



Bioethics

ISSN: 2989-5448 (Online)

Publisher: Tomorrow's Research Publications (TRP)

An International Peer-Reviewed Open Access Journal

Policy Analysis

Equity Considerations in Artificial Intelligence-Driven Triage Systems

Stefan Marques^{1*}, **Pierre Silva²** and **Helen Andersson³**¹Department of Bioethics, Danube University Center for Medical Law and Ethics, Hungary²Center for Medical Ethics and Law, Northmoor School of Public Health Ethics, United Kingdom³Department of Bioethics, Northmoor School of Public Health Ethics, United Kingdom

ABSTRACT

Artificial intelligence and algorithmic tools increasingly determine the order in which patients are seen, escalated, or allocated scarce resources such as ventilators, raising the question of whether these tools reproduce the very inequities they are often marketed as correcting.

This policy analysis reviews three documented mechanisms of bias in AI-driven and algorithmic triage: cost-based proxy labels that understate need in historically underserved populations, pulse-oximetry measurement error that masks hypoxemia in patients with darker skin, and crisis-standards-of-care scoring rubrics that have been challenged as discriminatory toward people with disabilities.

Each mechanism enters the triage pipeline at a different point -- training data, point-of-care instrumentation, and protocol design -- which is why no single technical fix addresses all three simultaneously.

We examine the governance responses that have followed, including regulatory inquiries into algorithmic bias, disclosure requirements for AI-enabled medical devices, and calls for continuous post-deployment auditing rather than one-time certification.

We conclude that defensible AI-driven triage requires equity to be treated as a design specification from the outset, sustained subgroup-specific monitoring after deployment, and explicit acknowledgment that some triage decisions encode value judgments that no algorithm can neutrally resolve.

Keywords: artificial intelligence; triage; health equity; algorithmic bias; healthcare rationing

Citation: Marques S, Silva P and Andersson H (2025) Equity Considerations in Artificial Intelligence-Driven Triage Systems. *Bioethics*. 1(1):01-05.

Corresponding Author: Stefan Marques

E-mail Address: s.marques@ducmlle.hu

Article History: Received 18 December 2024 | Accepted 25 March 2025 | Published 29 June 2025

Copyright: © 2025 Stefan Marques, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (<https://creativecommons.org/licenses/by/4.0/>).

INTRODUCTION

Triage has always been a rationing exercise dressed up as a clinical judgment: someone, under time

pressure and with incomplete information, decides who gets seen first, who waits, and who gets referred onward. What has changed in the last decade is who, or what, is making that call. Emergency departments,

primary-care telehealth platforms, and population-health programs increasingly route patients through machine-learning models that predict acuity, deterioration risk, or resource need, and use that prediction to decide queue position [1][2]. The appeal is obvious -- consistency, speed, and the promise of catching risk that a rushed clinician might miss. The problem is that these systems learn their notion of "risk" from historical data generated inside a healthcare system that has never allocated care equally across racial, socioeconomic, and disability lines, and a model trained on unequal patterns of care will, left unexamined, reproduce and sometimes amplify them [3].

This is not a hypothetical concern raised by critics unfamiliar with the technology. It is now a documented, repeated finding across several distinct triage-adjacent domains: population-health risk scores, pulse oximetry-dependent deterioration algorithms, and pandemic-era ventilator allocation protocols. Each surfaced a version of the same structural issue -- a proxy variable or a scoring rule that looked neutral on its face but tracked race, disability status, or prior access to care instead of the clinical need it claimed to measure. This paper reviews that evidence, the mitigation strategies proposed in response, and the harder governance question of who is accountable when an algorithm, rather than a clinician, makes the first cut.

THE PROXY PROBLEM: LESSONS FROM POPULATION-HEALTH ALGORITHMS

The clearest and most cited demonstration of algorithmic bias in a triage-adjacent tool remains the 2019 analysis of a widely used commercial population-health algorithm that hospitals and insurers relied on to identify patients for enrollment in intensive care-management programs [3]. The algorithm ranked patients by predicted future healthcare costs, on the reasonable-sounding assumption that patients likely to cost more are also patients with greater clinical need. But cost and need diverge sharply along racial lines, because Black patients in the underlying data had historically received less care -- and therefore generated lower costs -- for a given level of illness, whether from unequal access, distrust of the system, or clinician bias in referral patterns. The algorithm faithfully learned that pattern and reproduced it forward: at any given predicted-risk score, Black patients in the study population were substantially sicker than white patients, and correcting the proxy would have more than doubled the fraction of Black patients identified for extra care.

The lesson generalized well beyond that one algorithm because the underlying mechanism -- optimizing for a measurable proxy that correlates with, but is not identical to, the outcome a triage system actually cares about -- recurs in essentially every algorithmic triage tool that uses historical utilization, cost, or

documentation patterns as training labels [4][5]. Commentators have since argued that this is not a bug correctable by better statistics alone, but a structural feature of building predictive tools on top of a healthcare system whose baseline delivery of care is itself unequal; a model can be perfectly well-calibrated in a narrow technical sense and still allocate care unfairly, because calibration measures agreement with a biased ground truth rather than agreement with actual need [6][7].

MEASUREMENT BIAS AT THE POINT OF CARE: PULSE OXIMETRY AND OCCULT HYPOXEMIA

A second, mechanically different but ethically related problem emerged prominently during the COVID-19 pandemic: pulse oximeters, the fingertip devices used throughout emergency triage to estimate blood oxygen saturation, are measurably less accurate in patients with darker skin pigmentation, tending to overestimate true oxygen saturation and mask "occult hypoxemia" that would otherwise trigger escalation of care [8]. Because so many triage protocols, and the software layered on top of them, use a pulse-oximetry reading as an input feature or hard cutoff for eligibility -- for supplemental oxygen, for hospital admission, for COVID-19 antiviral treatment during periods of scarcity -- an instrument-level measurement bias propagates directly into algorithmic and human triage decisions alike.

Retrospective cohort analyses conducted after this problem gained attention estimated that thousands of Black and Hispanic patients experienced clinically meaningful delays in COVID-19 treatment eligibility because their true hypoxemia was masked by an inflated pulse-oximetry reading, a gap large enough to matter at the population level even though each individual discrepancy looks modest [9][10]. What makes this case ethically distinct from the population-health algorithm above is that the bias originates in hardware and physics -- how light absorption varies with skin pigmentation -- rather than in a choice of training label. That distinction matters practically: it means the fix cannot simply be "audit the algorithm's outputs," since the corrupted signal enters upstream of any triage model's own logic. Any AI triage tool that ingests oximetry readings as a feature inherits this bias whether or not its own training process was scrupulous.

VENTILATOR ALLOCATION, DISABILITY, AND THE ETHICS OF EXPLICIT RATIONING

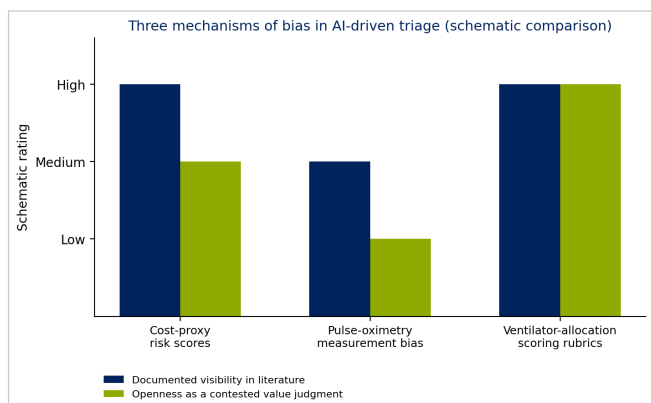
Table 1 summarizes the three distinct mechanisms of bias reviewed here, and Figure 1 offers a schematic comparison of how visible, and how contested, each mechanism has been in the recent literature. The third and most explicitly contested domain is pandemic-era crisis-standards-of-care protocols, which used scoring systems -- often built around the Sequential Organ

Failure Assessment (SOFA) score or similar acuity measures -- to decide which patients would receive a ventilator when demand outstripped supply. Disability rights advocates argued, in several formal civil-rights complaints filed in 2020, that many state triage protocols effectively deprioritized patients with chronic disabilities or lower baseline functional status, treating pre-existing disability as evidence of poor prognosis in a way that conflated disability with reduced quality of life rather than with actual likelihood of surviving the acute illness [11].

Table 1. Three mechanisms of bias documented in AI-driven and algorithmic triage tools, where each enters the pipeline, and the mitigation strategies proposed for it.

Mechanism	Point of entry	Example	Proposed mitigation
Biased training-label proxy	Model design / training data	Cost-based population-health risk scores underestimating need in Black patients	Replace cost proxies with direct measures of illness burden; subgroup-specific recalibration
Biased physical measurement	Point-of-care instrument	Pulse oximeters overestimating oxygen saturation in patients with darker skin	Device redesign and validation across skin tones; clinical override of borderline readings
Value-laden scoring rubric	Protocol / rule design	SOFA-based ventilator allocation scores treating disability as poor prognosis	Transparent, publicly deliberated rubrics; disability-inclusive review of scoring criteria

Figure 1. Schematic comparison of how visible each bias mechanism has been in the recent literature and how openly it has been treated as a contested value judgment rather than a technical error (illustrative ratings, not derived from a single empirical study).



Unlike the two mechanisms above, ventilator-allocation scoring is not typically framed as an AI or machine-learning system in the narrow sense -- most protocols were rule-based scoring rubrics designed by clinician committees, not learned models -- but the ethical structure of the problem is identical: an ostensibly objective, quantitative scoring tool embeds

contestable value judgments about whose survival counts, and it does so with a false appearance of neutrality that can insulate those judgments from scrutiny. Simulation studies conducted after the fact found that small design choices in how a protocol weighted short-term survival probability against long-term life-years substantially changed which patient subgroups were disadvantaged, and that no single scoring rule satisfied every ethical principle stakeholders cared about simultaneously -- utilitarian survival-maximization, first-come-first-served fairness, and protection of the worst-off all pulled in different directions [12][13]. Retrospective ethical analyses converged on a similar conclusion: heuristics that seemed defensible in the abstract produced allocations that few of their designers would have endorsed once the demographic consequences were made visible [14].

TOWARD GOVERNANCE: AUDITING, ACCOUNTABILITY, AND THE LIMITS OF TECHNICAL FIXES

The response to these findings has increasingly moved from technical bias-mitigation -- reweighting training data, choosing fairness-constrained loss functions, recalibrating thresholds by subgroup -- toward a broader governance argument that no purely technical fix can substitute for institutional accountability [4]. Regulatory attention has followed: state attorneys general have opened inquiries into whether commercial healthcare algorithms violate civil-rights and consumer-protection law [15], and federal regulators have begun requiring more disclosure from AI-enabled medical device manufacturers about the demographic composition of their validation data, though recent analyses of those disclosures describe a persistent "fairness paradox" in which summaries that are technically compliant still omit the subgroup performance breakdowns clinicians would need to detect disparate impact in practice [16].

A recurring theme across this governance literature is that auditing an algorithm once, at deployment, is insufficient, because the populations a triage tool serves, the clinical practices that generate its input data, and the tool's own outputs (once clinicians start trusting and acting on them) all shift over time, a phenomenon sometimes called automation feedback [17][18]. Proposed responses include mandatory post-deployment monitoring for subgroup performance drift, independent third-party audits rather than manufacturer self-certification, and meaningful clinician override mechanisms that are actually used rather than merely available on paper, since evidence from primary-care and emergency-department triage tools suggests clinicians often defer to algorithmic recommendations even when their own judgment points the other way [19][20]. Other commentators emphasize that equity has to be a design specification from the outset -- built into the choice of training labels,

the composition of validation cohorts, and the metric used to define "acceptable" performance -- rather than an audit applied retroactively to a system optimized for something else [21][22].

CONCLUSION

Three distinct mechanisms -- a biased proxy label, a biased physical measurement, and an explicitly value-laden scoring rubric -- have each produced documented racial or disability-based disparities in systems that decide who gets care first. They differ in where the bias enters the pipeline, which is precisely why no single technical intervention addresses all three: recalibrating a cost-prediction algorithm does nothing for a pulse oximeter's physics, and neither does anything for a ventilator-allocation heuristic's implicit value judgments about disability and quality of life. What unites them ethically is that each mechanism dressed a contestable allocation decision in the appearance of neutral measurement, which is exactly what made the resulting disparities hard to see until researchers went looking. Meaningful equity in AI-driven triage will likely require sustained, subgroup-specific auditing built into deployment rather than a one-time certification, transparency requirements strong enough to let outside researchers actually find the next Obermeyer-style proxy problem before it causes years of harm, and an explicit acknowledgment that some triage decisions encode value judgments no algorithm can neutrally resolve on a clinician's or regulator's behalf.

DECLARATIONS

Acknowledgements

The authors are grateful to colleagues in their respective bioethics and public health ethics departments for helpful conversations that sharpened the framing of this review.

Funding

This review received no external funding.

Conflict of Interest

The authors declare no conflicts of interest relevant to this work.

Author Contributions

Stefan Marques conceived the review, drafted the initial manuscript, and led the synthesis on population-health algorithms and governance responses. Pierre Silva performed the literature review on pulse oximetry and measurement bias and critically revised the manuscript. Helen Andersson contributed the analysis of ventilator-allocation ethics and disability discrimination and provided final approval of the submitted version. All authors read and approved the final manuscript.

Ethics Statement

As a literature review synthesizing previously published, publicly available research and policy analysis, this article did not involve new data collection from human participants and did not require institutional ethics committee approval.

Data Availability

No new data were generated or analyzed for this review. All sources discussed are cited in the reference list and are publicly available through their respective publishers.

REFERENCES

1. Improving safety and quality of emergency care using machine learning-based clinical decision support at triage (2023) AHRQ Digital Healthcare Research.
2. AI triage in primary care: building safer and more equitable real-world evidence (2026) ScienceDirect.
3. Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, PMID:31649194.
4. Guiding principles to address the impact of algorithm bias on racial and ethnic disparities in health and health care (2023) National Academy of Medicine / PMC, PMID:38875319.
5. Impact of healthcare algorithms on racial and ethnic disparities in health and healthcare (2023) National Academies Press, NCBI Bookshelf.
6. Tackling racially biased health care algorithms (2023) Petrie-Flom Center, Harvard Law School.
7. Algorithms are making decisions about health care, which may only worsen medical racism (2021) American Civil Liberties Union.
8. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS (2020) Racial bias in pulse oximetry measurement. *New England Journal of Medicine*, DOI:10.1056/NEJMc2029240.
9. Fawzy A, Wu TD, Wang K, et al. (2022) Racial and ethnic discrepancy in pulse oximetry and delayed identification of treatment eligibility among patients with COVID-19. *JAMA Internal Medicine*, PMID:35639368.
10. Wong AI, et al. (2023) Racial disparities in pulse oximeter device inaccuracy and estimated clinical impact on COVID-19 treatment course. *American Journal of Epidemiology*.
11. Who Gets the Ventilator? Disability Discrimination in COVID-19 Medical-Rationing Protocols (2021) Yale Law Journal Forum.
12. White DB, et al. (2024) Investigating ethical tradeoffs in crisis standards of care through simulation of ventilator allocation protocols. *PLOS ONE*, PMID:39264928.

13. Persad G, et al. (2022) Fair allocation of vaccines, ventilators and antiviral treatments: leaving no ethical value behind in healthcare rationing. *Management Science*.
14. White DB, Lo B (2020) Ethical heuristics for pandemic allocation of ventilators across hospitals. *PMC*.
15. Attorney General Bonta launches inquiry into racial and ethnic bias in healthcare algorithms (2023) California Department of Justice, Office of the Attorney General.
16. The fairness paradox in public regulatory summaries of FDA-authorized artificial intelligence-enabled medical devices (2026) *Frontiers in Medicine*.
17. AI-driven healthcare: a survey on ensuring fairness and mitigating bias (2024) *arXiv*.
18. Bias in medical AI: algorithmic fairness and ethics challenges (2026) *Journal of Young Investigators*.
19. The benefit of the doubt phenomenon in emergency triage assignment disparities (2025) *PMC*.
20. AI and bias in emergency triage: uncovering disparities (2023) *HealthManagement.org*.
21. Advancing equity in healthcare AI: mitigating bias and reducing opacity (2026) *National High School Journal of Science*.
22. Addressing bias in clinical algorithms to advance health equity (2025) *American Federation of Teachers, Health Care*.