PPG Sensor Bias in Remote Patient Monitoring

Measurement Bias in Wearable PPG Sensors: Ethical and Policy Implications for AI-Enabled Remote Patient Monitoring

Akintoye Oyedola · School of Computing, University of North Florida · CNT 6167 Internet of Things

UNF UNIVERSITY of NORTH FLORIDA

Overview & Motivation

I survey how measurement bias in wearable photoplethysmography (PPG) sensors propagates through AI-enabled remote patient monitoring (RPM) systems to produce unequal clinical outcomes across demographic groups (see Figure 3). I examine the full causal chain from the optical physics of LED-based sensors through machine learning modeling practices that amplify this hardware-level error. For LED-based sensors, melanin absorption degrades SpO_2 signal quality in patients with darker skin tones. The resulting disparity is not an algorithmic artifact but a structural bias encoded in device design before any patient data is collected. This is confirmed by clinical evidence spanning from the COVID-19 pandemic through 2025 prospective studies reporting a 20% occult hypoxemia error rate in dark-skinned patients [3]. I critically evaluate existing U.S. and EU regulatory frameworks to identify the policy gap that currently allows biased PPG devices to reach the market without mandatory demographic validation requirements.

Categorization of Existing Work

The literature was organized into three functional groups that together trace a "data-to-policy'" pipeline (see Figure 1):

Group 1 — Clinical & Empirical Evidence of Bias
These studies establish that PPG measurement bias is real and clinically consequential, not merely theoretical. Coppetti et al. [1] characterize the optical physics: melanin absorbs light at 660 nm (red) and 940 nm (near-infrared), degrading the signal and causing SpO_2 overestimation. Sjoding et al. [8] provide landmark clinical evidence from a large hospital dataset, and the EquiOx study [3] confirms the bias persists in 2025-era hardware.

Group 2 — Technical & AI Modeling Literature
These sources explain how sensor-level error enters, travels, and compounds inside AI pipelines. Mehrabi et al. [5] survey existing studies to create a formal taxonomy distinguishing *measurement bias*, *representation bias*, *label bias*, and *algorithmic bias*; each requiring different remediation.

Group 3 — Policy, Ethics, & Industry Frameworks
These artifacts characterize the major policy framework for the PPG sensor bias issue and its current gaps. The FDA's 2021 Safety Communication [9] acknowledges skin tone bias explicitly but still produces no enforceable standard. The EU AI Act [2] imposes binding data-governance obligations and financial penalties for non-compliance, yet lacks a PPG-specific hardware validation requirement.

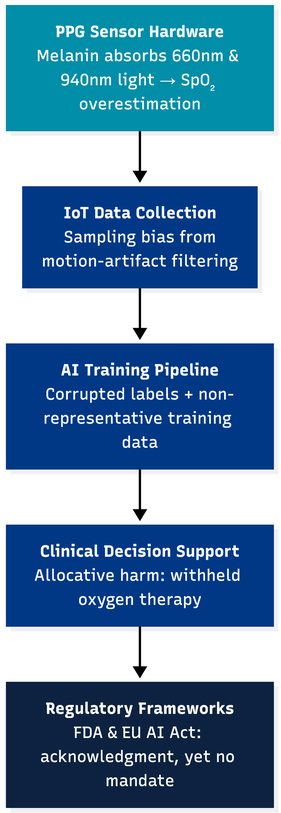


Figure 1: Data-to-Policy Bias Pipeline

Key Trends & Observations

Three dominant trends emerge consistently across the collected sources (see Figure 2):

Trend 1: Hardware-Level Bias Persists Over Time
Hardware-level SpO_2 overestimation in darker-skinned patients is consistent across a decade of PPG literature. This ranges from sensor physics characterizations in 2019 through the 2025 EquiOx prospective study [1, 3]. Ultimately, this indicates that recent hardware improvements have not addressed the fundamental optical calibration problem.

Trend 2: AI Pipelines Amplify Rather Than Correct the Bias
AI pipelines trained on non-representative datasets inherit and amplify this error at every stage. Biased oximeter readings produce corrupted training labels, and black-box deep learning architectures obscure the demographic source of prediction errors [5]. The harm manifests as allocative harm, described as supplemental oxygen withheld from hypoxic patients because neither the sensor nor the model flagged a problem. This pattern is documented at scale by Sjoding et al. and corroborated by Obermeyer et al. in adjacent risk-stratification systems [8, 6].

Trend 3: Regulatory Acknowledgment Without Enforcement
Regulatory responses have acknowledged the PPG sensor bias problem with no resolution. The FDA named skin tone bias explicitly in 2021 yet has issued no binding pre-market validation standard. Furthermore, the EU AI Act imposes enforceable data-governance obligations but contains no PPG-specific optical sensor requirement [9, 2].

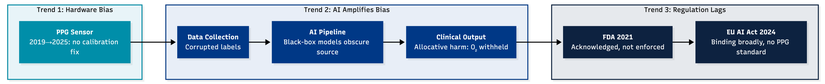


Figure 2: 3 trends mapped across the Data-to-Policy Pipeline

Summary & Recommendations For Future Work

Wearable PPG sensors carry a documented, physics-based measurement bias that current regulatory frameworks acknowledge but do not mandate manufacturers to correct before market access.

R1 — Mandate Fitzpatrick-Scale Hardware Validation
The FDA should finalize binding pre-market accuracy requirements for PPG devices specifying minimum performance thresholds across all Fitzpatrick categories, using arterial blood gas as ground truth rather than oximeter self-comparison.

R2 — Establish a Public PPG Fairness Benchmark
A publicly available, demographically annotated PPG dataset should be developed under NIH or equivalent sponsorship to enable reproducible, comparable evaluation of bias mitigation techniques across research groups.

R3 — Require Subgroup Performance Reporting
FDA and EU regulatory submissions for AI-enabled RPM tools should require disaggregated performance reporting by skin tone using standard fairness metrics, not aggregate accuracy alone.

R4 — Extend EU AI Act Scope to Sensor Hardware
The EU AI Act's high-risk medical AI classification should explicitly cover the sensor hardware feeding those systems, making optical calibration requirements enforceable alongside model governance obligations.

R5 — Adopt Wearable Data Consent Frameworks
Institutions deploying RPM systems that feed patient PPG data into AI training pipelines should implement explicit consent protocols disclosing this use and permitting opt-out without affecting clinical care.

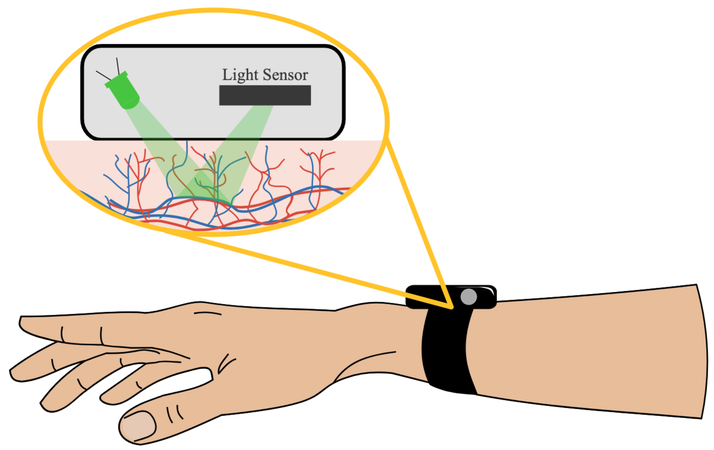


Figure 3: Photoplethysmography (PPG) Sensor on Skin.[4]

Identified Gaps

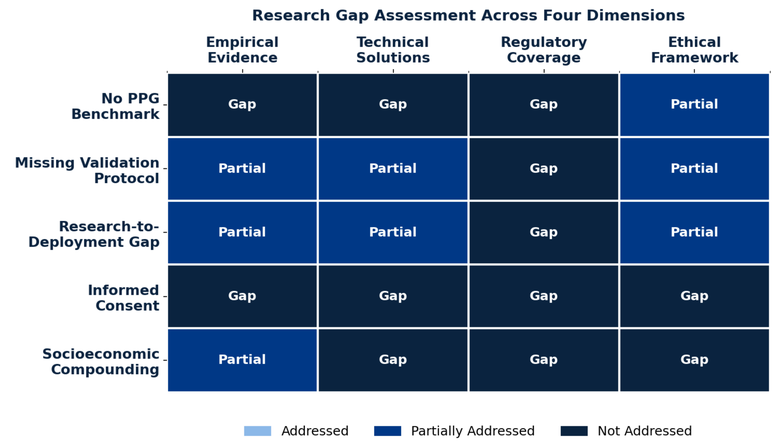
No Standardized PPG Fairness Benchmark
Unlike algorithmic fairness domains with established public datasets, no annotated PPG dataset with Fitzpatrick skin tone scores exists. This makes independent verification and cross-study comparison of bias mitigation techniques currently impossible.

Missing Pre-Market Validation Protocol
The FDA's 2023 discussion paper acknowledged the need for skin-tone-diverse testing but left unanswered the fundamental protocol questions: required sample sizes per Fitzpatrick category, testing conditions, and clinical-use-case-specific performance thresholds.

Research-to-Deployment Gap
Fairness interventions such as federated learning and post-hoc calibration have been evaluated almost exclusively in controlled settings, with little published evidence on real-world performance in live clinical workflows [7].

Informed Consent and Data Ownership
Patients whose wearable PPG data trains clinical AI models are rarely informed of this use, and consent frameworks governing the practice vary widely across institutions and jurisdictions with no PPG-specific standard in sight [7].

Socioeconomic Compounding
The patients most harmed by PPG measurement error are also the most likely to be monitored with lower-grade consumer devices, a compounding dynamic absent from both the technical literature and current regulatory frameworks. The gap assessment is summarized in Figure 4.



References

[1] Raffaele Coppetti et al. "Correction of SpO_2 Values Measured with a Pulse Oximeter in Patients with Dark Skin Pigmentation". In: *Journal of Clinical Monitoring and Computing* 33 (2019), pp. 609–616. DOI: 10.1007/s10877-018-0199-6.

[2] European Parliament and Council of the European Union. *Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence*. Official Journal of the European Union. 2024.

[3] Carolyn Hendrickson et al. "Pulse Oximeter Performance Varies by Skin Pigment: Results from the EquiOx Study". In: *American College of Cardiology Scientific Session*. Chicago, IL, Mar. 2025.

[4] KID-PPG. *How do smartwatches record your heart's activity?* URL: https://eslweb.epfl.ch/kid-ppg/what-is-ppg/.

[5] Ninareh Mehrabi et al. "A Survey on Bias and Fairness in Machine Learning". In: *ACM Computing Surveys* 54.6 (July 2021), pp. 1–35. DOI: 10.1145/3457607.

[6] Ziad Obermeyer et al. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations". In: *Science* 366.6464 (2019), pp. 447–453. DOI: 10.1126/science.aax2342.

[7] Jong Hyuk Park. "The Ethics of Using Artificial Intelligence in Medical Research". In: *Kosin Medical Journal* 39.4 (2024), pp. 255–264.

[8] Michael W. Sjoding et al. "Racial Bias in Pulse Oximetry Measurement". In: *New England Journal of Medicine* 383.25 (Dec. 2020), pp. 2477–2478. DOI: 10.1056/NEJMc2029240.

[9] U.S. Food and Drug Administration. *Pulse Oximeters: Recommendations for Patients, Health Care Providers, and Developers*. FDA Safety Communication. Feb. 2021. URL: https://www.fda.gov/medical-devices/safety-communications/pulse-oximeters-recommendations-patients-health-care-providers-and-developers.