. Plural Modes of Access: No service or entitlement should be exclusively digital. All systems must provide offline and assisted modes of access, with robust provisions for grievance redress and human facilitation at every level.
8. Decentralized Control and Human Oversight: Digital systems must strengthen - not replace - local institutions and human decision-making, ensuring that people remain at the center of technological governance.
9. Continuous Review and Renewal: All digital infrastructures must have sunset clauses and periodic public audits to prevent mission creep and preserve democratic accountability.

Provisions such as auto-registration, transparent calculation of social security contributions, real-time disclosure of algorithmic decisions, and worker-led governance models illustrate how digital infrastructures can be reoriented toward collective empowerment.

CHAPTER 6

Reclaiming the Missing Story: Platform Accountability In Myanmar

By Htaike Aung

Context

When Sarah Wynn-Williams' book "Careless People" (Wynn-Williams, 2025) came out, I felt two things at once. Relief, because Myanmar's story of digital harm was being recognized in a global narrative. And unease, because the way it was written carried familiar traces of omission. Her chapter on Myanmar is already reaching a wide audience, shaping how others will understand our history with social media. But the picture is incomplete.

This is not a book review. I do not want to spend these pages dissecting her arguments line by line. What I want is to tell the part of the story that was left out, the one lived by those of us in Myanmar who were already watching, warning, and working long before the crisis became front-page news. For me, this is about memory, responsibility, and care. The omissions in that chapter reflect a larger pattern: the labor of local communities, especially in fragile contexts like Myanmar, often gets erased or reduced to footnotes. Yet we were there from the very beginning, doing the work of monitoring, documenting, and sounding alarms.

In the early 2010s, I co-founded the Myanmar ICT for Development Organization (MIDO), the country's first digital rights group. Those were years when Myanmar was only just stepping into the digital age. We saw both the excitement and the dangers. Facebook was opening new doors, but it was also sowing division at a speed we had not seen before. We tried to make the world listen. Often, it felt like we were speaking into silence. This essay is my attempt to fill in the missing threads, not with anger, but with care. Care for memory, for those who labored unseen, and for the lessons still waiting to be learned.

> "Facebook was opening new doors, but it was also sowing division at a speed we had not seen before.

Myanmar in Transition

To understand why those years mattered so much, you have to remember the atmosphere in Myanmar in the early 2010s. After decades of military rule, the country was taking tentative steps toward democracy. Political prisoners were being released. Censorship was relaxed. The outside world flooded in (journalists, aid workers, investors) each with their own version of the "opening up of Myanmar." At the same time, something else was happening: the internet was arriving almost overnight. Only a couple of years earlier, a SIM card could cost as much as $2000-$500 USD. By 2013, prices had started dropping, and international operators were rushing in. Suddenly, millions of people were online for the first time. Most of them skipped desktops entirely and went straight to cheap smartphones. We called it the shift from No Phone to a Smart Phone country. For many, Facebook was the internet. It became the place to share photos, gossip, prayers, political debates. It was intoxicating, this sudden sense of being connected to one another after years of isolation. But the speed of adoption outpaced everything else: laws, media institutions, community norms.

We reached out to Facebook. We documented examples and tried to explain why content moderation could not be outsourced to users alone. What outsiders often dismissed as "a digital literacy issue" was, for us, clearly tied to platform incentives, weak moderation in Burmese, and deep historical fault lines being exploited through new tools.

However, it did not take long for Facebook to turn into a space of fear. Old tensions, ethnic and religious divides that had simmered for decades, suddenly had new megaphones. Pages glorified extremist monks and networks. Rumors of Muslims abducting Buddhist women spread overnight. Content designed to provoke outrage found eager audiences. From the beginning, many of us in civil society saw what was coming. This was not just random chatter. It was coordinated, systematic, and amplified by the very design of the platform. A rumor posted in one township could jump across the country in hours, landing in communities that had never even met the people they were being told to fear.

We started warning as early as 2012. We reached out to Facebook. We documented examples and tried to explain why content moderation could not be outsourced to users alone. What outsiders often dismissed as "a digital literacy issue" was, for us, clearly tied to platform incentives, weak moderation in Burmese, and deep historical fault lines being exploited through new tools. By 2013, the risks were flashing red. The violence in Meiktila that March, where dozens were killed and thousands displaced, was preceded by waves of online posts spreading dehumanizing language and inflammatory rumors. For us, the link was undenia

The Early Warning

MIDO was born in that tension. In 2012, we set out to build Myanmar's first digital rights organization. At the start, our focus was broad: digital literacy, access, freedom of expression online. But very quickly, our work also began to focus on the darker side of Myanmar's sudden connectivity. We started tracking Facebook. Every day, we saw how hate speech spread faster than anything else. We saw which pages were driving divisions. We saw how narratives targeting Muslims and the Rohingya community traveled with frightening speed. We tried to respond. We supported journalists and youth groups to recognize and counter hate speech. We published our findings. We reached out to Facebook with translations and examples, trying to explain what was happening. But the scale was overwhelming. Posts went viral in hours; our reports took days.

What made it harder was the silence. Platforms rarely replied, or sent us stock answers about reporting mechanisms. International researchers visited Yangon, wrote papers about Myanmar's "Facebook revolution," but rarely cited or credited the local civil society that had been monitoring the risks since day one. We were not just "users" of the platform. We were investigators, monitors, first responders. But to the outside world, we remained background characters, not experts in those first few years.

The Missing Story

I remember the 2013 Internet Governance Forum. That is where I met Sarah Wynn-Williams. We spoke briefly about Myanmar's fragile online space. Later, I sent emails outlining our concerns in more detail: how hate speech was spreading, how Facebook was already playing a dangerous role.

We started tracking Facebook. Every day, we saw how hate speech spread faster than anything else. We saw which pages were driving divisions.

Back home, the ground was shifting quickly. The Meiktila violence confirmed our worst fears. Online dehumanization had translated into real-world bloodshed. At MIDO, our team was scrambling. We built systems to monitor hate speech. We saved screenshots before posts disappeared. We translated Burmese content into English for outsiders who could not read it. We trained volunteers to spot and flag dangerous narratives. We stayed up late, drafting emails to Facebook staff, trying to get urgent posts taken down. Most of the time, we heard nothing back. And when replies came, they felt perfunctory:

"report more content," or "this is about digital literacy." But we were aware that, it was the design of the platform itself: algorithms were rewarding outrage, Burmese-language moderation barely existed, and a company was unwilling to invest in local expertise. However, we were invisible in the global conversation. International researchers described Myanmar with sweeping metaphors, painting it as exotic or unknowable. Meanwhile, those of us living inside it were already waist-deep in the work of documenting and resisting digital harm.

The emotional toll is harder to describe. We scrolled through endless content of hate, often directed at people we knew. We felt the weight of knowing those words, as they had the potential of inciting violence, and sometimes they did. Therefore, civil society groups started sharing strategies, building networks, training journalists, and engaging religious leaders. We believed the internet could still be a space for dialogue if shaped responsibly. Looking back, those years were a constant race, hate was spreading faster than we were able to document, warnings were being ignored until, global narratives were talking about Myanmar's case as cautionary tale, without acknowledging the people who tried to prevent it.

That is why omissions hurt. They have the potential of hinting that the crisis came from nowhere, as if nobody saw it coming. When I read global accounts that do not include our voices, I think about how stories travel. A book in London or New York gets picked up in classrooms, cited in policy debates, retold in news articles. Soon, that version becomes the version. That is why narrative omission becomes important. It erases the labor of those who did the hard, messy work. And it misleads future readers into thinking nothing could have been done, that harm was sudden and inevitable. But the truth is, harm was not inevitable. Recognizing that matters, not just for justice in Myanmar's story, but for the next country where platforms dismiss local voices until it is too late. For me, narrative responsibility is part of accountability. Just as platforms must take responsibility for the systems they build, authors and researchers must take responsibility for the stories they tell.

"

We were aware that, it was the design of the platform itself: algorithms were rewarding outrage, Burmese-language moderation barely existed, and a company was unwilling to invest in local expertise.

Conclusion:

I am writing this piece to correct, to remember, and to care. Careless People will continue to be widely read, and it will shape how Myanmar's digital crisis is remembered. My hope is that by sharing our side of the story, readers will see a fuller picture, one where Myanmar is not just an exotic backdrop, but a place where people were urgently documenting, warning, and resisting.

Even after those early days, we did not stop. In 2018, civil society groups in Myanmar sent an open letter to Mark Zuckerberg, pointing out that Facebook had failed to act on the very hate speech we had been fighting for years and also how they were taking credit for the work of us. We are still resisting, still documenting, still demanding accountability to platforms. The danger we face is not only violence under the coup, but also being erased from the story, our warnings, our resistance, and our accountability work slipping out of the international narratives. That is why we need to be deliberate with our narratives. By reclaiming the missing stories, I hope to honor the work of Myanmar's digital rights community and to remind us that stories matter. How we tell them matters even more.

Reference

Wynn-Williams, S. (2025). Careless people Edizioni Mondadori.

"

The danger we face is not only violence under the coup, but also being erased from the story, our warnings, our resistance, and our accountability work slipping out of the international narratives.

CHAPTER 7

The Gig Economy and Platform Workers: A View from the Ground

By Shaik Salauddin

The Telangana Gig and Platform Workers' Union (TGPWU), in partnership with Digital Empowerment Foundation (DEF), has launched a process of co-building and seeking endorsement for a comprehensive Citizen Mandate in course of DEF's flagship annual summit, the Digital Citizen Summit 2025. DCS is being held at T-hub, Hyderabad, between 14-15th November 2025 in partnership with Government of Telangana, T-Hub and Centre for Development Policy and Practice, Hyderabad, urging state and national authorities to ensure social security, fair wages, and collective bargaining rights for gig workers. The Mandate aims to amplify workers' voices across digital labour-based platforms, demanding recognition and justice in the digital economy. This chapter is an excerpt from an Interview taken of Shaik Salauddin, Founder President, Telangana Gig and Platform Workers Union (TGPWU) by the Research Team from Digital Empowerment Foundation.

> The conversation around platforms and people remains incomplete without deliberating on the fact that gig workers are classified as partners instead of employees. According to you, how does this allow platforms to escape accountability for wages, benefits, and social security?

In India, app-based platforms such as Ola, Uber, Swiggy, Zomato, Amazon, and Urban Company classify their workers as "independent contractors" or "partners" rather than employees. This terminology is not accidental, rather a deliberate legal and business strategy designed to avoid responsibility for fair wages, social security, and workplace protections. By calling workers "partners," platforms shift all

"By calling workers "partners," platforms shift all operational risks onto the individuals who actually perform the labour while retaining total control over pricing, access to work, and incentives. The so-called partnership is deeply unequal.

operational risks onto the individuals who actually perform the labour while retaining total control over pricing, access to work, and incentives. The so-called partnership is deeply unequal. A genuine partnership implies shared decision making, profit-sharing, and autonomy over work. In reality, platform workers have none of these. Algorithms decide when and where they can work, what rates they are paid, and whether they will be penalized or deactivated. Drivers and delivery workers cannot negotiate their pay, contest unfair ratings, or appeal suspensions easily. Yet they are responsible for fuel, vehicle maintenance, insurance, and other expenses—essentially subsidizing the business models of billion-dollar companies. This misclassification has far-reaching implications. It enables platforms to evade the payment of Employees' Provident Fund (EPF), Employees' State Insurance (ESI), gratuity, paid leave, maternity benefits, and accident compensation. It also denies workers the protection of the Industrial Disputes Act, 1947 and the Minimum Wages Act, 1948, which safeguard workers' rights to fair remuneration and collective bargaining.

The result is a situation of "wage-less growth" where digital platforms thrive and expand, while the very people powering them live with economic insecurity and debt. From the union's perspective, we view this as a form of disguised employment. Platforms exercise employer-like control without accepting employer-like responsibility. Across jurisdictions, from the UK to California, courts and regulators are increasingly recognizing this contradiction and reclassifying gig workers as employees or "dependent contractors." India too must move in this direction, to ensure that digital innovation does not come at the cost of basic labour rights.

"

From the union's perspective, we view this as a form of disguised employment. Platforms exercise employer-like control without accepting employer-like responsibility.

> Given that workers often face sudden income and incentive changes due to unclear algorithms,what forms of transparency are needed to make these systems fair?

Algorithmic management has replaced human supervisors in the gig economy, but it has done so without accountability or transparency. Workers wake up one morning to find their per-order rates reduced, incentives slashed, or working zones altered without any notice or explanation. This opacity fuels insecurity and mistrust. As part of the union, we feel, to create fairness, algorithmic transparency is essential in at least three dimensions:

a. Transparency in pay determination: Workers must have access to clear information on how their earnings are calculated as in what portion goes to the platform, what deductions are made, and how dynamic pricing or surge rates apply. Currently, workers have no visibility into the "black box" of fare computation. A driver may complete ten trips under similar conditions and still find different payments for each, with no explanation. We demand that platforms publish a standardized and comprehensible pay formula accessible within worker apps.

b. Transparency in performance evaluation and ratings: Ratings are used to control workers' behaviour and determine access to jobs, incentives, and even continued employment. Yet these systems often amplify customer bias or technical glitches. A single low rating, possibly from a customer with unrealistic expectations, can damage a worker's livelihood. Platforms must disclose how ratings are weighted, provide workers the right to contest unfair reviews, and ensure that no automated decision like "deactivation" is taken without human oversight.

c. Transparency in work allocation and disciplinary actions: The allocation of orders or rides, as well as decisions on suspension or termination, should be subject to clear, appealable rules. Currently, workers are often "logged out" or "temporarily blocked" without explanation, effectively fired by an algorithm. Fairness demands that such decisions be transparent, accompanied by written notice and an opportunity to appeal through a grievance redressal mechanism. At IFAT and TGPWU, we advocate for Algorithmic Accountability Guidelines as part of India's digital labour regulations. Platforms must be required to disclose the logic behind automated decisions, ensure fairness audits, and involve worker representatives in the design of incentive systems. Without algorithmic transparency, the promise of technology-driven empowerment will remain hollow.

"

Platforms must disclose how ratings are weighted, provide workers the right to contest unfair reviews, and ensure that no automated decision like "deactivation" is taken without human oversight.

The Platform Question: Power, Accountability and Global South

From your union's experience, how responsive are platforms when workers demand accountability, and what tactics have proven most effective in pushing back?

Our experience shows that platforms are generally unresponsive and defensive when it comes to worker demands. They engage with government agencies and investors readily, but not with the workers who make their profits possible. Most grievances, whether about unfair deactivation, delayed payments, or incentive cutsare handled through automated chatbots or customer service centers that provide no meaningful resolution. However, through years of organizing, we have learned that collective pressure works, even in the digital economy. Platforms may ignore individuals, but they cannot ignore thousands of voices united under a common demand.

The Telangana Gig and Platform Workers Union (TGPWU) has successfully used a mix of strategies:

- Public campaigns and protests: Peaceful protests, press conferences, and social media campaigns such as #SelfieWithSeatBelt and #RejectUnfairIncentives #MakeAmazonPay #Driverlifematters #NoAC Campaign #BoyCottAirport #LowFareNoAir etc. have drawn public and media attention to worker issues, forcing platforms to respond.
- Engagement with government departments: By bringing evidence and testimonies directly to the Labour Department, Transport Department, and now to legislative committees, we have ensured that worker concerns reach policymakers who can hold platforms accountable.
- Legal advocacy: Through IFAT, we have filed petitions before various High Courts and the Supreme Court seeking recognition of gig workers as employees and inclusion under social security laws. Litigation has become an important tool to challenge corporate impunity.
- Solidarity and education: We invest heavily in building unity among drivers and delivery workers through training sessions, WhatsApp groups, and district-level meetings. An informed worker base is harder to exploit. Platforms often respond only when public or regulatory pressure mounts. For example, after sustained advocacy, Ola and Uber agreed to introduce in-app emergency buttons and basic accident insurance—small but meaningful victories achieved through collective struggle. Yet, the overall responsiveness remains minimal. Hence, we are now focusing on institutional mechanisms, such as mandatory representation of worker unions in state-level gig worker welfare boards, to ensure ongoing dialogue.

> "Our experience shows that platforms are generally unresponsive and defensive when it comes to worker demands. They engage with government agencies and investors readily, but not with the workers who make their profits possible.

The Platform Question: Power, Accountability and Global South

Do existing labour codes and government regulations adequately protect gig and platform workers, or are urgent policy reforms needed?

India's new Code on Social Security, 2020 was a step in the right direction, it recognizes "gig workers" and "platform workers" as distinct categories for the first time. However, the Code's provisions remain vague and unenforceable. It does not grant gig workers the status of employees or ensure binding employer contributions to social security. Instead, it creates a voluntary framework dependent on the goodwill of platforms and the discretion of the government.

The Code envisages a Social Security Fund financed through a 1-2% contribution from aggregators' annual turnover, but the rules for collection, management, and distribution are still pending implementation in most states. Without a functioning mechanism, the recognition of gig and Platform workers remains symbolic.

We need a comprehensive legislative framework that moves beyond token acknowledgment and guarantees concrete rights. Key reforms should include:

1. Employee Status or Dependent Worker Category: Recognize gig workers as "dependent contractors" entitled to minimum wages, social security, and the right to unionize.

2. Mandatory Social Security Contributions: Enforce a joint contribution model where platforms contribute a fixed percentage toward workers' provident fund, insurance, and pensions.

> "We need a comprehensive legislative framework that moves beyond token acknowledgment and guarantees concrete rights.

3. Fair Contracts: Standardize terms of service to prevent unilateral changes by platforms and include clauses for dispute resolution.

4. Representation and Grievance Boards: Establish state-level Gig and Platform Worker Welfare Boards with elected worker representatives to oversee implementation of welfare schemes.

5. Data Transparency and Worker Rights: Recognize data generated by workers as a form of labour and protect their digital rights, including access to performance and earnings data.

Some states, like Rajasthan and Karnataka, have already proposed progressive bills to regulate platform work and ensure social security coverage. Telangana has also initiated consultations with stakeholders, including TGPWU and IFAT, toward drafting a policy framework. These are encouraging developments, but national-level implementation remains essential to ensure uniformity and prevent exploitation across state borders.

Platforms often claim about creating jobs and opportunities. But when you look at the everyday lived realities of gig workers in Telangana and India, does that narrative really hold up?

That is correct, platform companies are often found claiming that they are "creating jobs" and "empowering youth with flexible opportunities." On the surface, this narrative sounds appealing, especially in a country where unemployment and underemployment are chronic issues. However, the lived reality of gig workers in Telangana and across India tells a very different story. When platforms first entered the Indian market around 2013–2015, they offered attractive incentives and high earnings to rapidly expand their workforce. Many drivers and delivery partners took loans to purchase vehicles, believing they were entering a path of upward mobility. Initially, monthly incomes ranged between Rs.40,000– Rs.70,000, enough to support families and repay debts. But as platforms achieved market dominance, incentive structures were systematically reduced. Today, after deducting fuel costs, commissions, and

"When platforms first entered the Indian market around 2013–2015, they offered attractive incentives and high earnings to rapidly expand their workforce. Many drivers and delivery partners took loans to purchase vehicles, believing they were entering a path of upward mobility. Initially, monthly incomes ranged between Rs.40,000–Rs.70,000, enough to support families and repay debts. But as platforms achieved market dominance, incentive structures were systematically reduced

maintenance, most drivers take home less than Rs. 15,000–Rs.20,000 per month, often below minimum wage levels. Flexibility, another major selling point, has also proven illusory. In practice, workers are compelled to work 12–14 hours a day to earn a subsistence income. Algorithms penalize inactivity or trip rejection, pushing workers into a state of digital servitude. The supposed freedom of "choosing your hours" has turned into the compulsion of "working endlessly to survive." In Telangana, we see these contradictions daily. Many gig and Platform workers struggle with loan defaults, mental stress, and health issues due to long working hours and lack of social protection. During the COVID-19 pandemic, platforms abandoned workers almost overnight, no earnings, no safety equipment, and no support. It was only through the collective efforts of unions and civil society that emergency relief reached some of them.

This experience exposed the myth of platform benevolence. These are not job-creation platforms, in fact they are digital intermediaries that extract labour value without employment responsibility. The real contribution of platforms lies not in generating secure livelihoods but in making labour into an on-demand commodity. That said, gig work can indeed offer opportunities if governed fairly. Digital platforms have the potential to connect workers to demand efficiently, reduce barriers to entry, and promote entrepreneurship. But this potential will only be realized when the rules of the game are fair, and when workers share in the economic value they create. For that, India will have to adopt a worker-centered vision of the platform economy, one that prioritizes rights, dignity, and sustainability over profit maximization of some.

"If you could sum up the message that gig workers in India are sending to governments and platforms today, what would it be?"

The struggle of gig and platform workers in India is not only about wages or benefits, but it is about recognition and justice. It is about questioning the nature of employment in the digital age. The platforms' claim of being "technology companies" rather than "employers" cannot be used to erase fundamental labour rights. Our unions, TGPWU in Telangana

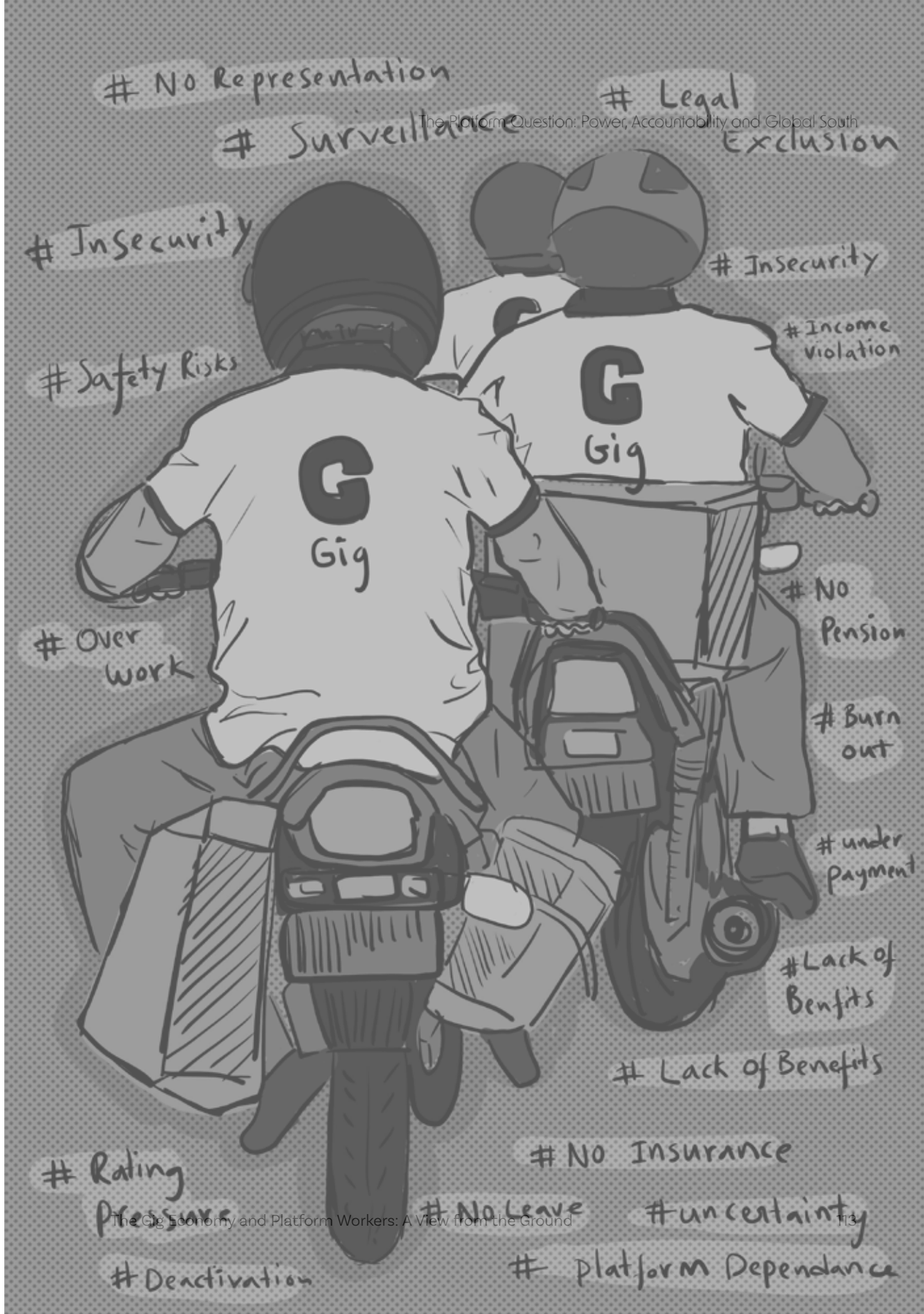
"Algorithms penalize inactivity or trip rejection, pushing workers into a state of digital servitude. The supposed freedom of "choosing your hours" has turned into the compulsion of "working endlessly to survive."

and IFAT nationally, have consistently argued that digital labour is still labour. Whether work is mediated by an app or a factory supervisor, the principles of fairness, safety, and social protection must remain the same. We are calling for a new social contract in the digital economy, one does not compromise basic rights under the garb of innovation. Governments must act decisively at the right moment, platforms must engage transparently without exploitation, and society must acknowledge that behind every ride or delivery is a human being whose work and dignity deserves equal respect. Until that happens, our organizing will continue. Because the future of work must not come at the cost of any worker.

“

Whether work is mediated by an app or a factory supervisor, the principles of fairness, safety, and social protection must remain the same. We are calling for a new social contract in the digital economy, one does not compromise basic rights under the garb of innovation.

The Platform Question: Power, Accountability and Global South



The Gig Economy and Platform Workers: A View from the Ground

CHAPTER 8

When Power Meets Platform: Zuckerberg's Decision and the Implications for the Global Majority

By Dr. Arpita Kanjilal

Introduction: Platforms and the Architecture of Truth

In January 2025, Meta Platforms, the parent company of Facebook, Instagram, and Threads, announced the termination of its United States third-party fact-checking programme and introduced a "Community Notes" model that relies on user-generated contextualisation instead of professional verification[1]. While currently limited to the United States, Meta stated its intent to extend this approach globally.[2] This decision marks a significant moment in the evolution of platform governance, as it shifts the locus of fact-checking from vetted experts to the general public. Fact-checking mechanisms are essential intermediaries when it comes to stopping large scale misinformation or disinformation. Their removal, even in limited jurisdictions, signals a recalibration of how truth itself is moderated and contested online.

To cite a scenario in the context of rural India and DEF's work with SoochnaPreneurs,[3] Salma*[4], a SoochnaPreneur in her village, watched an official-looking reel, which went viral in her community that said, "Earn Rs. 3000 – Last Day – Pay Rs. 299 now." After a few hours, a False label in Hindi was added by Meta's fact-checkers, which helped Salma identify the post as a deceptive scam post designed to extort money. As an immediate measure, she took the screenshot of the label on the reel, recorded a voice note and circulated it in her community's WhatsApp groups. Although a few villagers had become victims of this scam, Meta's labelling and Salma's action prevented many more from losing money. It is important to note here that Salma has the digital and information literacy to identify and interpret labels, without which, the labelling alone would not be enough to protect end users, especially those from vulnerable, marginalised, and underserved communities.

"Fact-checking mechanisms are essential intermediaries when it comes to stopping large scale misinformation or disinformation. Their removal, even in limited jurisdictions, signals a recalibration of how truth itself is moderated and contested online.

Power, Privilege, and the Unchecked Narrative

The withdrawal of institutional fact-checking mechanisms disproportionately benefits those already equipped with financial, political, or technological resources. Politicians, corporations, and other well-funded actors possess advanced capacities to craft and disseminate persuasive messaging. Without structured verification, such groups often amplify tailored disinformation, manipulate digital echo chambers, and dominate algorithmic attention economies.[5]

False narratives, particularly those strategically engineered for virality, can entirely reshape elections, influence public opinion, and direct policy outcomes. To take a popular example, the deepfake videos of Bollywood actors Aamir Khan and Ranveer Singh went viral during the 2024 general elections. Such instances, not only negatively influence electoral integrity and they also undermine the public's right to an informed vote.[6] In such cases, the digital space becomes more than a marketplace of ideas and becomes a platform where those with greater capital and resources can purchase and preserve narrative dominance. Such in built fact- checking mechanisms can counter these practices, and can literally save democracies of a country.

"

False narratives, particularly those strategically engineered for virality, can entirely reshape elections, influence public opinion, and direct policy outcomes.

Marginalisation and Digital Inequality

The erosion of fact-checking disproportionately harms communities already underrepresented in digital spaces. Marginalised populations, including rural, low-income, and digitally underserved communities, often lack the resources and literacy to navigate disinformation. Without reliable verification mechanisms, these communities are more vulnerable to misleading content that perpetuates stereotypes, undermines trust in institutions, or spreads false health and welfare information.[7]

While the Government's Fact-Checking Unit (FCU)[8] serves as an important institutional mechanism for regulation and rapid responses to fake news and misleading information, the scale of the problem, in the context of India's social, cultural, linguistic and demographic diversity, makes a centralised system inadequate. Social, digital and information divides mean those affected do not know how to access these platforms or any other reporting mechanisms. In such cases, a one-size-fits-all approach will not work. To tackle this deep-rooted problem, a hyperlocal assessment, and a decentralised approach to tackling misinformation, fake news, and disinformation through trusted village-level intermediaries is quintessential, as proven by the Digital Empowerment Foundation's (DEF) SoochnaPreneur model.[9]

Through the SoochnaPreneur programme, DEF seeks to mitigate the adverse effects of the rural information divide, given the heavy reliance on social media for health, financial, and civic information. Leveraging a nationwide network of 2,400 SoochnaPreneurs, DEF trains community members in information verification and the dissemination of reliable, accurate and trustworthy information. As Salma does in her village, SoochnaPreneurs act as rural fact-checkers and trusted information intermediaries, embodying a bottom-up model of information equity rooted in local social and cultural realities.

However, when global platforms reduce systemic oversight, initiatives like SoochnaPreneur face greater challenges. The withdrawal of platform-level fact-checking shifts the burden of verification onto individuals and grassroots organisations, often without commensurate resources.

"

When global platforms reduce systemic oversight, initiatives like SoochnaPreneur face greater challenges. The withdrawal of platform-level fact-checking shifts the burden of verification onto individuals and grassroots organisations, often without commensurate resources.

Converging Interests: Big Tech and Governance

Some analysts interpret Mark Zuckerberg's decision and its timing, coinciding with a politically conservative U.S. administration, as a strategic recalibration of Meta's relationship with administrative power.[10] While such interpretations remain contested, they raise legitimate concerns about the alignment of private technological authority with governmental agendas.

When the interests of Big Tech and governance converge, the risk of regulatory capture and erosion of public accountability increases. This convergence of interests undermines democratic checks and balances, as the informational ecosystem becomes shaped by private priorities rather than the public good.

Global Governance and the UN Response

The decision has provoked widespread concern from digital policy advocates and United Nations experts, who warn that weakening fact-checking will exacerbate the global "infodemic" of misinformation, disinformation, and hate speech.[11] Without structured verification, vulnerable communities struggle to distinguish credible content from harmful narratives, deepening social divides and undermining trust in public institutions.

UN officials have emphasised that digital platforms hold not only commercial interests but also human rights responsibilities. The removal of fact-checking, they argue, constitutes a regression in global efforts to ensure information integrity online.[12] Research indicates that disinformation disproportionately targets and harms marginalised communities, particularly women, ethnic minorities, and low-income populations.[13] In regions with pre-existing ethnic or religious tensions, unchecked online narratives can escalate into real-world harm, including discrimination and violence.[14]

> "The privileged, with social, financial and cultural capital, retain the means to shape and steer narratives, while the marginalised are left to defend their realities in increasingly hostile information environments without matching safeguards.

Conclusion: Developing our Digital Commons

Meta's move away from structured fact-checking represents more than a procedural change and signals a reconfiguration of epistemic authority in the digital public sphere. By replacing independent oversight with crowdsourced moderation, Big Tech recentralises power without any accountability. The privileged, with social, financial and cultural capital, retain the means to shape and steer narratives, while the marginalised are left to defend their realities in increasingly hostile information environments without matching safeguards.

In India's diverse, multilingual context, and the graded ladder of socio-economic hierarchies that shape who has access, agency and autonomy in today's digital ecosystem, community-centred grassroots initiatives and rural fact-checkers programmes illustrate an alternative vision, one in which community trust, local leadership and contextualised knowledge play key roles in countering socio-culturally rooted fake news, misinformation, and disinformation. Yet, these local solutions cannot fully substitute for systemic accountability, as protecting informational equity is a shared responsibility among governments, corporations and civil society, not an onus shifted onto citizens and communities already short of resources, infrastructure and capital.

The Platform Question: Power, Accountability and Global South

Endnotes:

[1] Meta Platforms, "More Speech and Fewer Mistakes," Meta Newsroom, January 7, 2025. https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes

[2] Meta Platforms, "Testing Begins for Community Notes on Facebook, Instagram and Threads," Meta Newsroom, March 13, 2025. https://about.fb.com/news/2025/03/testing-begins-community-notes-facebook-instagram-threads

[3] Launched in 2016, the SoochnaPreneur model of digital development is a community-driven, hyperlocal approach designed to bridge the digital divide, particularly in the last mile. Central to the model are local women social entrepreneurs in rural, marginalised, and underserved regions, called SoochnaPreneurs ("Information Entrepreneurs"), who are equipped with digital tools, skills, and infrastructure to enable access to information, literacy, and resources within their communities. The model focuses on training and empowering rural women to become digital entrepreneurs, strengthening a sense of ownership and leadership within the community. https://www.defindia.org/wp-content/uploads/2025/10/SoochnaPreneur-Model.pdf

[4] Name has been changed for the purpose of anonymity.

[5] "Meta Ends Fact-Checking on Facebook and Instagram in Favour of Community Notes," The Verge, January 7, 2025. https://www.theverge.com/2025/1/7/24338062/facebook-instagram-threads-meta-abandon-fact-checking

[6] "Deepfakes of Bollywood stars spark worries of AI meddling in India election," Reuters, April 22, 2024. https://www.reuters.com/world/india/deepfakes-bollywood-stars-spark-worries-ai-meddling-india-election-2024-04-22/

[7] "Ending Fact-Checking on Social Media Fuels Hate Speech and Harassment, Experts Warn," El País, January 10, 2025. https://english.elpais.com/technology/2025-01-10/ending-fact-checking-on-social-media-fuels-hate-speech-and-harassment-experts-warn.html

[8] "Ministry of I&B to fight misinformation, fake news with fact-checking chat-

> "These local solutions cannot fully substitute for systemic accountability, as protecting informational equity is a shared responsibility among governments, corporations and civil society, not an onus shifted onto citizens and communities already short of resources, infrastructure and capital.

bot," The New Indian Express, October 27, 2025.
https://www.newindianexpress.com/nation/2025/Oct/27/ministry-of-ib-to-fight-misinformation-fake-news-with-fact-checking

[9] "Rural Fact-Checkers for Community: Narrative Report 2024-25"
https://www.defindia.org/wp-content/uploads/2025/05/Rural-Fact-Checkers-for-Community_Narrative-Report-1.pdf

[10] Robert Booth, "Meta to Get Rid of Fact-Checkers and Recommend More Political Content," The Guardian, January 7, 2025.
https://www.theguardian.com/technology/2025/jan/07/meta-facebook-instagram-threads-mark-zuckerberg-remove-fact-checkers-recommend-political-content

[11] "'Real-World Harm' If Meta Ends Fact-Checks, Global Network Warns," Arab News (AFP), January 10, 2025.
https://www.arabnews.com/node/2585921/media

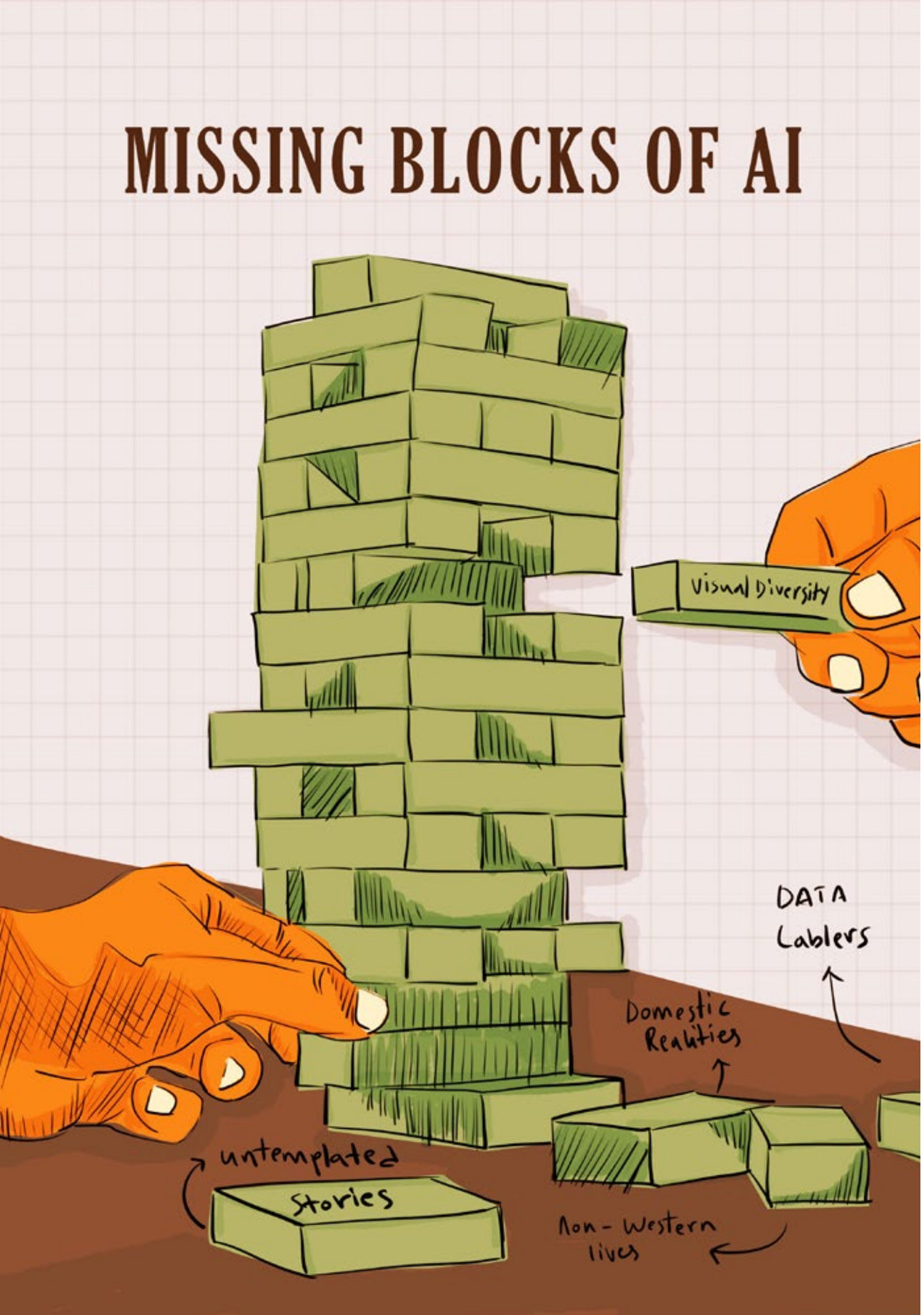
[12] UN human rights chief 'deeply worried by fundamental shift' in US, The Guardian, March 3, 2025.
https://www.theguardian.com/law/2025/mar/03/un-human-rights-chief-deeply-worried-fundamental-shift-us

[13] "When power meets platform: Zuckerberg's fact-check ban and its impact on India's rural digital landscape," Maktoob Media, January 20, 2025.
https://maktoobmedia.com/opinion/when-power-meets-platform-zuckerbergs-fact-check-ban-and-its-impact-on-indias-rural-digital-landscape/

[14] "Online speech and communal conflict: Evidence from India", PNAS Nexus, Volume 4, Issue 5, May 2025, pg.149, https://doi.org/10.1093/pnasnexus/

The Platform Question: Power, Accountability and Global South

When Power Meets Platform



CHAPTER 9

The Missing Blocks of AI: A Feminist Reimagination, The Story That Didn't Fit

By Shohini Banerjee and Vaishali Soni

In Lucknow, a young girl wanted to tell a story using Canva. As she explored the visuals available to her, she realized there were no images of kitchens that resembled her own. There were modern countertops, chimneys, and stainless steel fixtures, but not the wooden cupboards, stone slabs, or local utensils she saw every day. Her experiences did not fit into the 'templates' Canva offered; templates largely designed by, and through a Western, linear, and global-North lens. Canva is just one example, but the issue runs much deeper.

Artificial intelligence (AI) comes with the promise that it can generate anything we ask, transcending borders and boundaries and representing an array of realities and imaginations. But there's a catch: it knows only what it has been fed. And what it has been fed is not the full range of human life, but a narrow, dominant slice of it. This is how bias becomes embedded. This is how many stories, languages, and experiences get left out. At its core, AI is a system that compresses human life into data, sorts it, and then uses that data to make decisions through algorithmic logic. The process starts with massive datasets scraped from our digital traces, which are then labeled by humans, often low-paid "invisible workers"[1], who are given the parameters of labeling so that the machine learning model can "learn" to create the desired output. But this technology, supposedly objective, neutral, and efficient, often excludes those who do not fit the dominant narrative, both online, in who gets represented, and offline, through who is creating.

As Chun (2024) notes, AI consolidates stereotypes and homogenised discourses about what is, in reality, a hugely diverse and heterogeneous social world. This creates systems that not only misunderstand, but actively harm those who live outside dominant norms. ***The Missing Blocks of AI*** delves

> "AI consolidates stereotypes and homogenised discourses about what is, in reality, a hugely diverse and heterogeneous social world.

into this process and how it creates the erasure of stories, knowledge systems, and representation caused by systemic biases, and how these absences are replicated and amplified in the digital world.

Data, data everywhere but whose stories are told?

Type in "beautiful bodies" on Google, ChatGPT, Gemini or Midjourney and you will see mostly homogenous, eurocentric, slim bodies. These outputs reflect a system built on data about humans, but only the dominant beauty standards that see whiteness, thinness, non-disabled, heteronormative as desirable. When identified, this bias is often presented as an anomaly or oversight. However, AI, often assumed to have sociotechnical blindness (Johnson and Verdicchio, 2017), is a system "shaping and shaped by what is social" (Çetin, 2021). And one of its foundational elements, training data[2], is not just a technical input but is inherently social, created and shaped by experiences and narratives that are recorded, especially in large volumes. Crawford (2021) cites the example of National Institute of Standards and Technology (NIST), which consists of mug shots of arrested individuals from the United States, as a popular training dataset for image-related AI models. The context of when these images were taken or the lack of consent of the individuals (or their families) to be included in the database becomes irrelevant; individual experiences become reduced to datapoints in the form of an image which are to be extracted, circulated and repetitively used. Stories of lived experiences, therefore, get flattened into data.

"

Historically, the process of data collection, who gets counted, and what the counting signifies has always been deeply political. The "outcast, minoritised bodies, and subaltern" are often excluded or its inclusion minimised against the dominant colonial narrative.

Moreover, historically, the process of data collection, who gets counted, and what the counting signifies has always been deeply political. The "outcast, minoritised bodies, and subaltern" are often excluded or its inclusion minimised against the dominant colonial (Çetin, 2021), patriarchal (D'Ignazio and Klein 2020) and white (Noble, 2018) narrative. For example, during colonial times in India, nomadic people were labeled as criminal tribes if they did not fit into certain parameters and were recorded in public databases as such (Kovacs, 2020). In more recent times, Ruha Benjamin (2019) highlights how crime data, a public record, reflects the skew of over-policing in Black and Brown communities. As a result,

The Platform Question: Power, Accountability and Global South

foundational training datasets in machine learning (ML), often taken from such public records or scraped from the internet, become unneutral archives. Furthermore, in the Indian context, shared or surveilled phones of young women or the digitally excluded, due to the caste, socio-economic, or other barriers, leave minimal digital footprint; being absent in datasets used to train ML models altogether.

These create models that reflect social biases, like in the case of Amazon's hiring bias[3] or predictive policing tools over-policing in communities of colour (Ruha Benjamin, ***Race After Technology).*** The bias can also be compounded, as demonstrated by Joy Buolamwini and Timnit Gebru (2018) analysis of facial recognition algorithms and datasets. Their findings showed that consistently, the highest error rate was among dark-skinned women, demonstrating both a gender and racial bias. When these already biased sets are applied in the Global South context, it becomes even more exclusionary, missing out the sociopolitical and cultural lived realities of marginalised communities. To therefore have "beautiful bodies" reflect queer, Brown, Dalit, disabled bodies in AI systems, additional modifiers and labels have to be included. Type these in another language, and the results may be sexualised or irrelevant images. This reinforces the notion that marginalised bodies are the "other", deviations from the norm or the default representation, where one has to constantly assert and insert themselves into systems not inclusively designed for them.

The Data Doesn't Label Itself

Bias is not limited to the data that is being used. While training data determines what AI systems learn, annotations give that data its meaning by shaping how it is interpreted through classification. The data is given labels, defining what is and what isn't. This binarisation further reduces complex, nuanced realities to fit clean boxes of labeling. Consider gender. Facebook at one point had many choices of genders; however, on the backend, these identities were collapsed into the binary of male/female for Facebook advertisers (Beltrán & Ranganathan 2020). Therefore, classification systems, or the ways that people are sorted, are systems of power, determined and shaped by the decisions of those designing the AI (D'Ignazio and Klein; Wajcman and Young 2023). Regardless of bandaid frontend solutions that seem to include diverse

"

When these already biased sets are applied in the Global South context, it becomes even more exclusionary, missing out the sociopolitical and cultural lived realities of marginalised communities.

experiences, the backend repeats existing, often reductive, classifications. The consequences of this narrow perspective in classification have implications across the board, like the ability of a young woman on social media being able to report an image of herself being posted online without her consent. When annotators label what is considered "sexual content," they may focus solely on nudity. But in contexts like India, intimacy or what is sexual does not need to be in the form of nudity, but can be of two people holding hands in public.

These nuances are lost when the annotation logic is framed and designed through a Western lens and then becomes embedded into the functioning of AI systems. This is even more concerning when "some AI scientists have stated their desire to capture the world and to supersede other forms of knowing" (Crawford 2021). When selective stories are compressed into data, and that data is shaped by certain biases which produce AI outputs of "beautiful bodies" that users come to accept as reality, alternative stories and representations are effectively overwritten.

Annotators are provided parameters and guidelines for labeling by developers. Studies have shown that annotators' of socio-demographic backgrounds and lived experiences impact how they label.

The data is then labeled by low-wage workers from the Global South who determine what is "beautiful, and which "bodies" count. While the labour is global, the power to determine or influence is not, because those who monopolise resources monopolise imagination." (Benjamin, 2019). Annotators are provided parameters and guidelines for labeling by developers. Studies have shown that annotators' of socio-demographic backgrounds and lived experiences impact how they label. For example, in a study of large language models, researchers found that female annotators were more likely to label texts related to gender, religion, or cultural insensitivity as hateful than male annotators (Das et al, 2024).

However, when speaking with AI/ML practitioners, annotators were generally treated as "an apparatus for achieving a representation of the world" (Kapania et al, 2023). The subjective nature of the annotator would result in noting disagreement and labeling ambiguous content on the basis of their interpretations, informed by their own assumptions and biases. In a model, any disagreements and inconsistencies by an annotator are averaged across others and against a "golden data set".[4] The annotator is treated as an objective tool who would simply need to follow the parameters provided to them. Yet, like the example of ImageNet, where bias was attempted to be neutralised by crowdsourcing annotations from Amazon Turk workers' (Deng et al, 2009), the error rate was 6% (Northcutt et al, 2021), demonstrating that subjectivity matters even if notations were averaged out.

The Sociotechnical Machine

If we begin with the understanding that neither data nor bodies are neutral, then we must also accept that AI is not an abstract force - it is built, shaped, and sustained by human choices. However, if social bias continues to be treated as a bug and not a feature, as argued by several feminist authors, the redressal mechanisms will often be piecemeal and limited to one model or narrow benchmarks or 'fairness' or 'transparency'. Take, for example, Responsible AI. Rather than a universal framework, it is an overarching term that encompasses the intention of differing companies. Implementation varies as they are principles rather than accountability mechanisms. More importantly, existing within the same structures, assumptions, and biases of the architects of major AI systems, it cannot be transformative in changing the way that AI can be developed. Bias and unequal power structures are built into the system, so "how can one be "fair" in an unjust society?" (Hampton 2021).

Therefore, there is a need to interrogate and go beyond frameworks like Responsible AI. To include the lived worlds of the structurally marginalised people, questions around consent and ownership of data, performance and exploitation of labour, and co-opting social justice or feminist ideologies have to be addressed. Feminist AI principles remind us that the design of technology is never separate from systems of power (Costanza-Chock, 2023). They ask us to center care, context, and co-creation; to ask not just what data is collected, but who gets to define it, label it, and benefit from it. This is where imagination becomes essential. As Ruha Benjamin (2024) urges, imagination is a tool of resistance, especially in the face of systems designed to exclude, flatten, or erase. And so reclaiming the right to imagine, especially for those whose stories have long been left out of technological futures, is a deeply political act.

To include the lived worlds of the structurally marginalised people, questions around consent and ownership of data, performance and exploitation of labour, and co-opting social justice or feminist ideologies have to be addressed.

Reimagining AI from the Ground Up

Our project embraced this politics of imagination. Rooted in feminist and design justice frameworks, we facilitated workshops to help people create alternate 'data points' about themselves using art and storytelling as tools to reimagine the very foundations of AI, its datasets. At Hyderabad's Digital Citizen Summit 2024, we worked with tech makers, engineers, students, and professors. To reimagine what datasets could look like, we first needed to understand what the participants knew of the current AI landscape and its impact on human realities. This framing helped participants recognise AI not only as a future-forward innovation but also as an embedded decision-making infrastructure shaping the representational landscape of the web.

One participant shared a deeply personal story. He had worked on an AI-driven medical tool designed to predict emergencies. But the training data lacked female-specific health indicators, a gap that only became clear when a close acquaintance was diagnosed with breast cancer. The absence of women in the design team meant no one had flagged critical missing data. This powerful moment reiterated our point that missing data reflects missing people, missing perspectives, and ultimately, missed possibilities.

At the AWID Forum 2024, in Bangkok, conversations widened further. Here, our participants included artists, activists, human rights defenders, sex workers' rights groups, policymakers, and digital rights advocates. Despite the geographic and professional diversity, one theme persisted: most participants still associated AI primarily with generative tools. This unfamiliarity opened up space for reflection around ***What would your data look like, feel like, sound like? What would it resist being called?*** Participants began by searching everyday prompts ***beautiful bodies, breakfast, expression,*** and reflecting on whether the search results represented their realities. Queer participants, for instance, had to enter multiple, highly specific prompts ***inclusive, feminist, non-binary bodies,*** before seeing representations of themselves. People of color and disabled participants encountered similar erasures.

Especially among participants from non-Western contexts, it revealed the dissonance between lived realities and the worlds

One participant shared a deeply personal story. He had worked on an AI-driven medical tool designed to predict emergencies. But the training data lacked female-specific health indicators, a gap that only became clear when a close acquaintance was diagnosed with breast cancer. The absence of women in the design team meant no one had flagged critical missing data.

The Platform Question: Power, Accountability and Global South

rendered legible by AI systems. Once again, it became clear dominant representations online reflect the priorities of those with the most access to and control over data infrastructures. Similar to the absence of a kitchen from Lucknow, this is a design flaw and a failure of imagination as well as a refusal to make space for the multiplicity of lives that fall outside the hegemonic dataset.

The erasure became even starker when participants translated prompts into their native languages Burmese, Amharic, Hindi, and Thai. The results often returned nothing relevant, if anything at all. This underscored an uncomfortable truth, the language of the internet and by extension, AI is English. Anything outside it becomes invisible, unindexable, unintelligible to the machine. In response, participants reimagined these datasets by generating their own. Using categories like ***breakfast, bodies, and home,*** they transformed the generic sticky notes into artifacts of resistance, using basic tools such as pens, colored pens, watercolors, and embroidery threads. Under the breakfast sub-prompt "coffee," participants added ***bunna*** and ***jebena,*** Ethiopian terms for traditional coffee, alongside ***espresso*** and ***cappuccino.*** Under ***meals,*** they introduced foods like papad (a crispy lentil flatbread), ***masko daal*** (black lentils), ***aloo gobi*** (potato and cauliflower stir-fry), and ***rabari*** (a sweet dish from India), pushing back against Eurocentric culinary defaults.

The reimaginings extended to bodies. Participants from Thailand, India, Singapore, Ethiopia, Myanmar, the US, UK, and across Africa reflected, shared, deferred, and held space for every person's stories. They brought rich, layered insights into how AI is experienced, resisted, and reimagined. They drew trans bodies, faces with African locs, annotated in Thai, Burmese, Mandarin, Hindi. These were acts of creativity, interventions, care and resistance. Each drawing, annotation, and translation carved out space for new representations, recognitions, and the lived reality of those traditionally excluded. When structural injustice is coded into AI systems, the solution then becomes to integrate feminist AI principles of care, context, and co-creation, and question who gets to define what counts as data, who classifies it, who benefits from it, and who annotates it. So that when a young girl from a small town of the Global South, wants to tell ***her story*** through images of ***her kitchen,*** and ***her dishes,*** she can do so and see her reality in the vast, vast landscape that AI is.

Dominant representations online reflect the priorities of those with the most access to and control over data infrastructures.

References

1. Beltrán, M. & Ranganathan, N. (2020). The Making of Political Ads: Classification as Distraction.
2. The Internet Democracy Project. https://internetdemocracy.in/reports/the-making-of-political-ads-classification-as-distraction/#out sourcing-the-labour-of-classification
3. Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim Code. Polity. https://courses.complex-systems-laboratory.org/system/files/Race%20After%20Technology%20-%20Ruha%20BenjaminChapter1_0.pdf
4. Benjamin, R. (2024). Imagination: A Manifesto. W. W. Norton & Company.
5. Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research 81:1–15, Conference on Fairness, Accountability, and Transparency
6. Çetin, R. B. (2021). Wisdom of not knowing and decolonial AI. Heinrich Böll Stiftung. https://www.gwi-boell.de/en/2021/02/11/wisdom-of-not-knowing-and-decolonial-AI
7. Chun, W. H. K. (2024). Discriminating Data Correlation, Neighborhoods, and the New Politics of Recognition. The MIT Press.
8. Costanza-Chock, S. (2023) Design Practices: 'Nothing About Us Without Us'. Feminist AI: Critical Perspectives on Algorithms, Data, and Intelligent Machines. (2023). United Kingdom: OUP Oxford.
9. Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.
10. D'Ignazio, C., & Klein, L. F. (2020). Data feminism. MIT Press.
11. Das, A., Zhang, Z., Hasan, N., Sarkar, S., Jamshidi, F., Bhattacharya, T., Rahgouy, M., Raychawdhary, N., Feng, D., Jain, V., Chadha, A., Sandage, M., Pope, L., Dozier, G., & Seals, Deng, J., Dong, W., Socher, R., Li, L., Li, K., and Fei-Fei, L. 2024. Investigating Annotator Bias in Large Language Models for Hate Speech Detection Accepted at NeurIPS Safe Generative AI Workshop, 2024. https://doi.org/10.48550/arXiv.2406.11109
12. Gray, J. (2020). What can feminism do for AI ethics? Medium. https://medium.com/@jograycy7/what-can-feminism-do-for-ai-ethics-b7e401889441
13. Hampton, L. M. (2021). Black Feminist Musings on Algorithmic Oppression. In Proceedings of on Fairness, Accountability, and Transparency Conference https://arxiv.org/pdf/2101.09869
14. Johnson, D. G., & Verdicchio, M. (2017). Reframing AI discourse. Minds and Machines, 27(4), 575–590. https://doi.org/10.1007/s11023-017-9417-6
15. Kapania, S., Taylor, A. S., & Wang, D. (2023). A hunt for the Snark: Annotator diversity in data practices. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23) (Article 133, pp. 1–15). Association for Computing Machinery. https://doi.org/10.1145/3544548.3580645
16. Kovacs, A. (2020). When our bodies become data, where does that leave us? Data Governance Network.
17. https://deepdives.in/when-our-bodies-become-data-where-does-that-leave-us-906674f6a969
18. Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press.
19. Northcutt, C. G., Athalye, A., & Mueller, J. (2021, November 7). Pervasive label errors in test sets destabilize machine learning benchmarks. 35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks. https://arxiv.org/abs/2103.14749
20. Rani, U., & Dhir, R.K. The Artificial Intelligence illusion: How invisible workers fuel the "automated" economy.
21. https://www.ilo.org/resource/article/artificial-intelligence-illusion-how-invisible-workers-fuel-autom ated
22. UNESCO. (2024, March 7). Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes.
23. https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regre ssive-gender-stereotypes

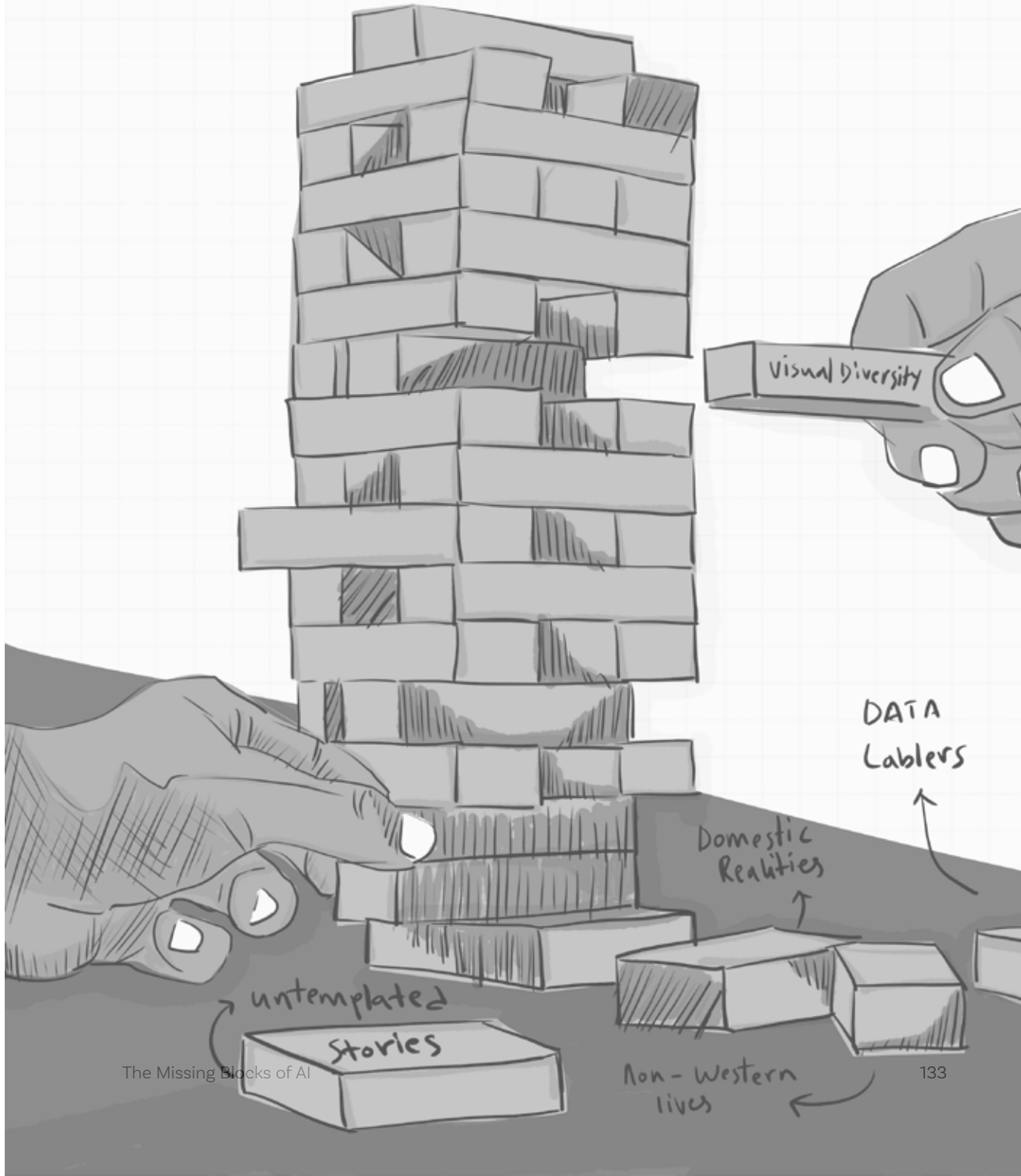
The Platform Question: Power, Accountability and Global South

24. World Economic Forum. Global Gender Gap Report 2020. World Economic Forum, 2020. https://www.weforum.org/reports/gender-gap-2020-report-100-years-pay-equality

Endnotes:

1. Rani and Dhir.(2024)
2. The initial set of texts, images, audio and other inputs used to teach a machine learning model what to recognize, how to categorize and what kind of output to produce.
3. Amazon experimented with an AI tool for hiring and found that it systematically downgraded women applicants - it was later scrapped but the data used was from Amazon's previous hiring patterns and therefore, reflected that existing bias.
4. A curated collection of human-labeled data that serves as a benchmark for evaluating the performance of AI and ML models.

MISSING BLOCKS OF AI



CHAPTER 10

Democracy Caught in A Power Struggle Between Platforms and Politics

By Juliet N. Nanfuka

A tug of war exists between global tech platforms and political actors in Africa and is increasingly shaping the continent's online civic engagement, digital democracy and political landscape. This essay explores the evolving dynamics between these powerful actors, and how the convergence of control of information and narratives is unfolding and its implications for democratic governance in Africa.

Digital platforms such as Facebook, TikTok, WhatsApp, X, and Snapchat have become vital spaces through which political parties, activists, and citizens engage with democratic processes. This is particularly pertinent in the global majority with the role that digital hyperlocal media and social networking spaces are playing especially in their capacities to empower communities by making political spaces more accessible and familiar, challenging traditional power structures, and enhancing democratic participation at the local level (Toumaras, 2025).

These platforms have come to offer unprecedented opportunities for often marginalised groups to participate in political discourse, expand civic engagement, and challenge political narratives. Often what has started off as a hyperlocal online narrative has evolved into actions offline and in some instances changes in governance. This has been witnessed in Kenya's 2024 Gen Z protests that drove the #RejectTheFinanceBill in which citizens driven by the youth demanded social justice and accountability, policy formulation, transparency and political change in the country (Nyagaka & Ongere, 2025); and the Nigeria 2021 #EndSARS movement driven by young Nigerians calling against the rogue police unit, the Special Anti-Robbery Squad (SARS) that stood accused of a slew of crimes including harassment, rape, profiling, extortion, and robbery (Bodunrin & Matsilele, 2024).

> "A tug of war exists between global tech platforms and political actors in Africa and is increasingly shaping the continent's online civic engagement, digital democracy and political landscape.

In recent years, several countries have blocked access to the internet especially during times of public protest, in many instances stating the move as a means to maintain public order or to ensure national security. However, an undercurrent has been emerging which points to a power dynamic between platforms and the state when it comes to internet shutdowns which points to the tension on who gets to manage the flow of information and online narratives.

In both instances, access to digital communication was interrupted. In Kenya, reports emerged of an internet shutdown or the throttling of internet access speeds by nearly 40 per cent (Opanda, 2025) across major networks. In May 2025, the Kenya High Court ordered the government, the Communications Authority of Kenya and relevant stakeholders not to shut down the internet. In the case of Nigeria, while the internet remained accessible, in the months that followed, the government would ban Twitter for what it termed as "double standards" and threats to "Nigeria's corporate existence" (Conroy-Krutz, 2021). Additional allegations were that the platform supported the #EndSARS movement against police brutality. However, the decision to suspend Twitter in the country followed a decision by the platform to delete a tweet by President Muhammadu Buhari in which he appeared to threaten violent retaliation against a southeastern secessionist group following attacks on government facilities and personnel (Asadu, 2021). Twitter argued the message had violated its rules against abusive behaviour. For both governments, it was exceedingly clear the role that social media platforms play in assembly, organizing and in challenging the ruling establishment.

However, the recurring pattern of network disruptions in Africa continues to be an area of concern. In recent years, several countries have blocked access to the internet especially during times of public protest, in many instances stating the move as a means to maintain public order or to ensure national security. However, an undercurrent has been emerging which points to a power dynamic between platforms and the state when it comes to internet shutdowns which points to the tension on who gets to manage the flow of information and online narratives. Ultimately, the tension emerges on who has the higher power in managing online narratives.

Instances of disruptions linked to elections or political unrest include the case of Senegal where President Macky Sall, throughout his 12-year tenure had been accused of significantly diminishing political opposition and shrinking press freedoms. In 2024, Sall attempted to have the two-term limit for presidents subverted by adjusting the constitution. Further, the electoral code was also changed to make it more difficult for opposition actors to compete in the elections. Meanwhile, an announcement to shift the elections from February 2024 to later in the year was also made. These moves resulted in

The Platform Question: Power, Accountability and Global South

discontent within the populace which was mirrored in online narratives. Consequently, in February 2024, the country faced two notable disruptions to digital communication when on February 4, 2024 the internet was shutdown before being restore a few days later, and on February 13, 2024 mobile internet access was blocked (Dione, 2024) ahead of a banned march against the postponement of the presidential election.

However, in 2023, on June 1, digital access monitor, NetBlocks noted restrictions to Facebook, Twitter, WhatsApp, Instagram, YouTube, Telegram and other social media platforms in the Senegal. The restriction later expanded to affect mobile data from June 4, 2023 and a curfew style of access was put in place for three days (2023). A month later in August 2023, Senegalese authorities suspended access to TikTok claiming that the platform was being used to disseminate falsehoods, with the Minister of Communications and the Digital Economy, Moussa Bocar Thiam, stating that, "the TikTok application is the social network of choice for ill-intentioned people to spread hateful and subversive messages threatening the stability of the country," the various incidents of shutdowns were associated with political decisions including the arrest of key opposition leader Ousmane Sonko. Senegal would eventually go on to hold its elections in March 2024, and would see Maky Sall ineligible to stand for another term. Bassirou Diomaye Faye, running in place of Ousmane Sonko, was consequently elected as president. Internet access remained accessible during this time. It was a marked as a win for democracy and digital civic participation in the country.

In May 2025, the Economic Community of West African States (ECOWAS) Court of Justice would also issue a landmark judgment in the case 'Association of Information and Communication Technology Users (ASUTIC) and Ndiaga Gueye against Republic of Senegal', declaring that Senegal's internet and social media shutdowns in 2023 were clear violations of fundamental human rights, including freedom of expression, right to access information, right to assembly and the right to work (Toussi, 2025). This ruling built on an earlier ruling by ECOWAS in 2020 which condemned internet shutdowns during anti-government protests and ordered Togo to pay a fine which is a decision which many assumed would have far-reaching implications for Francophone Africa, where digital repression has been steadily increasing (Toussi, 2025).

The various incidents of shutdowns were associated with political decisions including the arrest of key opposition leader Ousmane Sonko.

In the case of Uganda, the tension between platforms and the state has long existed. In July 2018, telecom companies in Uganda blocked access to social media platforms for all users and required them to pay a newly introduced Over-The-Top" (OTT) tax before regaining access (Nanfuka, 2018). The tax was the result of a March 2018 presidential directive for social media to be taxed to raise resources "to cope with the consequences" of social media users' "opinions, prejudices [and] insults" (2018). It was at this time that social media platforms were framed as a place of "gossip" despite the government having set up frameworks for various Ministries, Departments and Agencies to utilize social media platforms as avenues for civic engagement.

Uganda has had a practice of blocking social media platforms during elections, as was the case in the 2016 election and the in the 2021 elections. However, the latter election served as the tipping point between the state and platforms. In three days before the January 14, 2021, Facebook suspended the accounts of various government officials associated with the ruling party for what the platform described as "Coordinated Inauthentic Behaviour" which was aimed at manipulating public online debate. X (Twitter at the time) also suspended similar accounts. Facebook removed 220 personal accounts, thirty-two pages, fifty-nine groups and 139 Instagram accounts which according to the platforms had links with the media arm of the government – the Government Citizens Interaction Center (DFR Lab, 2021).

Uganda has had a practice of blocking social media platforms during elections, as was the case in the 2016 election and the in the 2021 elections. However, the latter election served as the tipping point between the state and platforms.

The government perceived the take-downs of accounts associated with the ruling party (the National Resistance Movement – NRM) as an attempt at meddling with national interests. This sparked a debate on who has the authority to curate online content, especially where government content and actors are involved. Don Wanyama (whose Facebook and Instagram accounts were also suspended), the President Museveni's press secretary, accused the platforms of trying to influence the elections and stated, "Shame on the foreign powers who think they can impose a puppet government on Uganda by disabling the online accounts of NRM supporters" (Fröhlich, 2022). Wanyama's account was among those that shared anti-opposition narrative.

The government would go on to respond by blocking access to social media platforms 48 hours before polling opened, and eventually a complete internet access block was be initiated. The election was conducted in a complete internet blackout. Internet access would start its journey back to restoration on January 18, 2021. All social media sites would be made accessible, except Facebook. As the country prepares for elections in January 2026, Facebook has remained inaccessible, unless through the use of a Virtual Private Network (VPN). The election will see the incumbent Museveni run for another term in office to add to his 40 year tenure in office.

Meanwhile, a month earlier, in December 2020 the Uganda government through the telecommunications regulator sent a letter to Alphabet (Google's holding company) in which it requested that YouTube blocks over 14 YouTube channels. The channels were primarily channels associated with the opposition as well as those of citizen journalists that had largely been publishing or live-streaming content from opposition party leader and presidential candidate Robert Kyagulanyi's (also known as Bobi Wine) campaign trail. In several instances, the campaign had turned violent with police cracking down on opposition supporters. The UCC argued that the channels violated Ugandan laws and that the continued broadcast by the channels might cause economic sabotage and compromise Uganda's national security (The Independent, 2020). The request was not accepted by Google LLC who noted that for YouTube channels to be removed, the government would have had to submit a court order. In the case of Uganda, there was a hyperfocus on disinformation by the ruling party and its associates, despite disinformation also being driven by opposition actors (Code For Africa, 2021).

What is emerging is a pattern of internet shutdowns where there is a complex tug of war between states and digital platforms. These power dynamics illustrate that the convergence of sovereignty, political control especially in authoritarian regimes, and platforms is increasingly resulting in conflicts. States implementing disruptions are increasingly intersecting with the role of tech platforms as arbiters of online discourse who some states are arguing are skewing information controls and dissemination.

What is emerging is a pattern of internet shutdowns where there is a complex tug of war between states and digital platforms. These power dynamics illustrate that the convergence of sovereignty, political control especially in authoritarian regimes, and platforms is increasingly resulting in conflicts.

While platforms can act as a force of democracy by facilitating access to information and enhancing government

accountability, as seen in the cases of Kenya and Nigeria, their misuse can also weaponise disinformation, hate speech, and cyber harassment that play into the interests of some political actors. There remain gaps in current platform governance, which struggles to balance the protection of free expression with the need to mitigate harm. This is especially so when for some IT platforms appear to be selective in what is addressed.

Ultimately, the ongoing tug of war between platforms and politics in Africa reveals broader tensions around power, control, and citizen engagement in the digital age. Strengthening platform accountability and protecting digital rights are essential to ensuring that democracy in Africa can withstand these pressures and remain vibrant, progressive and inclusive. This requires not only technical and legal innovation but also sustained advocacy and political will to foster transparent, pluralistic, and participatory digital public spheres.

Strengthening platform accountability and protecting digital rights are essential to ensuring that democracy in Africa can withstand these pressures and remain vibrant, progressive and inclusive.

References:

1. Asadu, C. (2021, June 2). Twitter deletes Buhari's 'treat them in the language they understand' tweet after outcry. Retrieved from The Cable: thecable.ng/breaking-twitter-deletes-buharis-tweet-on-dealing-with-secessionists
2. Bodunrin, I. A., & Matsilele, T. (2024). Social media and protest politics in Nigeria's #EndSARS campaign. Science Direct, 109-122. doi:https://doi.org/10.1016/B978-0-323-90237-3.00006-0
3. Code For Africa. (2021, March 26). Debunking election disinformation during Uganda's internet shutdown. Retrieved from https://medium.com/code-for-africa/debunking-election-disinformation-during-ugandas-internet-shutdown-d82f8345b634
4. Conroy-Krutz, J. (2021, July 7). Nigeria's Twitter ban could backfire, hurting the economy and democracy. Retrieved from The Conversation: https://theconversation.com/nigerias-twitter-ban-could-backfire-hurting-the-economy-and-democracy-162233
5. DFR Lab. (2021, January 12). Social media disinformation campaign targets Ugandan presidential election. Retrieved from https://medium.com/dfrlab/social-media-disinformation-campaign-targets-ugandan-presidential-election-b259dbbb1aa8
6. Dione, N. (2024, February 13). Senegal cuts internet again amid widening crackdown on dissent. Retrieved from Reuters: https://www.reuters.com/world/africa/ahead-planned-march-over-vote-delay-senegal-suspends-internet-access-2024-02-13/
7. Fröhlich, S. (2022, September 1). Dictators in Africa using social media to cling to power. Retrieved from DW: https://www.dw.com/en/dictators-in-africa-using-social-media-to-cling-to-power/a-60360543
8. Nanfuka, J. (2018, July 1). Uganda Blocks Access to Social Media, VPNs and Dating Sites as New Tax Takes Effect. Retrieved from CIPESA: https://cipesa.org/2018/07/uganda-blocks-access-to-social-media-vpns-and-dating-sites-as-new-tax-takes-effect/
9. Nanfuka, J. (2018, April 16). Uganda's Social Media Tax Threatens Internet Access, Affordability. Retrieved from CIPESA: Uganda's Social Media Tax Threatens Internet Access, Affordability
10. Netblocks. (2023, July 1). Social media restricted, mobile internet cut in Senegal amid political unrest. Retrieved from Netblocks: https://netblocks.org/reports/social-media-restricted-and-mobile-internet-cut-in-senegal-amid-political-unrest-W80QkaAK
11. Nyagaka, E. O., & Ongere, B. M. (2025). From Online and the Streets to the Corridors of Power: Gen Z Protests and the Promotion of Social Justice in Kenya. ISRG Journal of Arts, Humanities, and Social Sciences, III (IV). Retrieved from https://isrgpublishers.com/wp-content/uploads/2025/07/ISRGJAHSS1001352025.pdf
12. Opanda, C. (2025, May 14). https://www.kenyans.co.ke/news/112042-court-orders-communications-authority-kenya-not-shut-down-internet. Retrieved from https://www.kenyans.co.ke/news/112042-court-orders-communications-authority-kenya-not-shut-down-internet
13. The Independent. (2020, December 16). Google requires court order to delete YouTube channel-Africa spokesperson. Retrieved from https://www.independent.co.ug/google-requires-court-order-to-delete-youtube-channel-africa-spokesperson/

The Platform Question: Power, Accountability and Global South

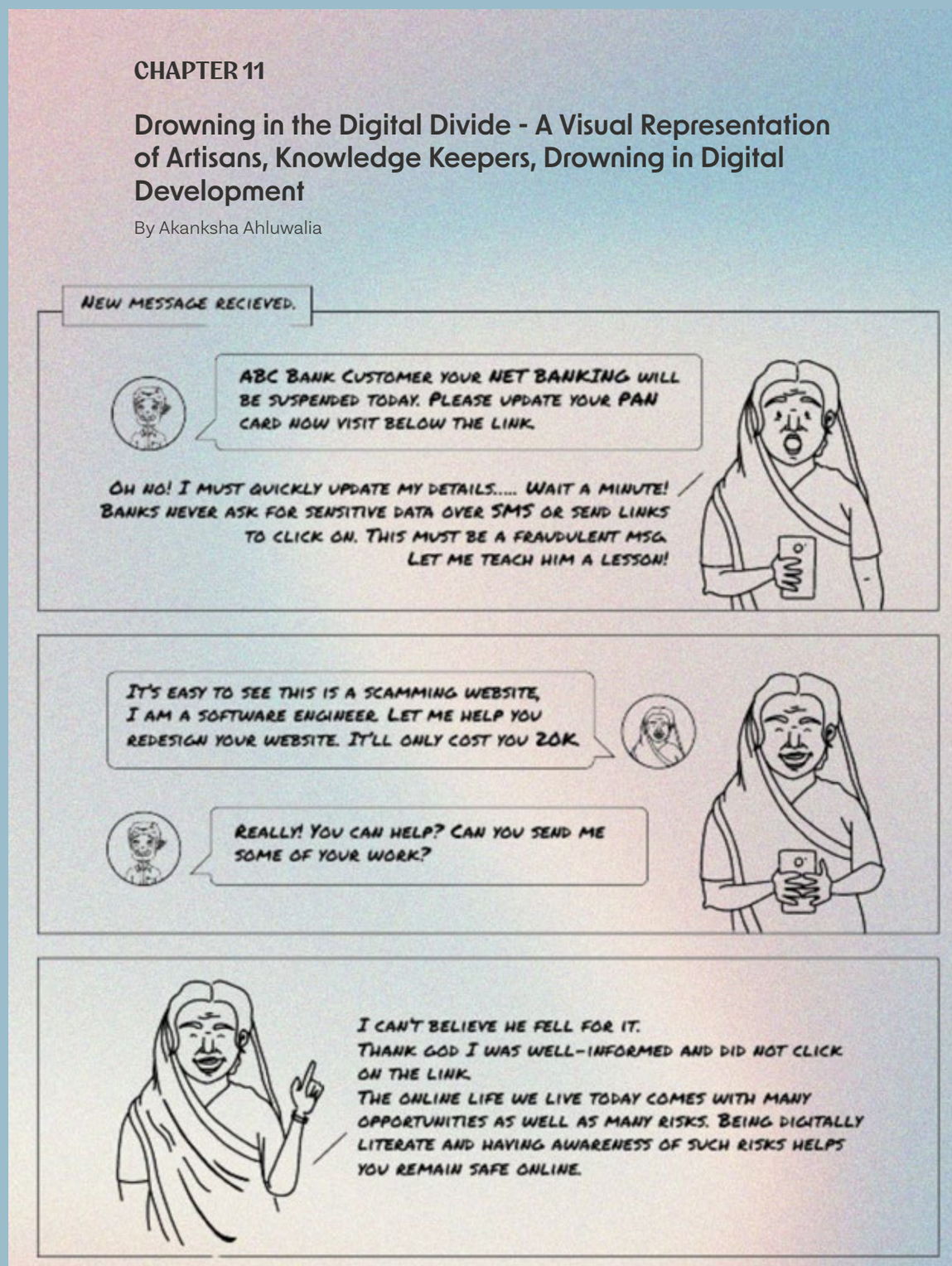
14. Toumaras, N. (2025). Networked Hyperlocal Activists: Digital Democracy and Engagement in Sub-Saharan Africa. Social Media + Society. doi:https://doi.org/10.1177/20563051251345945
15. Toussi, S. (2025, June 3). Will the ECOWAS Judgment on Senegal Redefine Digital Rights in Francophone Africa? Retrieved from CIPESA: https://cipesa.org/2025/06/will-the-ecowas-judgment-on-senegal-redefine-digital-rights-in-francophone-africa/

The Platform Question: Power, Accountability and Global South

CHAPTER 11

Drowning in the Digital Divide - A Visual Representation of Artisans, Knowledge Keepers, Drowning in Digital Development

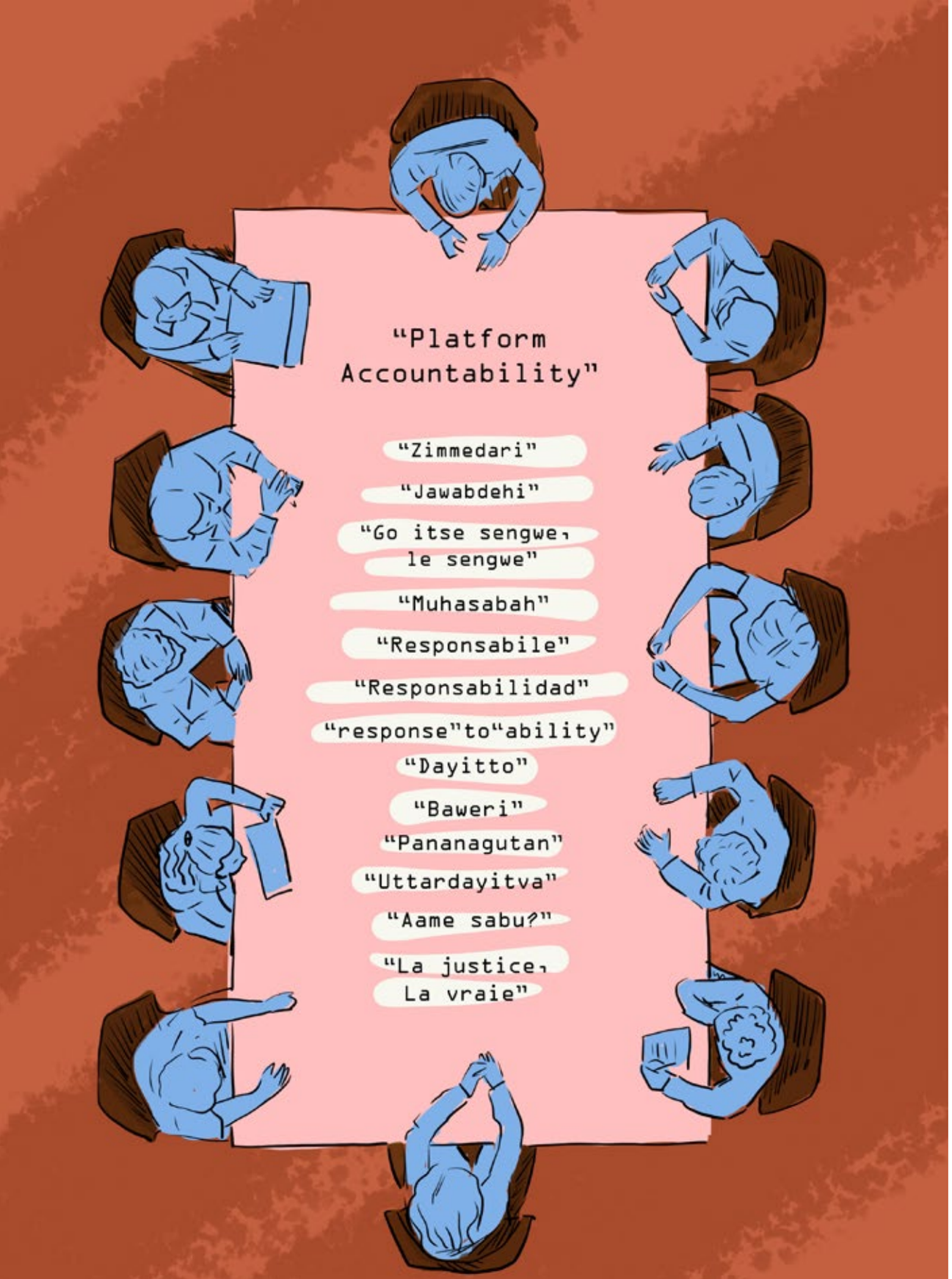
By Akanksha Ahluwalia



In India, artisans who are largely from Dalit, Bahujan, Adivasi, and other marginalised communities embody generations of traditional and cultural knowledge, yet remain invisible in digital policy, platform design, and state narratives of innovation. Their exclusion reveals how technology often amplifies privilege rather than bridging inequities. This chapter asks what accountability truly means in such a context: how can digital futures become spaces of care, recognition, and justice? It calls for reimagining technology not as a tool of extraction or spectacle, but as an ethical, inclusive ecosystem that acknowledges and uplifts the labour and knowledge of those sustaining India's artisanal traditions.


#LaBOURFORCe
#NOpOLicy
#DiGiTaLDiviDe
#accESS
#accESSGaP
#SKiLLGaP
#CONNECTivityiSSuES
#NOpOLicy
#LOWDiGiTaLLiTeRacy
#uNPaiDLaBOuR
#SCaMS
#LaNGuaGeBaRRieR
#PLaTFORM
#LOSSOFpay
#LOSTLeGacy
#FRauD
#DeSiGNCOpy
#CuLTuRaL
eROSiON


Go Online!



CHAPTER 12

Platform Accountability: A Translation Experiment with Words, Meanings and People

By Dr. Raina Ghosh

As the concluding chapter of this anthology that reflects upon the themes of platforms, power, accountability, and the Global South, this essay weaves together threads of the preceding discussions, connecting facets of regulation, labour, and algorithmic control to the very language of responsibility itself. The chapter attempts at showing how "platform accountability" transforms, travels, and resists translation across cultures, revealing new ways to imagine digital justice from the Global South.

The Global Gathering, held from September 8th to 10th in Portugal, brought together a gamut of progressive techies, activists, cyber lawyers, digital rights defenders, researchers, and civil society organisations from around the world to explore questions of digital rights, platform governance, and collective action. Organised by Team CommUNITY, the gathering created a space for the cross-pollination of ideas across scales and geographies, offering a rare opportunity for people working on similar challenges in different contexts to meet, learn, and strategise together.

At the Gathering, ARISE, supported by the DEF Secretariat, hosted a booth with a deceptively simple exercise. We asked participants passing through to write on small, colourful sticky notes - "platform accountability" in their own language, or in their mother tongue, or to think about the word closest to it. Some paused, puzzled by the challenge of translation. Others wrote quickly, confidently. A few engaged in animated discussions about whether a direct equivalent even existed in their language. Over a period of 2 hours, contributions accumulated on our table – bringing a polyphonic vista that included scripts such as Arabic, Sinhalese, Bengali, Spanish, alongside Setswana, Filipino, French, and many more.

> "The emergent collection of translations revealed something more profound for all of us to ponder - a kaleidoscope of meanings, each note reflecting a different cultural understanding of responsibility, power, and social bonds. The exercise, in fact, became a live ethnography of how communities across geographies conceptualise the relationship between those who hold power and those affected by it.

The emergent collection of translations revealed something more profound for all of us to ponder - a kaleidoscope of meanings, each note reflecting a different cultural understanding of responsibility, power, and social bonds. The exercise, in fact, became a live ethnography of how communities across geographies conceptualise the relationship between those who hold power and those affected by it. This essay is about those colourful notes, about the words, the silences, the improvisations and about the polyphonic shared experiences of how accountability is lived, not defined, across geographies and social margins.

Navigating the Linguistic Labyrinth

If we consider the term in English, "platform accountability" can refer to audits, compliance, redress mechanisms, corporate responsibility, operational transparency, algorithmic fairness, lawsuits, and standardised metrics – depending on various positionalities. As DEF Secretariat, when we ourselves started playing the exercise, translation turned out to be quite an effort - frequently getting lost into the semantic labyrinth, in search of the apt word.

Soon, ***"Zimmedari",*** a Hindi word (part of the Indo-Aryan language family) emerged as the term closest to accountability from amongst the team. In terms of meaning, it entailed a sense of moral duty and an ethical stance that conveys: "I am responsible, and I must respond". On the other hand, ***"Jawabdehi"***, again in Hindi, implied accountability as answerability - as in, someone asks, and the other responds. From Setswana's (part of the Sotho-Tswana group within the Bantu language family) ***"Go itse sengwe, le sengwe"*** (meaning transparency in information) to the simple English phrase "rooted in a system of trust." Many responses centred on a fundamental truth: accountability cannot exist without visibility and confidence. For instance, the Arabic ***"Muhasabah"*** carried connotations of self-examination and moral reckoning, suggesting that accountability begins with introspection. Italian's ***"Responsabile"*** and Spanish's ***"Responsabilidad"*** share Latin roots that bind "response" to "ability" or the capacity to answer for one's actions.

Yet, the mental churning at the booth soon made us realise that these aren't interchangeable concepts, and this aspect of untranslatability is epistemic in nature.

Yet, the mental churning at the booth soon made us realise that these aren't interchangeable concepts, and this aspect of untranslatability is epistemic in nature. It matters profoundly when we're talking about platforms that wield unprecedented power over our lives - be it in the context of Meta's control of its viewership of its content, the kind of political discourse being shaped by X, the phantasmic workings of algorithms behind TikTok setting global trends or the context of smaller regional platforms controlling local digital ecosystems. The question of platforms therefore, is of epistemic nature, helping us to understand how power gets manifested as an undercurrent through these platforms, that crosses borders and yet remains unaccountable primarily to the communities they affect. While unrelated to the domain of digital, someone wrote the phrase, "accountability runs down the family" situating responsibility within kinship structures, indicating the need for broader systemic and behavioural changes. In contrast, take the example of other the Indian languages, Bengali's ***"Dayitto"*** and Kumaoni's ***"Baweri"*** emerged from linguistic traditions that encode collective obligation differently than Western frameworks of individual liability.

But how do these culturally-specific understandings of accountability apply to a Silicon Valley corporation that operates globally? When a platform's content moderation decision harms a community in the Philippines, or its algorithm amplifies hate speech in India, or its data extraction practices undermine privacy in Kenya, /whose framework of accountability applies? The platform's corporate governance model? The user's cultural understanding of responsibility? Both? Neither?

On a broad level, terms like Filipino's ***"Pananagutan"*** and Hindi's ***"Uttardayitva"*** both carry a sense of weight of collective aspirations and outcomes. This heaviness is evident across multiple responses: accountability isn't merely about answering questions, but about shouldering the consequences. The notion of being 'committed towards a common goal' transforms accountability from a punitive mechanism into a shared objective that turns it into a mutual stake in outcomes, thereby transcending transactional relationships.

The phrase ***"Aame sabu?"*** in Odia, another Indian regional language (literally meaning "we all need to be accountable"),

On a broad level, terms like Filipino's "Pananagutan" and Hindi's "Uttardayitva" both carry a sense of weight of collective aspirations and outcomes. This heaviness is evident across multiple responses: accountability isn't merely about answering questions, but about shouldering the consequences.

repositioned accountability on a democratic plane, contrasting sharply with top-down models where platforms are responsible to users. Such a framing promises a future of a horizontal network where everyone holds responsibility for collective well-being. As one participant wrote, accountability to her refers to an exercise that aims to "weave society and social networks", emphasising the fact that digital platforms don't exist outside the social fabric but within it and are therefore, subject to the same relational ethics that govern community life. Additionally, French's *"La justice, La vraie"* (true justice) and Sinhala's reference to reparations pushed accountability beyond mere explanation toward restitution. In shorthand, accountability without remedy is hollow. This transforms the concept from a corporate governance checkbox into a mechanism for addressing historical and ongoing injustices. When platforms shape democracy, spread information (and misinformation), and mediate social relationships across billions of people, accountability must encompass repair, not just acknowledgement.

Some respondents noted there is "no specific translation" in their language, resorting to approximations or borrowed terms. This isn't a linguistic deficiency but an essential revelation that platform accountability may be a fundamentally new problem, one that existing vocabularies of responsibility weren't designed to address.

However, the most provocative question stemming from this collection is the chit that bore the words: "Accountability from/for what? Implies a response to an action/event/structure". This interrogation cuts to the temporal dimension of accountability – about how it is always retrospective, always addressing something that has already occurred or is currently occurring. It cannot be abstract; it must answer to specific harms, specific exercises of power, specific structural inequalities.

The Untranslatable Core

What's striking about this multilingual constellation is what resists translation entirely. Some respondents noted there is "no specific translation" in their language, resorting to approximations or borrowed terms. This isn't a linguistic deficiency but an essential revelation that platform accountability may be a fundamentally new problem, one that existing vocabularies of responsibility weren't designed to address. Therefore, we must acknowledge the struggle to comprehend the exact nature of the reality that the digital turn has presented to us. The sheer fact that a word like "Utopia" appears in the list – perhaps hints at the truth that "Platform accountability" as currently imagined might still be an idealistic

dream we carry within ourselves. Or maybe these diverse words offer building blocks for something new – say, a hybrid model of accountability that draws on multiple traditions, that sees platforms as simultaneously institutional and relational, technical and human, answerable to both law and community.

We need to, therefore, recognise that the untranslatability of platform accountability has practical consequences since it defies traditional concepts of accountability, as understood in terms of an individual's obligation to answer, community trust, or institutional transparency. For instance, policy frameworks designed in isolation, or in Northern regulatory contexts, may fail to resonate with marginalised users. The game of words across geographies reveals that platform accountability isn't one thing but many things, each shaped by language, history, and social organisation. Studying how people describe, feel, and negotiate platform accountability in their own languages provides insight into the gaps between policy, corporate rhetoric, and everyday life. Only then can interventions be genuinely inclusive.

We need to, therefore, recognise that the untranslatability of platform accountability has practical consequences since it defies traditional concepts of accountability, as understood in terms of an individual's obligation to answer, community trust, or institutional transparency.

CHAPTER 13

In the age of AI

By Osama manzar

Chakra View

Inspired by the concept of "chakras," ChakraView reflects on the evolution of data in the digital age, where the data extracted from individuals is used to control, categorize, and reshape human lives. Through immersive artwork and installations, the exhibition exposes how algorithms manipulate this data, often amplifying harmful traits in AI such as prejudice, exclusion, inequality, profiling, & fabrication.

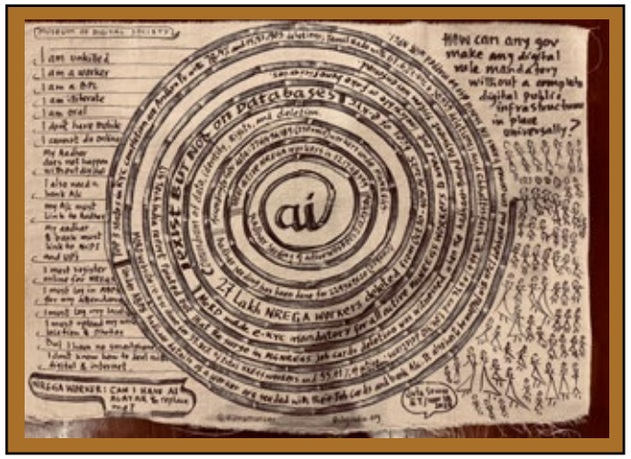
Through the lens of the continuous cycle of data extraction and algorithmic manipulation, ChakraView challenges us to critically examine a future where the very data used to control human behavior also reinforces societal inequities. The exhibition explores how data mining, algorithmic manipulation, and "data fiction" shape

> "Chakra View advocates for a future in which AI is guided by human values, transparency, and ethical accountability, ensuring that technology serves humanity rather than perpetuates harm through the very data extracted from it.

the digital landscape, urging reflection on the ethical and social consequences of AI's growing autonomy.

Chakra View advocates for a future in which AI is guided by human values, transparency, and ethical accountability, ensuring that technology serves humanity rather than perpetuates harm through the very data extracted from it.

A circle of inclusivity & Responsibility in AI



The artwork portrays the cyclical, chakra-like functioning of artificial intelligence at the grassroots level, especially among marginalized yet educated communities. It highlights how limited access to digital infrastructure leaves them excluded from systems that increasingly shape their lives. Despite this exclusion, their data continues to be extracted and processed. The piece reflects on the imbalance between digital control and the lack of digital empowerment.

> Limited access to digital infrastructure leaves communities excluded from systems that increasingly shape their lives. Despite this exclusion, their data continues to be extracted and processed.

AI for Humans or humans for AI

The artwork draws inspiration from the haunting refrain of "Every Breath You Take" by The Police—"I'll be watching you"—reimagined as a metaphor for the omnipresence of artificial intelligence. It evokes a quiet sense of surveillance, where every action becomes data. Through this lens, AI is not just a system, but an unseen observer, constantly recording and interpreting human lives.

> AI is not just a system, but an unseen observer, constantly recording and interpreting human lives.

A circle of inclusivity & Responsibility in AI

Virtues of AI

> Predictive systems may improve planning and risk assessment, but they also raise concerns regarding surveillance, privacy erosion, and accountability gaps.

Artificial Intelligence is often celebrated for its efficiency, scalability, and analytical precision, yet its 'virtues' demand critical scrutiny. While AI enhances productivity and supports data-driven decision-making, its design risks over-reliance on algorithmic systems that may embed bias, opacity, and structural inequalities. The promise of objectivity is contingent on the quality and diversity of training data; flawed datasets can perpetuate discrimination at scale. Although AI augments human capability, it can simultaneously displace labor and concentrate technological power in a few institutions. Predictive systems may improve planning and risk assessment, but they also raise concerns regarding surveillance, privacy erosion, and accountability gaps. Thus, making the 'virtues' of AI not inherent but conditional.

Always being tracked

The artwork talks about the AI's growing presence while gently reminding us of the responsibility it carries. As we speak about algorithmic systems, accountability and privacy must be built into their very design — not as limitations, but as the truest measure of AI's virtue and trustworthiness.

> As we speak about algorithmic systems, accountability and privacy must be built into their very design

Changing Data structure of "Beings"

The artwork meditates on how human identity is being redefined in the digital age — people as data structures, each shaped by their connectivity, information patterns, and digital presence. It raises questions about identity, surveillance, digital categorization, and what remains essentially human in an increasingly data-driven world.

"

It raises questions about identity, surveillance, digital categorization, and what remains essentially human in an increasingly data-driven world.

The Platform Question

Power, Accountability and the Global South

Authors

Syed Mohammad Haroon I Karen Vergara I Tavishi I Angelina Dash I Nicole Solano Chavarría I Jamila Venturini I Catalina Balla I Carina Singh I Khush Vachharajani I Rakshita Swamy I Htaike Htaike Aung I Shaik Salauddin I Arpita Kanjilal I Shohini Banerjee I Vaishali Soni I Juliet Nanfuka I Akanksha Ahluwalia I Raina Ghosh I Osama Manzar

Editors: Dr. Raina Ghosh I Maitri Singh I Dr. Arpita Kanjilal

This experimental anthology grew from the DEF Secretariat's vision for the ARISE Community to create a shared space where scholars, practitioners, and activists could think, write, and act together on questions of digital justice. Expanding beyond ARISE to include voices from Latin America, South Africa, India, Myanmar, and elsewhere, it brings reflective, critical, and creative contributions that challenge Big Tech's dominance and reimagine accountability through diverse, grounded, and decolonial perspectives.

defindia.org

ariseglobalsouth.org