Experiments with Social Good: Feminist Critiques of Artificial Intelligence in Healthcare in India

Radhika Radhakrishnan

Gender Research Manager, World Wide Web Foundation
radhika.radhakrishnan5@gmail.com

Abstract

In contemporary India, AI-enabled automated diagnostic models are beginning to control who gets access to what kind of medical care, with the most invasive systems being aimed at underserved communities. I critically question the dominant narrative of “AI for social good” that has been widely adopted by various stakeholders in the healthcare industry towards solving development challenges through the introduction of AI applications targeted towards the sick-poor. Using feminist theory, I argue that AI systems should not be seen as neutral products but complex sociotechnical processes embedded with gendered knowledge and labor. I analyze the layers of expropriation and experimentation that come into play when AI technologies become a method of using diverse bodies and medical records of the sick-poor as data to train proprietary AI algorithms at a low cost in the absence of effective state regulatory mechanisms. I posit that an overwhelming focus on “spectacular technologies” such as AI derails public efforts from solving the actual needs of populations targeted by the “AI for social good” narrative, and from the development of sustainable, responsible, situated healthcare solutions. Lastly, I offer social and policy recommendations that would enable us to envision inclusive feminist futures in which we understand and prioritize the needs of underserved populations over capitalist market logics in the development, deployment, and regulation of AI systems.

Keywords

AI for social good, feminism, healthcare, development, India

Introduction

The narrative of the relationship between humans and technology has undergone considerable change in the past century. This change has had much to do with feminist scholarship developing constructive approaches to engage with technology and with changing perceptions of technology itself (Achuthan 2011). In India, specific to artificial intelligence (AI) technologies, discourses describing AI co-exist as simultaneously positive (as found in “AI for social good” narratives of technologists endorsing AI to solve social challenges) and critical (as found in stances of feminist scholars of technoscience), both of which this paper will explore. These oscillating attitudes towards technology are particularly characteristic of contemporary postcolonial societies. It is in this postcolonial context that I attempt to bring fresh perspectives to these seemingly disparate theoretical framings of technology by critically examining the implications of perceiving AI technologies as objective solutions to social and development challenges in healthcare.

Within healthcare, AI research and deployment is driving technology applications in hospitals, pharmaceuticals, diagnostics, medical equipment and supplies, medical insurance, and telemedicine in India (Paul et al. 2018). Moreover, the development of such applications has received significant policy and regulatory support from the Indian state through its AI policy proposals, and providing conditions of possibilities for private technology companies to test and deploy their applications in the country (Artificial Intelligence Task Force 2018; NITI Aayog 2018). This almost all-encompassing emphasis on AI has been described as the “AI hype” (Hao 2019) with stakeholders across the board attempting to develop AI applications and the Indian government scrambling to produce national AI strategies.

Today, automated diagnostic models powered by machine learning (ML) are beginning to control who gets access to what kind of medical care, with the most invasive systems aimed at marginalized communities at the intersections of class, caste, gender, and other relevant social factors (Burt and Volchenboun 2018; Lecher 2018). These technologies (a subclass of AI),¹ trained on historical data to uncover patterns and learn from examples, are able to learn to predict and classify generalized future outcomes for the purposes of decision making. They are based on large datasets (Big Data) that reflect historical and systemic biases of their sociocultural environment, which affect the outcomes of decisions made by the data models. Backed by state support advocating AI usage for furthering public health policy objectives and addressing development challenges, these automated diagnostic systems aim to assist doctors in making patient diagnosis decisions—for example, to algorithmically detect diabetic retinopathy (Gulshan et al. 2016) in patients through retinal image capturing and processing. My research focuses on the use of technologies that harness AI to assist medical practitioners to diagnose diseases under the umbrella of “AI for social good,” with attention to

the intersectional lived experiences of patients in low-income marginalized contexts.

This paper is organized as follows. The second section describes the research methodology employed for this study, as well as its limitations and ethical challenges encountered. In the subsequent three sections, I argue that the “AI for social good” narrative and technology solutions under its umbrella perform as “spectacular technologies” (Nandy 1988) that derail from sustainably meeting the actual needs of populations targeted by their own identified agenda. This derailment happens because the initiatives inspired by this rhetoric perform the following functions. First, they replace labor with capital to frame social development as innovation opportunities to gain social legitimacy and state support, biasing healthcare investment and policy decisions (as described in “Critically Unpacking ‘AI for Social Good’”). Second, they fail to account for people’s socioeconomic and gendered lived realities and therefore produce limited knowledge and a consolidation of gendered care stereotypes under the guise of objectivity (as described in “Gendered Labor and Knowledge in ‘AI for Social Good’”). Third, they use diverse bodies and medical records of the sick-poor as corpus data to train proprietary AI algorithms at a low cost in the absence of effective state regulatory mechanisms (as described in “AI for Social Good or Experimentation?”). In the last section, I critically reflect upon the social and policy implications of the use of AI diagnostic technologies for the social good, with recommendations to constructively move forward in this scenario and imagine feminist futures that prioritize the healthcare realities of targeted underserved populations in the design, deployment, and regulation of technological solutions.

Research Methods

I carried out a series of qualitative, semi-structured, and in-depth interviews with different stakeholders in healthcare and technology settings in Bengaluru (Karnataka), Madurai (Tamil Nadu), and Thiruppuvanam (Tamil Nadu) in 2018. Bengaluru (with a population of over 10 million, making it the third most populous city in India) is the capital of the southern Indian state of Karnataka; Madurai (with a population of about 1.5 million) is a major city in the southern Indian state of Tamil Nadu; and Thiruppuvanam (with a population of about 21,000), in Sivaganga district in Tamil Nadu, is a town neighboring the city of Madurai. Bengaluru was chosen as a site because it is considered the country’s “IT capital,” housing the biggest technology companies and start-ups in India’s health-tech ecosystem. Madurai and Thiruppuvanam were chosen because hospitals and healthcare centers here are among the few sites where AI-enabled diagnostic systems were being tested at the time of conducting fieldwork.

The research tools I used for the study were interview schedules, and the analytical tool was manual coding. I used purposive and snowball sampling and

interviewed twenty-one men and seven women in a total of twenty-two interviews in English and Tamizh. The participants I interviewed include six leading AI data scientists from six technology companies, seven leading healthcare delivery and five paramedical personnel across four hospitals and clinics, four patients, two investors of healthcare technology start-ups, and two representatives of a people's health movement. Some names of participants have been changed as per their requests to protect their identities, and the changes have been mentioned where applicable. I also carried out two hospital tours as well as a tour of a rural health center for participant observations, and analyzed policy documents related to Indian government AI initiatives between 2017 and 2021.

All of the interviewed participants representing established national or global organizations are currently performing clinical validations in different settings,² some of which I observed during my fieldwork in southern India. Pre-deployment, or early-stage research on AI-enabled automatic diagnostic models, such as investigated in this study, is a crucial stage because once these digital systems scale up, they can be remarkably hard to decommission (Eubanks 2018).

One of the limitations of my study is that I was unable to collect data on the caste composition of research participants owing to complexities of navigating cultural sensitivities of caste in the field as a dominant-caste researcher. Therefore, I have not incorporated caste as a primary lens for analysis in this study despite caste being a significant axis of stratification in Indian society. My primary ethical concern for this study has been to negotiate with the politics of studying “up” as a female researcher. The medical personnel, data scientists, entrepreneurs, and investors I interviewed are predominantly cis-male persons holding positions of expertise in the field. There was a noticeable power dynamic in our interactions including my routinely being talked down to and, in one instance, stalked and solicited by a senior medical practitioner during my stay in the field. Another key challenge was to access institutional research participants since trade secrets, institutional secrecy, and corporate non-disclosure agreements made it insurmountably difficult to gain access to private institutions and I was often denied entry by their public relations teams. Thus, this study is exploratory in nature and can be taken forward by addressing these limitations and challenges.

Critically Unpacking “AI for Social Good”

The rhetoric of “AI for social good” has been adopted by organizations as well as the Indian state in its AI policy positions. Global technology companies such as Microsoft (n.d.), Intel (n.d.), and IBM (n.d.) have publicly articulated their philosophical stance conceptualizing AI as a tool for social good. Google's first AI principle is “to be socially beneficial”; their work aims to use AI to “explore and address hard, unanswered questions” (Google, n.d.). Similarly, the Wadhvani Institute for Artificial Intelligence, a non-profit research organization that partners

with the Bill and Melinda Gates Foundation in India (Gates Foundation 2019), primarily focuses on “AI for social good” by aiming to “transform the lives of...billions of poor, underserved people” through AI-based solutions (Wadhvani, n.d.).

To aim a critical lens at these claims of “social good,” I begin by interrogating how technology companies and healthcare organizations in India—which collaborate with each other to develop AI-enabled diagnostic tools—describe “AI for social good.” I unpack what “social good” means for their philanthropic approaches, and analyze whether and how these definitions shift on the ground in the healthcare spaces I studied by focusing on who has the power to shape AI-driven futures.

Among the six technology companies and four healthcare delivery partners offering AI-enabled diagnostics that I studied, all of the collaborative interventions are being developed with the rationale of improving healthcare access to rural India, which has an acute shortage of skilled doctors, as observed by one of the research participants at a global technology company: “It [the motivation] came from the need [for]...increasing access. You need a specialist to diagnose, and they're not available in many parts of the country, and even if they are, it's expensive...Even quality is suspect. We improve with both—we can reduce cost and increase access by quite a lot into areas that don't have that” (Mr. Gaurav, AI data scientist; name changed). In such under-resourced communities, AI-enabled diagnoses would “democratize access” to healthcare, according to the founder and head of an AI healthcare start-up in India (Mr. Wasim; name changed). A data scientist also mentioned that “it's going to be profitable for us in the long run because healthcare is one of those areas where you can make profits” (Mr. Gaurav, AI data scientist). These perspectives hint at a form of social good that is intended to catalyze social development, yet is driven by an interest in capital flows benefiting corporate interests. Noting the ideologies inherent in deploying democracy in the name of neoliberal, global capitalism reveals motivations underlying corporate philanthropy.

The rhetoric of social good is also expressed in the state's policy positions on AI. In 2017 the Artificial Intelligence Task Force was instituted by the Department of Industrial Policy and Promotion under the Union Ministry of Commerce and Industry of India (Artificial Intelligence Task Force 2018). Specific to healthcare, the report states that “AI has the potential to transform delivery of health services in rural areas” (Department of Industrial Policy and Promotion 2018, 21–23). Further, India's NITI Aayog (National Institution for Transforming India), a government think tank that makes policy recommendations and is involved in the policy's implementation, produced a *National Strategy for Artificial Intelligence* in 2018. The strategy document states its primary objective as to “leverage AI for economic growth, social development and inclusive growth” (NITI Aayog 2018). The NITI Aayog report also advocates for the use of AI in healthcare (2018, 24–29)

offering the same reasoning used in the Department of Industrial Policy and Promotion report geared towards social development sanctioned by state support for the same. These policy documents commonly reflect the optimism that AI can help India accelerate the pace of its socioeconomic development.

To further analyze how the “AI for social good” rhetoric reflects socioeconomic power structures, I now turn to interrogating who is targeted through these social good interventions and what factors have led to their position requiring these interventions.

I borrow the term *sick-poor* from Purendra Prasad’s (2007) sociohistorical appraisal of health systems in contemporary India, wherein the sick-poor comprise an under-resourced cross-section of India’s vast population that is cut off from meaningful access to healthcare. Understanding the reasons for this in a technological era and unpacking the experience of “AI for social good” for the sick-poor first requires a better sociopolitical understanding of the Indian healthcare ecosystem. Public healthcare activists I interviewed broadly locate Indian healthcare today between a completely dysfunctional public health system and a predatory private health industry. Both stakeholders target the sick-poor, the state aiming to “control and convert it into a certain idea of development” and the private sector driven by “a monetary idea of profit” (Mr. Vijay Seethappa, public health activist). Thus, these technological interventions are happening almost exclusively within an unaffordable (for the majority of the population) private healthcare industry that exists alongside a dysfunctional public healthcare sector.³

The generally poor quality of the Indian public health system is used to justify technological interventions within the private healthcare system, rather than to support an argument for improving the public healthcare system. For instance, given the widely acknowledged concern of scarcity of medical practitioners in public healthcare, the question one should be asking is, “If you’re not able to get doctors to sit in the most peripheral centers...what are you offering technology for? Why are doctors not sitting in those peripheral centers?” (Mr. Vijay Seethappa, public health activist). Though India has the capacity to produce 67,000 graduating doctors per year (Ministry of Health and Family Welfare 2017), public health activists suggest that, given the increasing costs of medical education, graduating medical students lack incentive to join public healthcare in remote parts of the country and can pursue more lucrative alternatives such as super-specializations in medical fields. Concurrently, state policies create greater opportunities for doctors to join the private healthcare system as opposed to public healthcare. Many of the efforts to enhance private healthcare are similarly rooted in the deployment of technological tools.

It is important to foreground this structural context to understand the reasons for labor shortages in healthcare in India and to assess how quick-fix technological solutions are inadequate and unsustainable. The status quo focus is on altogether replacing labor with capital in the form of advanced technologies in privatized spaces. This implies that the technological interventions being taken up do not necessarily arise from the actual needs of the targeted population: “What is the point of having an [AI-enabled] scanning machine in a place where there is no electricity?” (Ms. Akhila Vasan, public health activist). Many common technological solutions represent a capitalist profit logic. As feminist technology scholars have argued, the medical-industrial complex routinely channels resources into profitable areas with no connection to satisfying actual human needs (Wajcman 1991, 73).

Technology is also deeply embedded in India’s national imagination for development. Historicizing the role that science and technology have played in narratives around social development, Ashis Nandy (1988) argues that science has primarily been an instrument put to the use by the colonial state. Nandy posits that in postcolonial India, the focus of science has shifted to “spectacular science,” which has been used as a “reason of the state” to further political interests and, consequently, has given “spectacular technology” a central place in science. Such technology, he argues, has created conditions wherein the Indian middle classes believe that technology will tackle all their sociopolitical problems without questioning the domination of technocratic elites over the decision-making process (Nandy 1988).

Applying the idea of “spectacular technology” to AI-based diagnostic systems provides a framework to understand why technologies driven by AI are often more complex (and costly) than needed to address the problem. For instance, to measure birth weight in poor rural homes, Wadhvani AI (n.d.) could create a simple, inexpensive, accurate measuring scale rather than an expensive, AI-based smartphone application to provide three-dimensional models of newborn babies. However, Wadhvani AI (n.d.) identifies that a key ethos in the spectacular technology approach is to employ AI in global health to empower maternal and child health. This strategy demonstrates a conscious effort on behalf of technology designers to focus on developing “spectacular technology” even if less elaborate, more cost-effective, and sustainable solutions can be found for the same social problem. Thus, the “AI for social good” narrative and initiatives inspired by it perform themselves as a “spectacular technology”; this approach is used by AI developers and associated stakeholders to frame social development as an opportunity to innovate by appropriating the needs of the poor, while gaining social legitimacy and state support.

Further, this approach reveals a particular kind of post-development privatized future for the domain of healthcare. It is a future in which the inequities of the

existent healthcare system remain in place under the valorizing guise of “spectacular technology” while also incentivizing the creation of new privatized medical subspecializations (i.e., AI-enabled diagnostics) that further channel resources away from public healthcare towards specialized privatized healthcare futures. These medical subspecializations offer doctors in privatized healthcare spaces with newer areas to explore and lay expert claim to, thus also attracting funding and support for the development of the field (Pfeffer 1987). This professional incentive is evidenced by Dr. Lalith (name changed), a consultant ophthalmologist at a national medical establishment chain that collaborates with a global technology company for an AI-enabled diagnostic solution:

“In 2015, there was an evolution of my work. I didn’t know about AI before...Once I got to know about it, it’s been a fabulous experience...AI became an obvious layer that would add amazing value to the work we did and...that is something that got our teams excited.”

Claiming expert technical knowledge is one of the methods of legitimizing subspecializations within healthcare because funds and professional acclaim are often distributed within the medical profession according to the technological sophistication of the specialty (Wajcman 1991). In such a context, the introduction of AI-enabled healthcare technology has the effect of driving an increase in its usage by its mere existence, which in turn drives the development of more advanced technologies (through collaborations between private medical establishments and AI technology companies) as the healthcare space becomes more technologized. These effects bias investment in public healthcare as noted by Ms. Akhila Vasan, a public health activist:

“The amount of resources that are devoted to the higher-level private facilities is much higher than the first-level PHCs [Public Healthcare Centers]...So, you’re making the first level dysfunctional and pushing people to seek care at higher levels, which is wasteful and resource-intensive.”

Therefore, the increase in deployments of AI-based “spectacular technologies” channels resources away from public healthcare towards more technologized super-specialized futures, which are shaped and driven by private stakeholders without heed to the needs and experiences of the sick-poor. The “AI for social good” agenda, as such, derails sustainable problem solving for underserved populations in need.

Gendered Labor and Knowledge in “AI for Social Good”

The advent of the “AI for social good” narrative drives the technologization of medical services targeting peripheral populations. The authority of the medical

expert—which has thus far predominantly served central/urban populations in privatized healthcare, as seen earlier—is being extended by deploying data-driven diagnostic systems in under-resourced areas. As a result, more bodies (of the sick-poor) are being brought into formal medical institutions in the private sector, enabling them to be governed. I use the term *governance* here to refer to a biopolitical form of power over the bodies and lives of the sick-poor by controlling their access to and experiences of equitable healthcare.

To better understand the nature of this governance, it is crucial to inquire who are the experts designing such systems of diagnosis and governance, whose laboring bodies operationalize these systems, and the nature of knowledge produced through the systems. After providing a conceptual framework to situate these enquiries, I engage with them through observations from my fieldwork and feminist scholarship on epistemology and labor: first, through an analysis of the gendered nature of labor in AI-integrated healthcare systems in India, and, second, by unpacking the construction of knowledge systems in this ecosystem.

Reframing AI Technologies as Complex Sociotechnical Processes

To critically investigate the role that gender plays in the design and deployment of AI-enabled diagnostic systems, a fundamental shift is needed in the traditional conceptualization of AI—and more broadly, technology—as objective and neutral. The diagnostic decisions produced by algorithmic models are widely held, especially in industry and government spheres, to be objective. To wit, a research participant mentioned the motivation for developing their AI solution as “it will help the [doctor] provide evidence-based treatment to the patients without any bias” (Mr. Indra, AI data scientist at a global technology company; name changed). Another research participant from one of the global leaders in data technologies said:

“The thing about healthcare is that it’s a very imprecise science. Because the decisions there have a lot of ambiguity and a lot of mistakes...To improve the objectiveness of medicine, [we are] bringing in more data to make decisions...[and] a little bit of mathematics and statistics and engineering” (Mr. Gaurav, AI data scientist; name changed).

Feminist scholarship has consistently debunked this myth of objectivity in science and technology (Adam 2000; Harding 1986; Longino 1987). Contemporary feminist scholarship has raised several concerns under the broad umbrella of “fairness, accountability, and transparency,” including the framing of technology to center the choices of the makers and modelers of technology in evaluating the fairness of its outcomes. Feminists have shown that data reflects historical and systemic biases, and discrimination in the sociocultural environment that it represents affects the outcomes of algorithmic decisions, which are particularly disparate for marginalized communities, and naturalize prejudices against them

(Noble 2018; O’Neil 2016; West, Whittaker, and Crawford 2019). AI in particular is shown to not only reflect but also amplify biases by combining multiple generalizations on training or input data to maximize the accuracy of the model.

For example, a consultant ophthalmologist at one of India’s leading national hospital chains mentioned that “you need to take all of this as a double-edged sword...The technology will do good if you tell it to do good; it will do bad if you tell it to do bad” (Dr. Lalith; name changed). This reductionist reasoning relating to complex technologies reflects the understanding that automated decision making by AI-healthcare tools renders AI technology an inherently neutral product until it is transformed through human involvement for “good” or “bad” applications.

Feminist scholars of technology such as Judy Wajcman (1991, 60–63) have shown that it is crucial to address the contexts in which technologies are produced. Rosalind Pollack Petchesky (1987) and Michelle Stanworth (1987) viewed technology as neutral or ambivalent with the potential to empower or disempower based on its contexts and use. They argued that instead of problematizing technologies themselves, feminists should problematize the power relations and institutional settings within which these technologies are applied. Such theories can be broadly classified as theories of the social determination of technology (Winner 1980).

Theories of the social determination of technology, though widely popular in the 1990s, invisibilize the ways in which technologies *themselves* can embody forms of power and authority (Longino 1987). On the other hand, the theory of technological politics draws attention to the response of modern societies to certain technological imperatives, and acknowledges that there are sociopolitical processes at play in the common conceptualization of technology as a neutral product which resist the notion that technical artifacts have political qualities (Winner 1980). Rather than insist that we reduce technologies to the interplay of social forces (as suggested by the theories of social determination of technology), the theory of technological politics identifies technologies as political phenomena in their own right.

By framing AI technologies as complex sociotechnical processes, I reject the core assumption of a social determination of technology approach, which suggests that technology itself is neutral. When we permit AI technology to be viewed as a finished neutral *product* separate from the *processes* of its design, we risk allowing the knowers of the subject (i.e., expert stakeholders involved in the creation of AI applications) to escape scrutiny under the guise of unbiased and objective decision making. In framing technological systems as complex sociotechnical processes, we question the widespread notion that automated decision making translates into better decision making, by foregrounding the subjectivity and

personal, political, and disciplinary commitments of the AI developers and their institutional settings. I now turn to this analysis, addressing how these complex sociotechnical processes are gendered, constructed with biased labor and epistemology.

Gendered Labor in “AI for Social Good”

Medical professionals and data scientists who are engaged in the process of collaborating on designing automated diagnostic tools, as well as using them to diagnose patients, mostly comprise of upper-class cis-men. One of the AI teams from a global company developing an AI diagnostic tool for India stated that only 15 to 20 percent of the people working on their project were women, and that this figure was even lower in other projects: “Two of the doctors we work with are women, but a lot of the specialists we have consulted are men” (Mr. Gaurav, AI data scientist; name changed). In addition to the lack of gender diversity, there were few disciplines beyond computer science and medicine represented on teams building these systems. I did not find any social scientists or ethnographers (trained in understanding the social impact of processes and products) to be directly involved with any of the development and deployment teams that I interviewed for this study.

I also visited remote vision centers to observe the workflows in these sites and interview key stakeholders.⁴ Here, only young women are recruited as mid-level ophthalmic personnel (MLOPs) at the hospital and its vision centers, revealing an age and gender bias among employees. Prior to the advent of AI-enabled diagnosis, the usual work of MLOPs ranged from clinical tasks such as refraction or assisting, and non-clinical tasks such as housekeeping and organizing medical records. Since the introduction of AI-enabled technologies, MLOPs in smaller clinics collaborating with base hospitals are additionally in charge of tasks such as scanning the patient with AI-enabled tools for the automated diagnosis.

Despite the increase in workload, there are no additional wages offered to MLOPs to work with AI systems. In fact, there is no wage difference between MLOPs irrespective of their work. They are recruited after the completion of their twelfth standard education (usually at the age of seventeen or eighteen years) on a contract of five years. Their salaries go into a bank account from which they are not permitted to withdraw money without the explicit permission of their parents or senior medical personnel at the base hospital. At the end of a specific employment duration, they are entitled to receive a salary bonus that “could be used for their marriage or education,” as I was informed by the hospital’s engagement coordinator. The restrictions on a young MLOP’s finances limit her negotiating potential, which contributes to a form of financial control that healthcare systems wield over female link workers. Such forms of financial control have historically been used to socially and culturally “train” female workers into being more submissive and therefore less likely to unionize against the employer (Elson and Pierson 1981). In the digital economy, this patriarchal form of financial

control cements the undervalued nature of gendered labor that forms the backbone of AI systems on the ground.

Moreover, I was informed by a doctor at the base hospital that MLOP positions were open only to women because “women are more compassionate than men,” and it was believed that “such things [care labor] can be done only by women” (Dr. Arun, general physician; name changed). Referring to the socioeconomic context of the town, the doctor also stated that “in these areas, women are usually the caregivers. Fathers are not very involved. Usually, they have male doctors and female nurses. It is fitting for the hospital’s culture.” Men were recruited in these positions only to “accompany women” when the hospital could not “send a single female candidate with [patient younger than fifty years of age] for medical follow-up camps” (Dr. Arun, general physician; name changed).

Job opportunities here reflect gendered care stereotypes, where devalued care work is seen as “women’s work” and the high-paying and valued expert roles of the doctors are largely occupied by men. Moreover, the attitudes of doctors evidenced a savior ideology—suggesting that recruiting young women to these jobs immediately after their schooling when “they are ready for marriage” offered “self-empowerment, safety, and dignity through their work,” saving them from a life of “marriage and agricultural labor” (Dr. Arun, general physician; name changed). This logic suggests that merely increasing work participation for women leads to their empowerment. Such reductive framing is problematized by Amar Jesani (1990) who argues that the labor market favors men over women, and, more relevant here, that the division of labor within occupations is sex-biased. This patronizing logic is commonly held across the domain of healthcare. Since this logic is so endemic, it is normalized and therefore not recognized as biased. The gendered labor that is constitutive of the sociotechnical processes of AI-diagnostic systems is thus largely invisibilized under the guise of objectivity, reinforcing gendered notions of care work (Atanasoski and Vora 2019), and making them harder to contest.

Gendered Knowledge in “AI for Social Good”

Gendered hierarchies of labor are sustained by gendered hierarchies of knowledge in the design of AI systems. Diagnostic predictive AI requires the ability to codify knowledge in the form of rules that can be used to design the system. There are “clinical guidelines...established by the clinical community” to be followed for enabling the “algorithmic prediction to map to a particular diagnosis” (Mr. Gaurav, AI data scientist; name changed). Moreover, it is generally agreed in the Indian AI-healthcare ecosystem that evidence-based diagnoses are more accurate and thus preferred. Thus, a medical practitioner’s experiential and subjective knowledge of healthcare is considered less valuable, and even detrimental, to the project of designing an unbiased data-driven system. This worldview reflects a firm hierarchy of evidence over contextualized knowledge that subverts human expertise and experience and denies them agency.

Further, within this hierarchy, the knowledge of some is more valued than others. As Alison Adam (2000) shows, in classical epistemological approaches, the identity of the knowing subject is not important, which disguises an “implicit hierarchy of knowers” in the system, and privileges the AI designers who are mostly white, male, middle-class Global North professionals over the perspectives of women. This privileged “deleted knowing subject” can be correlated to the model of the “expert” that is dominant in healthcare technology (Mol 2008). This hierarchy leads to invisibilized relationships between the knowledge represented and the subject doing the knowing (Adam 2000, 240), privileging the knowledge and viewpoints of the “expert.” The AI-healthcare industry, globally and in India, is largely populated by such male, upper-class, upper-caste professionals, but also engineers and data scientists who are often far removed from the socioeconomic contexts of those for whom they design tools. Such models reproduce ideologies that are damaging for marginalized communities and reinforce prejudices about them. For example, a consultant ophthalmologist at one of India’s leading national hospital chains implementing AI decision-making in their hospitals said,

“I would put patients in different buckets of intellect...You will not expect a rural individual to come in and then expect them to understand AI...For those who understand, for those who are from the IT sector...they are thrilled” (Dr. Lalith, consultant ophthalmologist; name changed).

This doctor differentiates rural patients and IT experts to justify why the hospital does not provide the former with complete explanations about the role of AI in the automated decision making of their diagnosis. By denying patients this information, such hierarchization limits a patient’s ability to fully understand their own diagnosis.

Moreover, technology designers have little awareness of the critique and assessment that restrain development. For example, some employees of global technology companies I interviewed have never allowed an independent study of their AI diagnostic tools; nor have they published any scientific papers demonstrating how the tool affects patients. None of the data scientists I spoke to were aware of how the technology was being used outside of the laboratory, as evidenced by Mr. Gaurav’s remarks about consent:

“For the validation we’re doing, we give the consent forms to the [healthcare] partners. A lot of the hospitals did not care about it. I don’t know what exactly they told the patients...The partners had their own processes...I don’t know how seriously they take it...I never saw the practice in action” (Mr. Gaurav, AI data scientist at global technology company; name changed).

As such, technology designers fail to inquire as to whether these diagnostic technologies are in fact improving the health of the poor—an approach that is patronizing and denies patients the right to fully consent to their diagnosis and treatment.

In keeping with this trend, the sick-poor “targets” of outreach initiatives are marked by their class location, and almost always belong to the same strata of marginalized communities. Their social locations are disadvantaged vis-à-vis private stakeholders who are responsible for medical data collection. My fieldwork showed that the underserved communities that are targeted through such AI solutions are never consulted by technology designers to understand their actual needs, and feedback about their experiences with the technology are not sought by technology designers when the technology is in use. Technology companies value performance of their systems over the experiences of people with them, as indicated by a data scientist at a global technology company:

“The first goal is performance, definitely. And we have some indicators where the AI is already better than certain doctors on an aggregate level. So the question is, Should we go ahead with this [implementation] or should we wait for explanations to patients?...I don't think explaining to patients is critical...They might not be most informed to understand it...They may even misunderstand it and cause panic” (Mr. Gaurav; name changed).

Thus, the lives of the sick-poor are rarely accessed as sources of knowledge; instead, they are expropriated as the targets—as opposed to beneficiaries—of outreach campaigns offering solutions without attention to their needs and experiences.

This abdication of the lived experience of the poor violates key feminist theory that seeks to empower marginalized communities by centering their knowledge. For example, Sandra Harding’s (1992) “feminist standpoint epistemology” insists that some knowledge—such as that of those who are oppressed—can only be accessed via lived experiences or their standpoint; this knowledge is invisible to others, leaving the dominant groups in an epistemologically disadvantaged position due to their privileged positions in societies stratified by gender, race, class, sexuality, and other divisions. Further, since “social good” is meant to translate to healthcare outcomes relevant to “social justice,” Chandra Talpade Mohanty argues that an “experiential and analytic anchor in the lives of marginalized communities...provides the most inclusive paradigm for thinking about social justice” (2003, 510).

Building on Harding and Mohanty, I argue that the “AI for social good” rhetoric reflects the dominant group’s limited knowledge and drives the development of mismatched classes of technology for solving the social developmental challenges

it has set out to address. It also masks gendered knowledge and labor in its design under the guise of “objectivity” of AI. When poor, middle- and low- level stakeholders are not consulted in decision-making processes that determine their own futures, this is “spectacular technology” at play. These dynamics engender an expropriation of the lived experiences of the poor for whom the technology does not generate any value, and also fails to meet the state’s social development goals.

AI for Social Good or Experimentation?

“AI for social good” initiatives not only expropriate the lived experiences of the sick-poor through their gendered dominant group epistemology and labor workflows (as analyzed so far), but are also a site for experimentation upon the sick-poor. The development of AI-enabled diagnostic systems, which requires large datasets, is facilitated by collaborations between the medical establishment and technology companies, which arguably have two primary purposes: sharing of patient data in the form of medical records, and having access to the bodies of patients to test the AI-diagnostic tools. For technology companies, the introduction of AI technologies thus becomes a method of using bodies and medical records of the sick-poor as data to train AI algorithms.

India is an attractive site for data collection for many reasons, and AI application experiments are the latest in a long history of clinical trials conducted in the Global South, and India specifically, by US pharmaceutical companies (AI Jazeera 2011). India is a sought-after testing ground for experimental pharmaceutical drug trials due to spoken English, an established medical infrastructure, welcoming attitudes toward foreign industry, and, most importantly, legions of poor, illiterate test subjects who are willing to try out new drugs (AI Jazeera 2011). In this context, I now investigate the ethical problematics and risks in using India as a testing pool for experimental AI technologies, how this plays into a prioritization of corporate profit over respect for patient rights and knowledges of the poor, and its potential impact and future created. Specific to the case of automated diagnosis systems in healthcare, I analyze three primary factors that aid the process of automation and impact national health and development plans by setting in motion a privatized techno-utopian future: the diversity of Indian populations; cost effectiveness of collecting data in India; and an unregulated healthcare ecosystem.

One reason for locating these algorithmic validation processes in India is the diversity of populations in the country, as noted by a research participant (who heads the medical team for their AI project) from a national hospital chain:

“India is a hub for diabetic retinopathy. You don’t see these variations of diabetic retinopathy patients back in the US. They [technology collaborator] wanted to train the software with the numbers that we have—our hospital” (Dr. Naresh, hospital director; name changed).

India's diversity covers much of the spectrum of the world's populations, and the Indian state, in collaboration with market agents, makes these diverse Indian populations available to Global North corporate interests as "experimental subjects" (Rajan 2005). "Spectacular technology" is once again at play in the articulation of these idealized "experimental subjects."

Second, I observed that the data collection for training AI-enabled diagnostic algorithms is being combined with patient treatment.⁵ For example, patient data collected for treatment is simultaneously used as input data to train the AI algorithm, and the algorithmic result is manually correlated with a doctor's diagnosis. If there is a match between the manual and automated diagnosis, the patient diagnosis is confirmed and reported to the patient. If not, the automated diagnosis is reviewed by a doctor, and the feedback is entered back into the algorithmic system as part of the clinical validation.

By combining patient diagnosis with experimental trials, the cost of data collection for the technology companies and healthcare providers is significantly reduced as there is no additional cost to conduct a separate set of procedures to design the algorithm. This scenario raises ethical concerns through a conflict of interest between the technology company's market-driven interests and the patient's best interests, such that divergent interests have to be catered to by the doctor. Since the marginalized socioeconomic contexts of the sick-poor reduce their bargaining power and ability to demand their right to effective healthcare within a medical-industrial complex, such a conflict of interest would normally favor vested market-driven private interests over patient interests. During the debates around the Karnataka Private Medical Establishments (Amendment) Bill 2017, public healthcare activists that I spoke to had demanded that hospitals should not only declare all the clinical trials underway at a given facility, but also that the clinical trial facility should be physically separated from the treatment facility in a hospital. However, these demands were not taken on board due to pushback from private stakeholders. As a result, the reduction in the costs of data collection for private stakeholders is offset by a compromise on patient interests and their right to seek healthcare without participating in experimental algorithmic trials.

Third, the weak regulatory landscape in India contributes to what could be called a "low rights environment" where there are few expectations of political accountability and transparency (Eubanks 2018). For instance, though patients are asked to sign a consent form for their data to be sent to the base hospital for training the algorithm, I observed that most patients did not understand what they were consenting to, and where their diagnosis was coming from—through human versus machine evaluation. This was observed in the following interaction at a local clinic during my fieldwork (translated from Tamizh):

Patient: Can you give me a Xerox copy [photocopy] of this [consent] form? In my house, my kid is studying, he can read English.

Nurse: We cannot provide a copy. Why don't you read this Tamizh version of the form?

Me: Do you understand what the form says?

Patient (after reading the Tamizh form): I don't know.

Patients end up consenting to the collection of their data for training the AI algorithm without understanding the scenario because:

"consent...is given under so much duress with so little information. You can never beat that information asymmetry in medical situations. And this is where everything is stacked against the patient" (Ms. Akhila Vasan, public health activist).

As such, consent of the sick-poor is already assumed. Moreover, patients have no real agency in choosing or rejecting to opt into these experimental trial treatments involving AI since the alternative is no medical care at all. One medical practitioner remarked, "conditioning that you are putting into the patient's mind...is going to go a long way in how they accept an AI insight" (Dr. Lalith, consultant ophthalmologist; name changed). A lack of choice translates to a lack of patient opposition to data collection, which is construed as consent to the use of these technologies, as this statement from the chief medical officer of a leading hospital chain in India highlights:

"We were wondering if patients are going to accept the technology...But people are accepting it...There is nothing to avoid or hesitate to sign [the consent form]. Everybody signs...They are willing to accept what is there" (Dr. Kiran, chief medical officer; name changed).

In this context, Prasad (2007) argues that in India health has not been a rational choice but rather a preference imposed upon the sick-poor. Through a historical account of the plurality of medical systems in India, with a specific focus on structures of caste and class, Prasad argues that current debates in the health sector revolving around the choices between public and private healthcare are not meaningful choices for the sick-poor. Many of the sick-poor in India are desperate due to the lack of medical care, and are hence coerced into entering into experimenting trials to get treatment, thereby nullifying any informed consent.

Thus, whereas in the traditional case of clinical trials, pharmaceutical companies are complicit in the process of expropriation for testing experimental drugs on the sick-poor, the advent of AI-enabled data systems recreates a similar scenario wherein technology companies take on the role of pharmaceutical companies for testing the outcomes of their algorithmic technologies on the sick-poor. In this case, the training of the algorithm on the bodies of the sick-poor people is used to improve the diagnostic technology. While information asymmetry is likely to exist between pharmaceutical companies and healthcare delivery partners, here it is amplified due to the specificities of technology company operations, as well as specificities of data-driven medical technologies that differ vastly from those within the domains of healthcare. This asymmetry can also manifest in the form of human prejudices about people and communities that are deeply embedded in the technology, practices and use-cases around it, making them seem “natural,” as indicated in the medical practitioner’s reference to the assumed lack of intellect or capacity to understand AI-assisted medical services among the rural poor.

In the face of such expropriation, there is limited (if any) effective legal protections available for the sick-poor. For instance, there is no consensus (amongst the seven medical practitioners I interviewed from leading national hospital chains in India) regarding questions of accountability in case of an error in the automated diagnosis from the AI-enabled system, despite the technology already being tested upon the bodies of the sick-poor. Some medical practitioners sidestepped questions of accountability by suggesting an “intent to do only good...even if there’s something that’s gone amiss” (Dr. Lalith, consultant ophthalmologist; name changed). The chief medical officer of a leading hospital chain remarked, “I don’t think anybody knows the answer to who should be held responsible for errors...It’s a million-dollar question” (Dr. Kiran, chief medical officer; name changed). This deliberate ignorance of expert medical practitioners justified by their intent geared towards social good reflects an evasion of ethical responsibility towards the sick-poor.

Moreover, the state plays a foundational role in erasing the sick-poor from all decision making regarding the management of public healthcare, and their legal rights. The state is in fact complicit in producing socio-legal-technical conditions under which the sick-poor emerge as an available “data source” for experiments. Kaushik Sunder Rajan (2005) argues that there exists a “hybrid state-corporate assemblage” in India that is aimed at the nation-state becoming a “global player” in the global marketplace through corporate entities whose conditions of possibility are largely enabled by the state. This dynamic is akin to Nandy’s (1988) accusation that science and development demand sacrifices from and sufferings of ordinary citizens. Finally, state support for such private-industry-led AI-enabled healthcare technologies is evident in the key decision by the Telangana government in 2017 to adopt Microsoft’s Artificial Intelligence–based eyecare screening technology as part of the Rashtriya Bal Swasthya Karyakram program

under the National Health Mission (Microsoft News Centre India 2017). Thus, instead of providing equitable public healthcare access to all, the state plays the role of an enabler for private technology companies and healthcare providers to operate within market logics that experiment upon and expropriate the needs of the sick-poor.

While a harm-based framework is often employed to assess impact, it is difficult to have evidence of harm for emerging technologies (i.e., at the stage of development and testing) because these harms are often not tangible and observable given the opaque nature of the technologies themselves and the institutions they are produced in. In such cases, we need to move away from the need for evidence of harm and still be able to pre-empt the dangers from such technological interventions. We can highlight such dangers by focusing on potential risks arising from practices around the technology's design, attending to the ethics involved in health technology testing, and heeding patient consent and choice. Virginia Eubanks (2018) proposes a useful framework to evaluate the impact of an automated tool that is targeted towards the poor; the framework questions whether the tool increases the self-determination and agency of the poor, as well as whether the tool would be tolerated if it was targeted at non-poor people. As has been analyzed in this section, automated diagnostic tools generally neither generate any value gains for the targeted sick-poor nor increase their agency, but in fact worsen both these conditions of possibility. This scenario arises through, first, the use of their personal medical data for training AI algorithms that do not generate any value gains for the sick-poor, and second, by burdening the sick-poor with risks emanating from uncertain and unforeseen results of such untested technology through the experimental use of automated diagnostic systems in medical services.

Social and Policy Implications: Envisioning Feminist Futures

Expropriation through technological outreach initiatives under the umbrella of "AI for social good" is rooted in the marginalized class and gender dynamics of patients and workers in the healthcare ecosystem and is facilitated by alliances between private technology and healthcare industries at the site of these social dynamics. In other words, private technology companies supported by private healthcare providers and the state are exploiting AI as a tool for profit and the governance of the bodies of the sick-poor rather than harnessing it as a means to enable better care and development within the domain of healthcare.

The "law of the hammer" (as it is commonly referred to) is a cognitive bias that involves an overreliance on a familiar tool: "I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail" (Maslow 1966, 15). Viewing AI as a "tool" (similar to a hammer), stakeholders developing AI

applications view social and developmental problems as nails that can be fixed through the usage of the tool. As has been argued, what unites many of these AI applications is their focus on using “AI for social good” and consequently targeting underserved populations in the country.

As noted earlier, this fetishization of AI takes us into technocratic futures wherein collaborations between the state and private technology and healthcare companies aid the exploitation of the sick-poor justified by an agenda of development. These powerful stakeholders are taking the lead in ideating and shaping possible AI-led futures for the country, with the targeted communities having no say in these futures. Further, such data-driven systems are becoming ubiquitous in the lives of the sick-poor in enabling their limited access to healthcare and associated services. For instance, digital identification (in the form of Aadhaar, a unique identity number for Indian citizens) is mandatory for the poor to access social welfare schemes (Khera 2017), and the Indian government has passed legislation to create a national digital infrastructure under the National Digital Health Mission for health data in the country (Radhakrishnan forthcoming). These technologies are crucial to the state’s imagined healthcare future in which the health of the sick-poor would be managed technologically through collusions and collaborations with the private healthcare industry and technology companies so that the state does not itself have to provision for sustainable solutions to development challenges. Instead of managing sickness and poverty, the state thus invests in technologies that manage the individual sick-poor. Eubanks terms this uptake of data-based tools for the poor “technological innovations in poverty management” (2018, 13). Thus, the deployment of AI tools aids in the creation of a sick-poor population by channeling resources away from public healthcare towards technologization and privatization of healthcare and incentivises the move towards a future in which these conditions are maintained by centering the role of AI technologies in the state’s imagined healthcare future.

To realize a more feminist future that truly empowers underserved communities in accessing equitable, quality healthcare—which is the stated rationale for many “AI for social good” initiatives—I recommend some structural and policy-level changes in this section as follows.

Feminist scholars have mounted the challenge of what kinds of problems can be ethically solved using AI without exacerbating marginalizations (Eubanks 2018; Keyes 2018; Noble 2018; Taddeo and Floridi 2018). Building upon their work, instead of asking, “How can AI solve this problem?,” I propose a methodological reframing to foreground the question of “What problems can AI solve?” by keeping in mind the feminist futures we wish to achieve. This reframing would enable stakeholders to channel resources towards “healthcare for all” (which is a worthwhile social development agenda) instead of “AI for all” (which is the current

focus of the Indian government) (NITI Aayog 2021). Such a vantage point allows us to focus on finding need-based solutions to problems that are within the scope of technologies such as AI to solve. Such solutions are much more suitable than those developed by expropriating the needs and lived experiences of the underserved to serve market-driven interests of technology developers looking for opportunities to innovate towards “spectacular technology.”

For such solutions to be effective, we must move towards a praxis of interactive methodological approaches at the grassroots level that not only take into active consideration but center the needs and lived experiences of underserved communities in the quest to find solutions in a bottom-up, reflexive, open manner. The feminist methodology of participatory action research offers one such approach that promotes mutually respectful relationships, shared responsibilities, and an emphasis on local capacity building (Read et al. 2014). A successful technological intervention carried out using this methodology in India is reflected in the initiative of Gram Vaani (Moitra et al. 2016). Another such grounded methodology more specific to the deployment of AI technologies is the “AI on the ground” approach that focuses on using ethnography to observe, listen to, and speak with people on the ground (ACM 2021). Through these approaches, stakeholders can carefully challenge their biases and commitments by engaging meaningfully with impacted communities to understand their histories and needs.

Further, if designers are drawing upon the experiences of underserved communities, they should ensure that the applications they build benefit those communities in return and increase their agency through the use of the applications (Eubanks 2018). Diversifying the teams that develop such technologies by actively including women and social scientists who are trained in understanding social implications in the lives of people, and building institutions that encourage such diversity in their workplace culture, will be immensely helpful in this process (West, Whittaker, and Crawford 2019).

Given the state’s collusion with this ecosystem as analyzed earlier, policy-level changes are also necessary. In a policy brief on transparent algorithmic decision making in India, Amber Sinha and I (Radhakrishnan and Sinha 2020) propose four thresholds for AI regulation that are applicable to AI diagnostic systems and would be useful for policymakers to ask themselves while regulating these systems: (1) Is there a vast disparity between the primary user and the impacted party? (2) Is AI trying to model or predict human behavior? (3) Is there either a likelihood or high severity of potential adverse human impact of the AI solution? And (4) are there no reliable means for retrospective adequation? If the answer to one or more of these questions is affirmative, the authors propose greater regulatory scrutiny of the AI system. In the absence of reliable evidence of satisfactory solutions, we propose that the regulatory response must align itself to

appropriate limitation or prohibition on the AI system until such evidence is gathered (Radhakrishnan and Sinha 2020).

Furthermore, as noted earlier, there are limited legal protections available for the sick-poor. In this context, I propose that there must be an established consensus in India on who would be held responsible in case of an error in diagnosis, malfunction of the technology, or the use of inaccurate or inappropriate data in the AI system; how the acceptable level of responsibility and liability for the stakeholders would be determined; and what the process for recourse would be. There must also be established procedures for obtaining patient consent meaningfully, backed by legal regulation enforcing penalties for non-compliance. Further, meaningful alternatives to healthcare services must be available to the sick-poor to reject AI-based diagnostics. All of these legal protections must also apply to AI systems at the pre-deployment stages of data collection, testing, and clinical validation. In line with the demands made by public health activists for the Karnataka Private Medical Establishments (Amendment) Bill 2017, all clinical trials underway at a given health facility must be publicly declared, and there must be a physical separation of clinical trial facilities from treatment facilities in a hospital. These demands must also be extended to algorithmic testing facilities. Such measures would ensure that the rights of patients are safeguarded and given precedence over capitalist interests, and that patients would have meaningful recourse when their rights are violated.

On a more fundamental level, the questions we must ask of AI systems are not purely technical questions requiring technical solutions (such as independent algorithmic auditing), but should also focus on the social dimensions and the epidemic scale of the problem affecting millions of sick-poor persons and setting up dangerous futures for them. The gendered nature of labor that characterizes AI system development, and the biased, patriarchal knowledge that it reflects and reproduces takes place within a national and global context of experimentation and expropriation on the targeted sick-poor.

On one level, this shift in focus has implications for the sick-poor for whom value is not generated through these technological interventions, but rather is expropriated through the experimentation of untested medical technologies upon their bodies. On another level, this shift also has crucial implications for a feminist understanding of these systems as gendered, sociotechnical processes located within a global political economy of data that is actively enabled by state regulation and lack thereof. These harmful implications include shifts in healthcare outcomes for millions of patients (who are treated as “experimental subjects”) such as poorer understanding of their own AI-aided diagnosis due to patronizing attitudes by medical practitioners, denial of rights such as agency, choice, and meaningful consent in accessing healthcare through the data-based governance of their bodies, and further entrenchment of inequitable class- and

gender-based power dynamics in the domain of healthcare. This re-focus further implies structural shifts in public healthcare investments by channeling resources away from public healthcare towards technologized and privatized healthcare futures.

Highlighting and foregrounding these concerns in our understanding of AI systems and becoming aware of their implications (especially in the lives of targeted sick-poor communities), then, not only destabilizes the “AI for social good” narrative—especially its tendencies to naturalize our social conventions of and responses to AI systems—but also helps feminists to mount challenges to these conventions through critical interrogation and resistance.

Acknowledgments

All research and labor are social and collective. Thus, while the views expressed herein are personal and the shortcomings my own, many stimulating conversations over the past four years with experts, colleagues, and friends have deeply informed my perspectives and challenged my viewpoints. This research was carried out originally at the Advanced Centre for Women’s Studies, Tata Institute of Social Sciences, Mumbai, India, during my master’s in women’s studies (2017–2019) and taken forward at the Centre for Internet & Society, New Delhi, India, during my employment as a programme officer (2019), undertaken as part of the Big Data for Development network, established and supported by the International Development Research Centre, Canada. I am endlessly indebted to Dr. Asha Achuthan, whose advice, guidance, and constructive criticism during my research at Tata Institute of Social Sciences have significantly shaped this inquiry. I am incredibly thankful to Sumandro Chattapadhyay, research director, Centre for Internet & Society, for reviewing this paper prior to the peer review. I am grateful to the Special Section editors and reviewers at *Catalyst* for taking painstaking efforts to suggest improvements to this work. Most importantly, I am indebted to all research participants who were a part of this study for sharing their time and thoughts with me; your experiences and expertise have entirely driven this research. I am deeply thankful especially to the participants from marginalized and underserved communities; I hope this work has done justice to your lived experiences and takes forward the struggle to catalyze social change through feminist research.

Notes

¹ Machine learning (ML) is a subset of artificial intelligence (AI). AI is considered the broad discipline of creating intelligent machines, while ML usually refers to the development of machines that can learn from experience. Most AI applications involve the usage of ML because developing what is commonly known as “intelligent behavior” requires a considerable corpus of “knowledge” in the form of datasets, and (machine) learning is the easiest way to obtain that

“knowledge.” In this paper, while I specifically study medical automated diagnostic applications that are developed using ML, I extend my critique to the broader domain of AI, and hence use the term *AI* to refer to the technology powering such automated medical tools.

² Broadly, the main stages for designing and deploying an AI algorithm in healthcare include collecting and preparing data, selecting a model to train the data, evaluating the model’s performance with testing data, validating the model in clinical settings, and, finally, commercially deploying the model in clinical practice. For instance, one of the technology companies I interviewed that is developing an automated diagnostic tool for the detection of a specific disease uses an image classification deep learning model to train a retrospective development dataset of images of the potentially infected organ, which were graded by medical experts to check if the said disease exists. These images were retrospectively obtained from existing medical records of various Indian hospitals among patients who had been earlier presented for screening of the disease. The resultant algorithm was then validated using another set of graded datasets. They are currently completing further testing processes and applying the algorithm in clinical settings in Indian hospitals to determine whether the algorithm’s usage could lead to improved patient outcomes compared with current medical expert assessment.

³ While privatization of the industry is not a new phenomenon for the country, what is being observed today is not only an aggravation of the manner the industry has been privatizing, but also consolidation in the market for medical services with large corporate giants swallowing up smaller private entities and centralizing delivery of healthcare through franchisee units in various private hospitals. Such a market consolidation of medical services lends to corporate institutions exercising an almost monopoly-like control over data (medical data of patients) collection and analysis within privatized and centralized data infrastructures. For detailed analysis on the state of affairs in the Indian healthcare industry today, please refer to Gadre and Shukla 2016.

⁴ Under this healthcare delivery model, each doctor remotely overlooks eight vision centers, which receive roughly between 100 and 150 patients per day. A patient’s medical data (such as imaging scans) collected at the vision center is transmitted to the remotely situated doctor for evaluation at the base hospital, who provides the diagnosis over a video link (diagnostic telemedicine). The diagnosis report is then printed out at the vision center and handed over to the patient. A doctor visits a vision center once a month. Each vision center serves roughly 25 to 30 villages within a 5-kilometer radius. Given that the total population required in a region to set up a vision center is estimated to be a minimum of 50,000, vision centers are typically set up in regions where there are no hospitals or practicing doctors nearby.

⁵ To train an automated diagnostic tool to detect diabetic retinopathy through retinal images, the technology provider that builds the algorithmic model requires a large body of “training data,” which it obtains through collaborations with large eye-care hospitals across the country. These large base hospitals have collaborations with other local clinics in the region at which the AI-enabled diagnostic systems are tested on patients visiting the local clinics. An auxiliary nurse uses an AI-enabled camera to capture six retinal images (three per eye) of each patient for the image recognition algorithm to detect any diabetic-retinopathy-related abnormalities, and this patient medical data is sent to the base hospital to train the AI algorithm. At the base hospital, a technical image grader manually grades the received images as per standard clinical guidelines and compares the manual diagnosis to the automated diagnosis of the algorithm.

References

- Achuthan, Asha. 2011. *Re:wiring Bodies*. Bangalore: Centre for Internet & Society.
- ACM (Association for Computing Machinery). 2021. “AI on the Ground Approach: Critical Methodological Reflections and Lessons from the Field.” Streamed live March 4, 2021. *YouTube*. <https://www.youtube.com/watch?v=MhpzTDyopll>.
- Adam, Alison. 2000. “Deleting the Subject: A Feminist Reading of Epistemology in Artificial Intelligence.” *Minds and Machines* 10 (2): 231–53. <https://doi.org/10.1023/A:1008306015799>.
- Al Jazeera. 2011. Outsourced: Clinical Trials Overseas. Documentary video posted online, July 11, 2011. <https://www.aljazeera.com/program/fault-lines/2011/7/11/outsourced-clinical-trials-overseas>.
- Atanasoski, Neda, and Vora, Kalindi. 2019. *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures*. Durham, NC: Duke University Press.
- Burt, Andrew, and Volchenboum, Samuel. 2018. “How Health Care Changes When Algorithms Start Making Diagnoses.” *Harvard Business Review*, May 8, 2018. https://hbr.org/2018/05/how-health-care-changes-when-algorithms-start-making-diagnoses?mc_cid=944a576258&mc_eid=92f95b707d.
- Chandra Talpade Mohanty. 2003. “‘Under Western Eyes’ Revisited: Feminist Solidarity through Anticapitalist Struggles.” *Signs: Journal of Women in Culture and Society* 28 (2): 499–535. <https://doi.org/10.1086/342914>.
- Department of Industrial Policy and Promotion. 2018. *Report of the Artificial Intelligence Task Force*. <https://dipp.gov.in/whats-new/report-task-force-artificial-intelligence>.
- Elson, Diane, and Ruth Pearson. 1981. “‘Nimble Fingers Make Cheap Workers’: An Analysis of Women’s Employment in Third World Export Manufacturing.” *Feminist Review* 7 (1): 87–107. <https://doi.org/10.1057/fr.1981.6>.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Gadre, Arun, and Abhay Shukla. 2016. *Dissenting Diagnosis: Voices of Consciousness from the Medical Profession*. Random House India. E-book.

Gates Foundation. 2019. "Wadhvani Institute for Artificial Intelligence Foundation." <https://www.gatesfoundation.org/about/committed-grants/2019/11/invo03758>.

Google. n.d. "AI for Social Good." Accessed March 30, 2021. <https://ai.google/social-good/>.

Gulshan, Varun, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, et. al. 2016. "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs." *JAMA* 316 (22): 2402–10. <https://doi.org/10.1001/jama.2016.17216>.

Hao, Karen. 2019. "Five Questions You Can Use to Cut through AI Hype." *MIT Technology Review*, May 15, 2019. <https://www.technologyreview.com/s/613535/five-questions-you-can-use-to-cut-through-ai-hype/>.

Harding, Sandra. 1986. *The Science Question in Feminism*. Ithaca, NY: Cornell University Press.

———. 1992. "Rethinking Standpoint Epistemology: What Is 'Strong Objectivity?'" *The Centennial Review* 36 (3): 437–70. <https://www.jstor.org/stable/23739232>.

IBM. n.d. "Science for Social Good." Accessed March 30, 2021. <https://www.research.ibm.com/science-for-social-good/>.

Intel. n.d. "AI for Social Good." Accessed March 30, 2021. <https://www.intel.com/content/www/us/en/artificial-intelligence/ai4socialgood.html>.

Jesani, Amar. 1990. "Limits of Empowerment: Women in Rural Health Care." *Economic and Political Weekly* 25 (20): 1098–1103. <https://www.jstor.org/stable/4396290>.

Keyes, Os. 2018. The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. *Proceedings of the ACM on Human–Computer Interaction* 2 (CSCW), article no. 88, 1–22. <https://doi.org/10.1145/3274357>.

Khera, Reetika. 2017. "Impact of Aadhaar in Welfare Programmes." *Economic and Political Weekly* 52 (50). <https://www.epw.in/journal/2017/50/special-articles/impact-aadhaar-welfare-programmes.html>.

Lecher, Colin. 2018. "A Healthcare Algorithm Started Cutting Care, and No One Knew Why." *The Verge*, March 21, 2018. <https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy>.

Longino, Helen. 1987. "Can There Be a Feminist Science?" *Hypatia* 2 (3): 51–64. <https://www.jstor.org/stable/3810122>.

Maslow, Abraham. 1966. *The Psychology of Science: A Reconnaissance*. New York: Harper & Row.

Microsoft. n.d. "AI for Good with Microsoft Artificial Intelligence." Accessed March 30, 2021. <https://www.microsoft.com/en-us/ai/ai-for-good>.

Microsoft News Centre India. 2017. "Government of Telangana Adopts Microsoft Cloud and Becomes the First State to Use Artificial Intelligence for Eye Care Screening for Children." *Microsoft*, August 3, 2017. <https://news.microsoft.com/en-in/government-telangana-adopts-microsoft-cloud-becomes-first-state-use-artificial-intelligence-eye-care-screening-children/>.

Ministry of Health and Family Welfare, Government of India. 2017. Rural Health Statistics 2016–17. <https://www.nrhm-mis.nic.in/>.

Moitra, Aparna, Vishnupriya Das, Gram Vaani, Archana Kumar, and Aaditeshwar Seth. 2016. "Design Lessons from Creating a Mobile-Based Community Media Platform in Rural India." *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development*, article no. 14, 1–11. <https://doi.org/10.1145/2909609.2909670>.

Mol, Annemarie. 2008. *The Logic of Care: Health and the Problem of Patient Choice*. New York: Routledge.

Nandy, Ashis. 1988. "Introduction: Science as a Reason of State." In *Science, Hegemony and Violence: A Requiem for Modernity*, edited by Ashis Nandy, 1–23. Oxford: Oxford University Press.

NITI Aayog. 2018. *National Strategy for Artificial Intelligence*. https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf.

NITI Aayog. 2021. *Responsible AI. #AIForAll. Approach Document for India Part 1 – Principles for Responsible AI*. <https://niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf>.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.

Paul, Yesha, Elonnai Hickok, Amber Sinha, Udbhav Tiwari. 2018. *Artificial Intelligence in the Healthcare Industry in India*. Centre for Internet & Society. <https://cis-india.org/internet-governance/files/ai-and-healthcare-report>.

Petchesky, Rosalind Pollack. 1987. "Fetal Images: The Power of Visual Culture in the Politics of Reproduction." *Feminist Studies* 13 (2): 263–92. <https://doi.org/10.2307/3177802>.

Pfeffer, Naomi. 1987. "Artificial Insemination, In-Vitro Fertilization and the Stigma of Infertility." In *Reproductive Technologies: Gender, Motherhood and Medicine*, edited by Michelle Stanworth, 81–97. Cambridge, UK: Polity Press.

Prasad, Purendra. 2007. "Medicine, Power and Social Legitimacy: A Socio-Historical Appraisal of Health Systems in Contemporary India." *Economic and Political Weekly* 42(34): 3491-3498. <https://www.jstor.org/stable/4419944>.

Radhakrishnan, Radhika, and Amber Sinha. 2020. "Towards Algorithmic Transparency." Policy brief. Centre for Internet & Society, New Delhi. <https://cis-india.org/internet-governance/algorithmic-transparency-pdf>.

Radhakrishnan, Radhika. Forthcoming. "Health Data as Wealth: Understanding Patient Rights in India through a Feminist Digital Ecosystems Approach." Data Governance Network, Mumbai.

Rajan, Kaushik Sunder. 2005. "Subjects of Speculation: Emergent Life Sciences and Market Logics in the United States and India." *American Anthropologist* 107 (1): 19–30. <https://doi.org/10.1525/aa.2005.107.1.019>.

Read, Clancy, Jaya Earnest, Mohammed Ali, and Veena Poonacha. 2014. "Applying a Practical, Participatory Action Research Framework for Producing Knowledge, Action and Change in Communities: A Health Case Study from Gujarat, Western India." In *M2 Models and Methodologies for Community Engagement*, edited by Reena Tiwari, Marina Lommerse, and Dianne Smith, 91–105. Singapore: Springer.

Stanworth, Michelle, ed. 1987. *Reproductive Technologies: Gender, Motherhood and Medicine*. Cambridge, UK: Polity Press.

Taddeo, Mariarosaria, and Luciano Floridi. 2018. "How AI Can Be a Force for Good." *Science* 361 (6404): 751–52. <https://doi.org/10.1126/science.aat5991>.

Wadhvani AI. n.d. "Artificial Intelligence for Social Good." Accessed March 30, 2021. <https://www.wadhwaniai.org/>.

———. n.d. "Maternal, Newborn, and Child Health." Accessed March 30, 2021. <https://www.wadhwaniai.org/work/maternal-newborn-child-health/>.

Wajcman, Judy. 1991. *Feminism Confronts Technology*. University Park, PA: Penn State Press.

West, Sarah Myers, Meredith Whittaker, and Kate Crawford. 2019. "Discriminating Systems: Gender, Race and Power in AI." *AI Now Institute*. <https://ainowinstitute.org/discriminatingystems.html>.

Winner, Langdon. 1980. "Do Artifacts Have Politics?" *Daedalus* 109 (1): 121–36. <https://www.jstor.org/stable/20024652>.

Author Bio

Radhika Radhakrishnan is a qualitative researcher currently at the World Wide Web Foundation as a Gender Research Manager. With interdisciplinary academic training in Gender Studies and Computer Science Engineering, her research focuses on understanding the challenges faced by gender-minoritized communities with digital

technologies in India and finding entry points to intervene meaningfully using feminist methodologies.