Ethical Evaluation of AI Deployment in Low-Resource Healthcare Systems

Tresna Dey

Abstract

As Bangladesh accelerates toward its ambitious "Smart Bangladesh 2041" vision, the stethoscope is no longer the sole symbol of medical care; it is increasingly augmented by the algorithm. AI-powered diagnostic tools—from portable X-rays for Tuberculosis (TB) to AI-assisted clinical decision apps—are making their way into remote Upazilas and community health centers, offering the promise of faster, more accurate, and wider-reaching healthcare. By 2026, these tools are no longer experimental; they are operational components of Bangladesh's underfunded healthcare system.

Yet this integration presents a critical ethical crossroads. In settings where resources are scarce, deploying AI carries hidden risks. The same technology that promises efficiency and wider coverage can inadvertently entrench inequalities, favoring those already better served while marginalizing the most vulnerable. Ethical considerations—particularly those around fairness, accountability, transparency, and human oversight—must therefore guide every deployment.

The first ethical challenge is the invisibility trap, which manifests as a data fairness crisis. AI systems are only as unbiased as the datasets on which they are trained, and Bangladesh's digital health footprint remains fragmented. Globally, most AI models are trained on datasets from high-income countries, with patient profiles, disease prevalence, and environmental factors that differ significantly from local realities. For instance, cardiovascular risk calculators calibrated on European cohorts systematically underestimate risk for South Asian populations. When applied without adjustment, these algorithms risk misdiagnosis, delayed treatment, or inappropriate interventions. Moreover, if models are primarily trained on urban hospital data—such as private hospitals in Dhaka—they will be "blind" to the clinical patterns of patients in rural regions like Kurigram or remote riverine chars. This creates a form of digital exclusion, whereby AI tools systematically favor urban, affluent populations while leaving the rural poor underrepresented. Global studies indicate that biased AI algorithms can produce a 20–30% reduction in diagnostic accuracy for

underrepresented populations, transforming a potentially life-saving tool into a source of systematic error. In Bangladesh, where health disparities are already stark, this bias is not merely technical—it is a moral issue.

Closely related is the risk of over-reliance on AI in under-resourced facilities, a phenomenon known as automation bias. Government hospitals in Bangladesh are often severely understaffed, with a single physician managing 100 or more patients in a day. In such environments, there is an understandable temptation to trust machine recommendations over human judgment. While AI can improve throughput and reduce diagnostic delays, excessive reliance may erode the clinician's own decision-making skills. This creates a dual problem: first, the accountability vacuum. If an AI-powered triage tool misclassifies a critical pediatric case, who bears responsibility? Is it the software developer, the overburdened nurse, or the Ministry of Health? Current policy frameworks in Bangladesh, including the draft National AI Policy of 2026, are still grappling with such liability gaps, leaving frontline providers vulnerable. Second, there is the risk of deskilling. Repeated dependence on automated interpretation—such as X-ray reading or preliminary triage—may gradually diminish clinicians' core skills. In the event of power outages, server failures, or system downtime, healthcare providers could find themselves unprepared to deliver accurate care, creating a potentially catastrophic vulnerability in the system.

To mitigate these risks, ethical AI deployment in low-resource healthcare must embrace the "Human-in-the-Loop" principle. AI should function as a decision-support tool, not a decision-maker. Explainable AI (XAI) is central to this approach. Algorithms must provide interpretable outputs, allowing clinicians to understand why a patient was flagged for a condition such as pneumonia or anemia. This transparency builds trust and enables critical oversight. Furthermore, the "Stop Button" principle must be institutionalized: any healthcare professional should have the authority—and legal protection—to override AI recommendations when they conflict with clinical judgment. Human oversight is not a redundancy; it is a safeguard against the unintended consequences of automation.

Governance structures play a critical role in ensuring ethical deployment. Bangladesh is laying the groundwork through initiatives such as the National Strategy for Artificial Intelligence (2022–2026) and the forthcoming Artificial Intelligence Act (expected 2028). For healthcare, these policies highlight four critical priorities. First, algorithmic impact assessments (AIA) are essential. Before deployment in public hospitals, AI tools must undergo equity audits to ensure they do not inadvertently discriminate based on gender, age, socioeconomic status, or geographic location. Second, data sovereignty is a national imperative. Protecting the health records of 170 million citizens from exploitation by foreign entities is both a privacy and national security concern. Third, digital health literacy must become part of professional training. Clinicians need a baseline understanding of AI ethics, model limitations, and audit procedures. Fourth, continuous monitoring and feedback mechanisms must be built into AI systems, ensuring that models adapt to local realities over time, rather than remaining static reflections of biased historical data.

The ethical stakes extend beyond technical accuracy to societal trust and equity. If AI is perceived as opaque or elitist, it risks reinforcing existing disparities. Communities may resist AI-assisted care, particularly if it seems to prioritize urban centers or private hospitals over local clinics. Conversely, ethical deployment—grounded in fairness, transparency, and human oversight—can extend specialist-level diagnostic capabilities to rural Upazilas, riverine chars, and marginalized communities previously excluded from high-quality care. The goal is not simply to digitize healthcare but to democratize it, ensuring that technological advancement benefits all segments of society, rather than entrenching privilege.

In addition, ethical AI in healthcare must consider the unintended consequences of algorithmic decision-making. Predictive models, if applied without nuance, can perpetuate stereotypes. For example, an AI system predicting disease prevalence based on neighborhood data might inadvertently stigmatize certain areas or communities, discouraging residents from seeking care. Similarly, overemphasis on efficiency—such as prioritizing cases with higher predicted survival—can marginalize patients with chronic or complex conditions. Ethical frameworks must therefore incorporate mechanisms to assess both direct and indirect harms, balancing efficiency with equity.

Practical strategies for ethical deployment include integrating local datasets into model training, continuously validating algorithms against real-world outcomes, and establishing multi-stakeholder oversight committees. Clinicians, patients, AI developers, and policymakers should participate in regular audits to ensure that AI recommendations align with ethical norms and clinical realities. By embedding feedback loops, AI systems become adaptive partners in care rather than rigid authorities, capable of evolving with the local healthcare landscape.

Finally, ethical evaluation must address the human experience. Healthcare is inherently relational, relying on empathy, trust, and human judgment. AI can augment but not replace these dimensions. Ensuring that AI interventions do not depersonalize care is essential. Training modules for healthcare providers should emphasize collaboration with AI, ethical decision-making, and communication with patients about how technology informs their treatment. Patients, in turn, should be informed about the role and limitations of AI in their care, fostering a culture of informed consent and mutual understanding.

In conclusion, AI in Bangladesh’s healthcare system is neither a panacea nor a risk-free innovation; it is a tool with profound potential and sharp ethical edges. Its deployment in 2026 presents both an unprecedented opportunity to extend specialist-level care to marginalized populations and a real risk of widening the gap between urban and rural, rich and poor, digitally included and excluded. Responsible integration requires rigorous attention to fairness, transparency, accountability, and human oversight. By ensuring that AI serves as a decision-support system rather than a substitute for human judgment, by auditing algorithms for bias, and by cultivating digital health literacy among clinicians, Bangladesh can harness AI to reduce disparities rather than deepen them. In an era where algorithms increasingly shape

health outcomes, ethical stewardship is not optional; it is the ultimate operating system guiding Smart Bangladesh toward a future in which healthcare is equitable, accessible, and human-centered.