# Can AI Make Fair Triage Decisions? Navigating Algorithmic Bias in Digital Health

Rasit Dinc

*Rasit Dinc Digital Health & AI Research*

Published: January 24, 2022 | Digital Therapeutics

DOI: 10.5281/zenodo.17998046

## Abstract

The integration of Artificial Intelligence AI into healthcare is rapidly transforming clinical workflows, with one of the most critical applications being me...

The integration of Artificial Intelligence (AI) into healthcare is rapidly transforming clinical workflows, with one of the most critical applications being **medical triage**. AI-driven systems promise to enhance efficiency, reduce human error, and standardize patient prioritization, particularly in high-pressure environments like emergency departments. However, the central question remains: Can AI truly make **fair triage decisions**, or does it merely automate and amplify existing human and systemic biases? This is a critical challenge for the future of **digital health** and a major focus for researchers in **healthcare AI**.

## The Promise and Peril of AI in Triage

AI's potential in triage is undeniable. By analyzing vast datasets of patient history, symptoms, and outcomes, algorithms can often predict the severity of a patient's condition with greater speed and consistency than human clinicians. This capability is particularly valuable in mass casualty events or overwhelmed healthcare systems, where rapid, consistent **clinical decision-making** is paramount. Studies, such as those reviewed in *The Role of Artificial Intelligence in Enhancing Triage Decisions*, suggest that AI can significantly reduce variability in human triage, which is a key step toward more **equitable care**. The consistency offered by AI contrasts sharply with the known cognitive biases that can affect human triage nurses, such as anchoring bias or availability heuristic.

However, the very foundation of these systems—the data—is also their greatest vulnerability. AI models are trained on historical patient data, which often reflects decades of systemic inequalities in healthcare access and treatment. If the training data contains a disproportionate number of cases from certain demographic groups, or if it reflects historical under-treatment of specific populations (e.g., racial or socioeconomic minorities), the resulting AI model will inevitably inherit and perpetuate these biases. This is the core

problem of **algorithmic bias** in healthcare, a phenomenon where the model's output systematically disadvantages certain groups.

## The Challenge of Algorithmic Bias in Clinical Decision-Making

**Algorithmic bias** in triage can manifest in dangerous ways, leading to disparities in patient care and potentially life-threatening outcomes. For instance, an AI system trained on data where certain symptoms in one demographic were historically downplayed or misdiagnosed might subsequently assign a lower urgency score to a patient from that same demographic, even if their condition is critical. This is not a failure of the algorithm's logic, but a reflection of the flawed human data it was fed. The algorithm is simply optimizing for the historical, and often biased, outcome.

A 2023 review published in *PLOS Digital Health* highlighted that bias can arise at every stage of the AI development pipeline, from data collection and labeling to model deployment and evaluation. The review emphasizes that the goal is not just to create an accurate model, but an **equitable AI** model that performs equally well across all patient populations. The economic and ethical implications of these biases are profound, as they can lead to delayed treatment, poorer outcomes, and a further erosion of trust in the healthcare system. The challenge is to ensure that AI-driven tools enhance, rather than undermine, the principle of distributive justice in medicine, which dictates that resources should be allocated fairly.

## Strategies for Achieving Equitable AI and Mitigating Bias

Achieving **AI fairness** in triage requires a multi-pronged approach focused on transparency, data quality, and continuous monitoring:

1. **Data Curation and Auditing:** Developers must actively seek out and integrate diverse, representative datasets. This includes auditing existing data for proxies of protected attributes (like race or socioeconomic status) that could inadvertently lead to biased outcomes. Techniques like re-weighting or oversampling underrepresented groups are crucial for mitigating this initial data bias, ensuring the model learns from a complete and balanced picture of the patient population. 2. **Algorithmic Transparency (Explainable AI - XAI):** Increasing the interpretability of AI models allows clinicians and regulators to understand *why* a specific triage decision was made. **XAI** techniques provide a window into the model's reasoning, making it easier to spot and correct biased outputs before they impact patient care. This transparency is vital for building trust among both clinicians and the public, moving away from the "black box" problem. 3. **Human-in-the-Loop (HITL) Systems:** AI should function as a decision-support tool, not a replacement for human judgment. The final triage decision must remain with a trained clinician who can contextualize the AI's recommendation and override it if necessary, especially in cases where bias is suspected or the patient presents with atypical symptoms. This hybrid approach leverages the speed of AI with the ethical reasoning of a human expert, creating a necessary safeguard against algorithmic error.

The journey toward truly fair **healthcare AI** is complex, requiring a collaborative effort between data scientists, clinicians, ethicists, and policymakers. The technology holds immense promise, but its ethical deployment hinges on our ability to confront and correct the biases embedded in our historical data.

For more in-depth analysis on the ethical frameworks and practical strategies for deploying AI in **digital health**, the resources and expert commentary at [www.rasitdinc.com](www.rasitdinc.com) provide essential professional insight into the future of **clinical decision-making** and **healthcare AI**.

## Conclusion

The answer to whether AI can make fair triage decisions is a cautious "yes," but only if we design, train, and deploy these systems with an unwavering commitment to equity. AI is a powerful mirror reflecting the biases of the human systems that created it. By proactively addressing **algorithmic bias** and prioritizing **AI fairness** alongside accuracy, we can harness AI to build a more equitable and efficient future for **medical triage** and **digital health** as a whole. The ethical imperative is clear: the pursuit of efficiency must not come at the expense of equity, ensuring that the benefits of this powerful technology are distributed justly across all populations.

**Keywords:** AI fairness, medical triage, algorithmic bias, digital health, healthcare AI, equitable AI, clinical decision-making, XAI.

---