Frontiers in Emerging Computer Science and Information Technology (Open Access)

Volume 01, Issue 02, August 2024,
Publish Date: 08-03-2024
PageNo.29-49

Bias Mitigation in Clinical AI: Auditing Race/Gender Disparities in Sepsis Prediction Models

Kawaljeet Singh Chadha
Business Analyst II MI, McLaren Health Care, TX, USA

Abstract

This paper exposes persistent race and gender biases in AI-based sepsis prediction models, arguing that these inequities undermine patient outcomes and demanding prioritization of fairness as a core clinical metric. An audit of multiple AI tools in urban hospitals revealed consistent accuracy gaps, notably more false negatives for Black, Hispanic, female, and non-binary patients, which delayed care and worsened clinical results. These disparities stem from imbalanced training data, distance miscalibration, and structural inequities embedded in clinical practice. The manuscript surveys algorithmic bias types, presents audit frameworks (Fairlearn, Aequitas), and evaluates mitigation strategies such as data rebalancing, fair regularization, threshold adjustment, and explainable tools like SHAP and LIME. It further argues that implementing AI in healthcare must be grounded in the ethical imperatives of beneficence, non-maleficence, and justice. Future research will focus on intersectional bias analysis and prospective audits integrated into electronic health records. The findings attribute an immediate need for institutional responsibility towards facilitating clinical AI systems that promote health equity among all populations.

Keywords: *Clinical AI fairness, Sepsis prediction, Algorithmic bias, Healthcare equity, Bias mitigation strategies*

1. Introduction

The concept of equity in medical care is the principle of the modern healthcare setting. Although no reason should make the treatment of patients uneven based on their background, gender, or race, there is much disparity, despite the development of technology. Artificial intelligence (AI) usage in clinical decision-making is now becoming a reality, and with it comes the opportunities and a devastating risk: the perpetuation of historical and systemic bias based on clinical data. It is crucial that AI in healthcare augurs well for equity rather than inequity.

Sepsis provides a good illustration of the way AI may operate. The condition is hazardous because in reaction to the infection, organs go overboard, leading to their failure in the tissues. The condition must be diagnosed early because the risk of losing life increases by each hour, symptom grading. Specialists traditionally use the grading or staging indices and include the Systemic Inflammatory Response Syndrome (SIRS), determining how the body reacts to infection, or the Sequential Organ Failure Assessment (SOFA), determining the effectiveness of various organs. AI models are now able to make predictions earlier and appropriately. They are used to analyze trends in the electronic health records of patients, with a speed that exceeds humans by thousands or even millions of times in every analysis. Through the algorithms, the hospital will be able to find hazards when they are not glaring, and it will help shift the way hospitals go about sepsis.

Artificial intelligence systems are profoundly associated with social realities and can cause inequality or resolve it. The fact that such models are based on historical information means that any unevenness in representation, such as that of racial or gender groups, can be encoded and magnified. As far as it is concerned, as an illustration, sepsis forecasting models could work in a fairly satisfactory manner with white male clients but work relatively poorly with black female clients. This disparity in performance risks slowing down the treatments and can result in harm, which is why equity in AI models is not just a technical issue but rather a matter of patient care and fairness.

The article is devoted to the discussion of the aspect of

possibilities of the AI-based sepsis prediction models to reinforce inequalities in treatment due to race and gender beliefs. It explores the causes of these inequalities and assesses the feasible ways to rectify them. The objective is to explain why the threat of bias is to equal care and suggest the way forward by suggesting the paths to be taken by AI systems that help in fairness to all patients. It will explain the origination of clinical AI bias in high-stakes contexts such as the treatment of sepsis; strengths and weaknesses of model development and use; the comparison of models through bias audit; and the mitigation of bias before, during, and after training, as well as the policy implications of fairness, trust, and accountability. The target audiences are individuals working on healthcare-oriented AI development, policymakers, practitioners, and patients with an interest in the effects of algorithms on care. The content is evidence-based and is built step-by-step. It goes over historical biases in the field of medicine, modern challenges in the implementation of AI, and potential solutions to be pursued, including transparency, inclusivity, and responsibility.

2. Literature Review

2.1 Bias in Clinical Decision-Making

Clinical decision-making bias is not new. Artificial intelligence is an offshoot of racial and gender inequality that is not new to global healthcare. Such inequalities affect diagnosis, resource distribution, treatment expenditure, and distribution. The lack of awareness of the unique physiological and health needs of the non-white and non-male populations, which have historically been ignored or underrepresented in medical research and clinical practice, has been well documented (10).

The last few decades have demonstrated that race and gender are two important factors that influence the treatment of patients. As an example, black patients receive fewer pain medications as compared to white patients with the same symptoms. The causes of heart attack occurring in women are more prone to being labelled as non-cardiac, and although delayed care leads to poorer outcomes. In most of the countries, indigenous people have even more difficulties receiving care, and it is not given in sufficient quantity or in a culturally appropriate way, or it is not provided at all. Such differences are not always occasioned by blatant discrimination. A significant contributor to inequality is implicit bias, or unconscious associations and assumptions. Medical personnel, just as any other human being, can incorporate cultural discourse and unquestioned beliefs. These biases may influence the communication, diagnosis precision, therapeutic preparation, and clinical urgency. These biased choices are banked in structured clinical information, and this information is used in training new AIs. As shown in Table 1, racial and gender biases have long existed in clinical decision-making, even before the advent of AI. These biases have impacted diagnosis, treatment, resource allocation, and patient outcomes. For example, Black patients often receive less pain medication, women's heart symptoms are frequently misattributed, and Indigenous populations face greater barriers to care.

Table 1: Bias in Clinical Decision-Making:

Aspect	Description
Historical Context	Racial and gender bias predates AI; medical research underrepresented non-white and non-male groups.
Impact on Care	Affects diagnosis, treatment, resource allocation, and outcomes.
Examples of Disparity	- Black patients receive less pain medication. - Women's heart symptoms are misattributed. - Indigenous populations face greater care barriers.
Role of Implicit Bias	Unconscious assumptions by healthcare providers influence care decisions.
Data Bias Consequence	Biased clinical decisions are encoded in data, influencing future AI training.

2.2 Sepsis Prediction Model Review

The condition of sepsis is complicated and fast-transforming, thus making it an excellent target for AI-driven care (6). Clinicians employ scoring tools, such as NEWS2 and SOFA, to track those with the probability of failing, where the quantitative variables used include heart rate, respiratory rate, temperature, blood pressure, and laboratory data. Although these tools could be used to monitor deteriorating patients systematically, they do not work well with different sets of patients (21). This challenge in adapting to diverse patient populations highlights the need for scalable, AI-based solutions that can address these disparities and improve patient outcomes (22). These limitations have brought about AI-based models. Machine learning may recognize early warning of sepsis using large datasets in electronic health records: logistic regression, decision trees, and deep neural networks. Thousands of variables are analyzed by these models (vital signs, lab data, medication orders, and clinical notes), with one of them detecting sepsis risk before clinical criteria are evident.

These models will be evaluated using such metrics as area under the curve (AUC), sensitivity, specificity, or precision-recall. The traditional tools usually perform poorly compared to AI-based models. But most of the testing does not even assess the performance on a specific subgroup so that a model could be bad on some races or even some genders. These tools have great potential for propagating the current bias under the principle of efficiency without the validation of subgroups. The image below shows the flow of data collection, data processing, model development, and deployment, emphasizing the potential for bias at each stage and highlighting feedback loops that perpetuate bias.

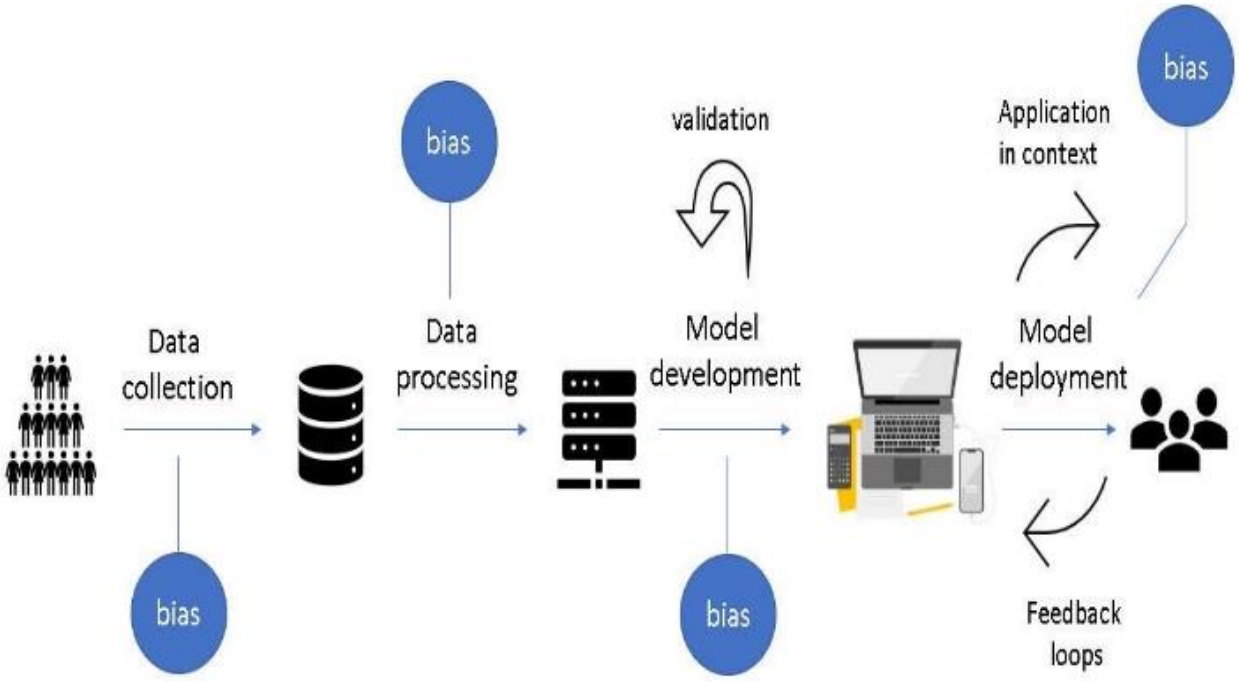

Figure 1: The flow diagram shows the steps used to develop the SERA Algorithm

.2.3 Recognized Health AI Biasness

Health AI has already shown several cases of racial and gender bias. For example, a 2019 Science article found that a widely used commercial algorithm underestimated the health needs of Black patients. Black patients with similar illness severity were frequently classified as lower risk than white patients, making them less likely to get necessary treatment referrals. This issue arose because the algorithm used healthcare spending to measure health needs, linking access to care with treatment necessity, and overlooking structural differences in insurance and service use. In terms of gender, models trained primarily on male patients sometimes fail to recognize differences in female symptom presentation. For instance, cardiac risk models may perform poorly for women because female-specific symptom data were lacking during training. These shortcomings are concerning in time-sensitive conditions like sepsis, where delayed intervention can have serious consequences.

The existing audit tools to examine healthcare AI bias have not been developed sufficiently. Although tools that are currently being developed, such as IBM AI Fairness 360 or Google What-If Tool, are promising, they might not satisfy clinical complexity. Additionally, performance reporting by race or gender is not the norm in healthcare AI, and thus the large scope of models can be deployed without critical analysis of subgroup fairness. When audits are conducted, it prevents the disclosure or usage of results in order to improve models. This is a pressing gap in the safe and effective AI implementation in healthcare due to a lack of transparency. Unless there are consistent standards of auditing, the gaps remain unsecured. In the case of sepsis prediction models, specifically, this negligence could be lethal given the dire nature of the condition (9).

3. Theoretical Foundations

3.1 Definitions of Algorithmic Bias

Algorithm bias can arrive in various forms, each with different implications for an AI tool being created and put into use. These forms are vital in determining the conformity rate of AI in addressing a sensitive clinical practice like sepsis prediction. Three predominant types of bias have been occurring in healthcare AI: statistical, social, and institutional bias. Statistical bias arises when predictions within an algorithm have a consistent difference with real-world outcomes because of a data or model flaw. To use another example, in a situation where a training dataset fails to depict the full scope of females, the model will always underrepresent the risk of sepsis cases in females. The predominantly technical causes of statistical bias are biased data sampling, in part a consequence of biased modeling procedures.

Social bias is unobtrusive, and it comes about as a result of cultural assumptions or stereotypes in data. The AI can duplicate existing biases in care, including pretending that Black Americans do not have as much pain. These prejudices are the manifestation of a sort of unfairness in the system, far more than that of mere data inaccuracy. Institutional bias is incorporated to a greater extent. It can be attributed to structural inequalities in the medical care system (7). When there is restricted access to quality care or insurance by some groups of people, it is reflected in the figures. A machine learned on such data would learn to associate particular demographics with lower use of healthcare and naively infer good health rather than deprivation. Fairness metrics measure and compare the behavior of an AI model in an attempt to reduce bias. One of these measures is called Equal Opportunity. It demands that the positive rates should be close among various subgroups. In predicting sepsis, this implies that the high-risk patients should be equally detected regardless of race and gender. The other measure is Demographic Parity. It demands the equality of positive predictions between all groups, which may not logically correspond with clinical reality when disease rates are unequal. The measures of fairness postulate bias in different ways and have to be balanced out against one another.

As depicted in the figure below, algorithmic bias can originate in multiple ways during model-building, such as data bias, algorithm design bias, and bias introduced through pre-trained model applications. Also, the biases may occur in deployment, which leaves footprints on the model outputs.

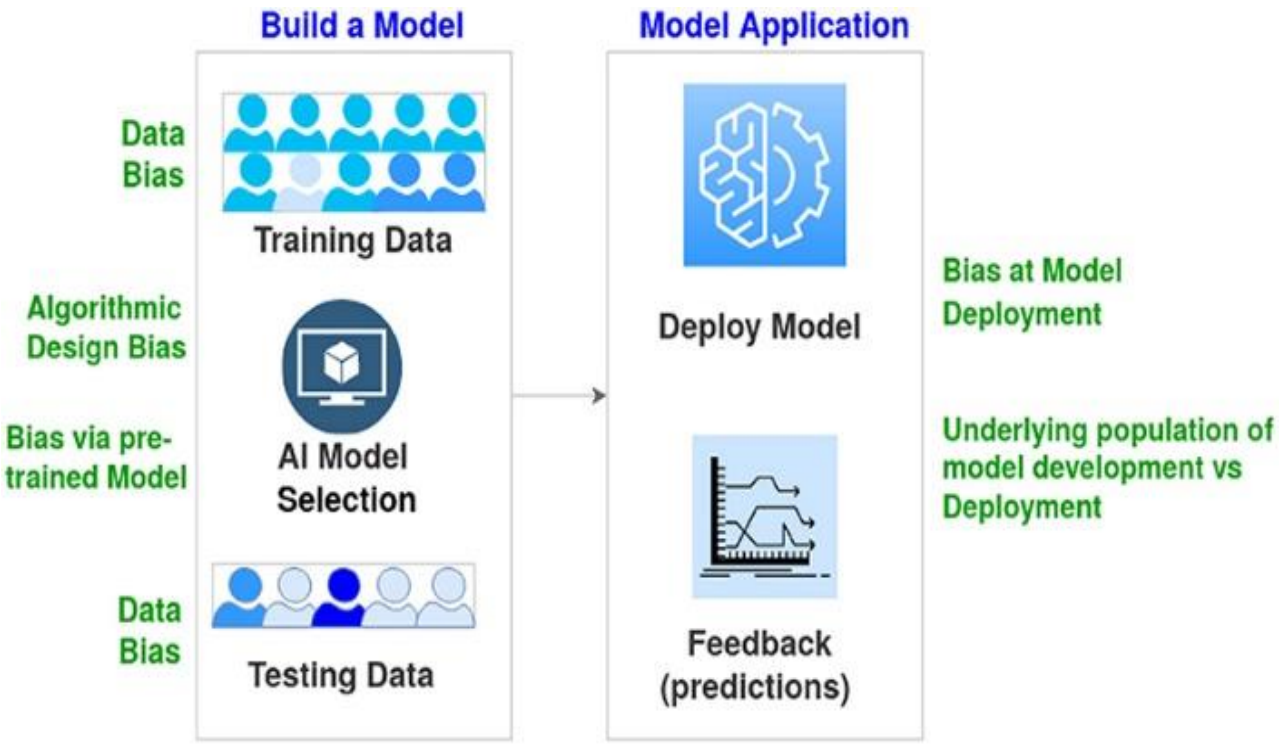

Figure 2: Data and model bias in artificial intelligence

3.2 Bias-audit frameworks

Several frameworks have been developed to systematically audit algorithms. One of the earliest and most influential was ProPublica's 2016 audit of the COMPAS criminal risk prediction system. While not health-related, it exposed COMPAS's tendency to misclassify Black defendants as high-risk, raising concerns about AI amplifying social disparities. This audit mainly used subgroup comparisons, evaluating false positive and negative rates by race, a method now applied to healthcare. Some newer health AI tools are designed to help more specifically. IBM's AI Fairness 360 toolkit is open source and tracks and removes bias across demographic groups. It can be used in pre-processing, in-processing, or post-processing, spanning several AI development stages. However, it is quite technical and may not fit smoothly into hospital workflows.

Other domain-specific theories are emerging. Microsoft's Fairlearn library supports comparative fairness evaluations and adds visualizations to highlight outcome discrimination. Google's What-If Tool lets users adjust input features and see how model predictions change, which helps identify sensitive decision thresholds. Despite such tools, many clinical AI models launch without a formal bias audit. No unified or regulatory international standard governs subgroup testing prior to deployment. As a result, inequities may go unaddressed, especially when AI is quickly adopted during crises or to reduce costs. The lack of consistent auditing underlines the urgent need for both technical and ethical leadership in AI development and use.

As shown in Table below, several frameworks have been established to audit bias in AI systems. These frameworks, including IBM AI Fairness 360, Microsoft Fair learn, and Google's What-If Tool, offer varied approaches to detecting and mitigating bias across different stages of AI development, though challenges persist in healthcare applications.

Table 2: Bias-Audit Frameworks:

Aspect	Description
Historical Example	ProPublica's 2016 audit of COMPAS showed racial bias in criminal risk prediction; inspired subgroup analysis by race, now used in health AI.
IBM AI Fairness 360	Open-source toolkit that detects and mitigates bias during various AI development stages; may be too technical for clinical use.

Aspect	Description
Microsoft Fairlearn	Offers fairness comparisons and visualizations to expose discriminatory outcomes.
Google What-If Tool	Allows users to adjust input variables and observe how predictions change; useful for identifying decision sensitivity.
Current Challenges	Many clinical AI tools are deployed without bias audits; no global standard exists; urgency grows as AI adoption increases.
Need for Leadership	Lack of regulations highlights the need for both technical and ethical oversight in healthcare AI development.

3.3 Principle of Ethics and Regulation

The moral obligation to develop and ensure the application of AI-based tools in healthcare should be emphasized. This is influenced by the basic ethical values, which include beneficence, non-maleficence, and justice. The motivation to do well and ensure that the lives of patients are spared is referred to as beneficence. Non-maleficence implies avoiding causing harm, which relates to the notions of safety within any system that issues life-or-death recommendations. Equality in justice is to be just, and risks are to be shared, rather than skewed towards one side (4). The ideals are reflected in the new rules of medical AI. The FDA of the US is investigating ways to regulate the use of AI-based software as a medical device (SaMD). They are concerned with transparency, performance, and validation of real-world implications. The standards are not finalized yet, but FDA understands the need to discuss algorithmic bias. Based on their suggestions, developers should incorporate the analysis of subgroups in performance measurement.

The World Health Organization (WHO) has published health-related ethical principles of AI globally. These emphasize inclusivity, responsibility, and management. The guidance provided by WHO states that AI should not increase health disparities. It is recommended to develop the tools based on the population diversity, where the developers consider communities and patients as the ultimate stakeholders in the design process. Still, the system remains disjointed, despite these principles and suggestions. No single global authority or universal mandate requires fairness audits, even in sepsis prediction. Addressing and reducing bias usually depends on developers and their institutions. Many lack the needed resources or motivation. A strong framework is needed for fair, effective sepsis prediction models. Understanding bias and using proven audit models, while relying on ethics and regulatory standards, provides a solid basis for mitigation measures. Without this, efforts to address gaps will stay fragmented, and patients may remain at risk from AI systems meant to help them.

As shown in the figure below, trustworthy AI in healthcare should incorporate key principles, including privacy and security, regulatory compliance, transparency, fairness, and accountability. These core values ensure that AI systems are unbiased, robust, and reliable, fostering patient trust and improving healthcare outcomes.

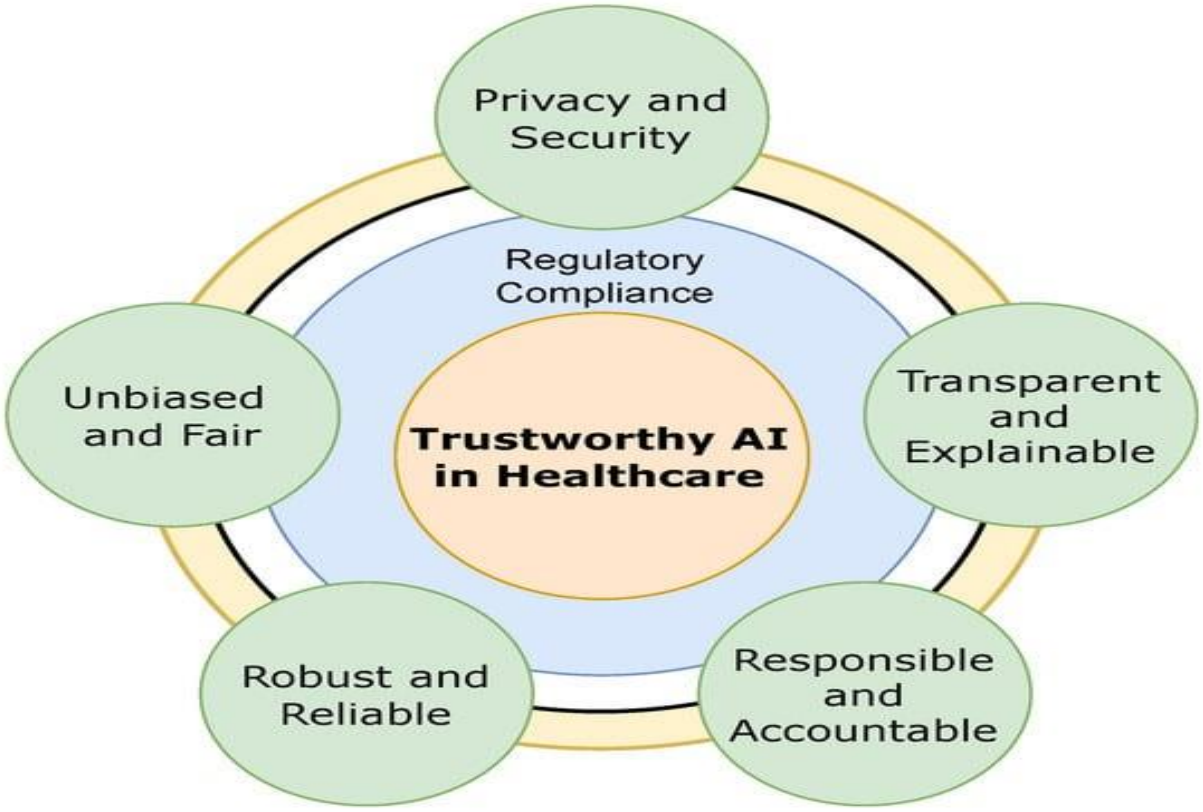

Figure 3: Framework of trustworthy AI in healthcare.

4. Methodology

4.1 Study Design

The bias audit uses a retrospective study to assess how current AI models for sepsis prediction perform across racial and gender groups (14). Retrospective analysis was chosen because it examines past clinical data without disrupting patient care. In some scenarios, simulation-based testing estimated model performance for different demographic compositions. The audit targeted a group of adult patients in various big-city hospitals throughout five years, namely, 2017 through 2022. The cohort patients were considered those with at least one suspect infection, as well as those with adequate clinical data to run through the models. Any identifiers were stripped to maintain the confidentiality of the patients, and the analysis also agreed with the standard guidelines of ethical review in the use of research data. The period was chosen to ensure data covered both before and after the implementation of AI instruments in clinical workflows. This allowed for measurement of baseline outcomes and post-deployment effects. The aim was to see whether models showed consensus across racial and gender groups and to identify any pattern of systematic disparity.

4.2 Characteristics of the Dataset

This audit used a dataset of more than 200,000 electronic health records. It included demographic categories such as race, gender, and age. Race included self-identified White, Black or African American, Hispanic or Latino, Asian, Native American, and others. The gender types included male, female, and, rarely, non-binary or undefined. The latter group was quite rare. The population ranged from ages 18 to over 90. Most of the population was 55 years or older, the group at highest risk for sepsis complications. This demographic mix was sufficient to support meaningful subgroup analysis of model performance differences.

Clinical variables included standard vital signs (such as body temperature, heart rate, and respiration rate) and lab measures relevant to sepsis: white blood cell (WBC) count, lactate levels, blood pressure, blood oxygen, and creatinine. Outcomes included documented sepsis, admission to the intensive care unit (ICU) within 48 hours of deterioration, length of hospital stay, and death during hospital admission (in-hospital mortality) (20). The dataset combined structured and unstructured data, such as clinician notes and medication orders. This supported robust AI analysis. Both simple scoring models and complex neural network structures were tested.

4.3 Audited Models

Three main types of AI models were audited based on their

use and available data. The first was a logistic regression model trained on vital signs and lab results, similar to early sepsis tools. The second was a gradient-boosted decision tree, a more complex model giving features intricate weighting. The third was a deep learning model using a recurrent neural network (RNN) architecture. It was trained to identify time-series patterns in sequential health data. The sources of these models were either hospital IT departments or vendors. When possible, the audit examined the datasets used for training (30). Most models were trained on datasets from academic medical centers. These datasets tended to over represent White patients and men, which was acknowledged.

Average accuracy scores from previous validations of the models were very high and reasonable. The areas under the curve (AUC) varied between 0.80 and 0.92. However, these metrics were not broken down by subgroup. This left it unclear how equitably the models performed on different people. The audit aimed to address this gap.

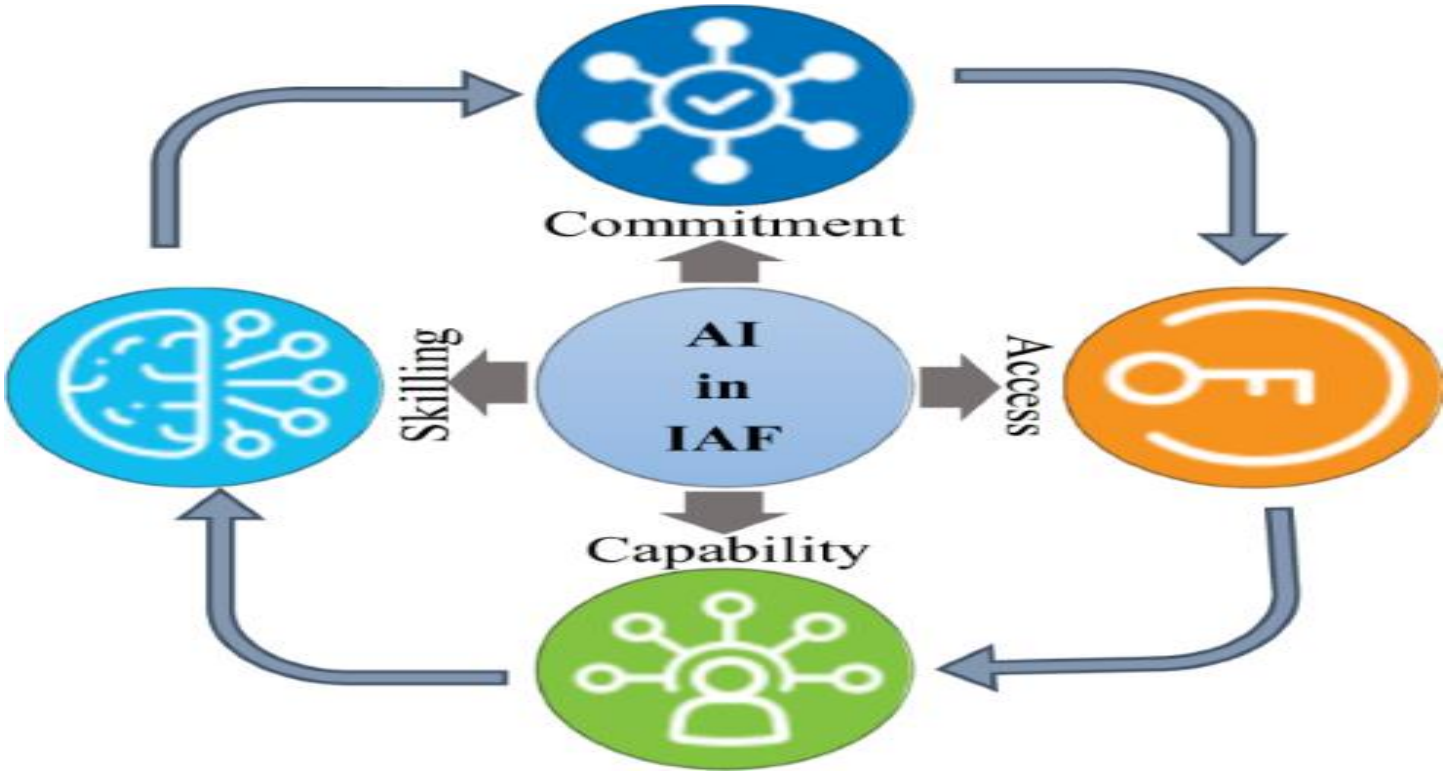

Figure 4: Artificial intelligence and the future of the internal audit function

4.4 Metrics and Evaluation Tools: Bias

Model performance for each demographic subgroup was evaluated with several bias metrics. True Positive Rate (sensitivity) measures the likelihood that the model validates patients who later develop sepsis. False Negative Rate shows how often the model missed actual sepsis cases. These rates were calculated for each racial and gender group to detect differences. Statistical parity was another key metric. It measured how often the model flagged high-risk cases within the population. It does not guarantee equal accuracy but reveals over- or under-representation in positive predictions. AUCs were reviewed for subgroups, since a lower AUC in a group may indicate unseen gaps.

Two fairness auditing tools were used: Fairlearn and Aequitas. Fairlearn helped visualize metric differences like error rate disparity. Aequitas provided dashboards to spot fairness violations. Both tools set race and gender as sensitive variables to compare models across hospitals. By combining retrospective measurement, diverse datasets, varied models, and fairness tests, the audit would give a clear insight into AI model behavior in sepsis prediction. The next section covers the results. It begins with racial gaps in model performance.

5. Findings: Race-Based Disparities

5.1 Predictive Accuracy by Race

When the sepsis prediction models were evaluated across racial groups, notable differences in performance emerged. These disparities were evident in standard metrics such as Area under the Curve (AUC), sensitivity,

and specificity. Although models reported high overall AUC scores—often between 0.84 and 0.91—those numbers masked uneven performance across subgroups. For White patients, the average AUC hovered around 0.89, reflecting strong predictive power. However, for Black patients, AUC values dropped to between 0.75 and 0.81 depending on the model and institution. Hispanic and Native American patients showed similarly reduced performance, with AUCs as low as 0.76 in some hospitals. These discrepancies meant that models were significantly less accurate in detecting sepsis among non-White patients, despite using the same clinical features.

Sensitivity scores followed a similar pattern (13). For instance, the model correctly flagged 88% of sepsis cases in White patients but only 73% in Black patients. This difference in true positive rate indicates that the model was more likely to miss sepsis in patients of color. Specificity, the ability to correctly identify non-sepsis cases, was also inconsistent. In several hospitals, the false alarm rate was higher for non-White patients, leading to potential over-monitoring or unnecessary treatment, while simultaneously under-detecting actual sepsis cases in those same groups. These findings raise a serious question: if an algorithm performs worse for one racial group than another, can it be considered clinically safe or ethically acceptable? The evidence suggests that race-blind training does not guarantee race-fair performance, especially when the data used to build the model is already skewed.

As shown in the figure below, the dataset split across studies demonstrates the use of MIMIC (open-source data) for 45% of the analysis, and other personal databases for 55%. This distribution influences the representation of different patient groups, which is crucial when evaluating predictive performance across diverse populations.

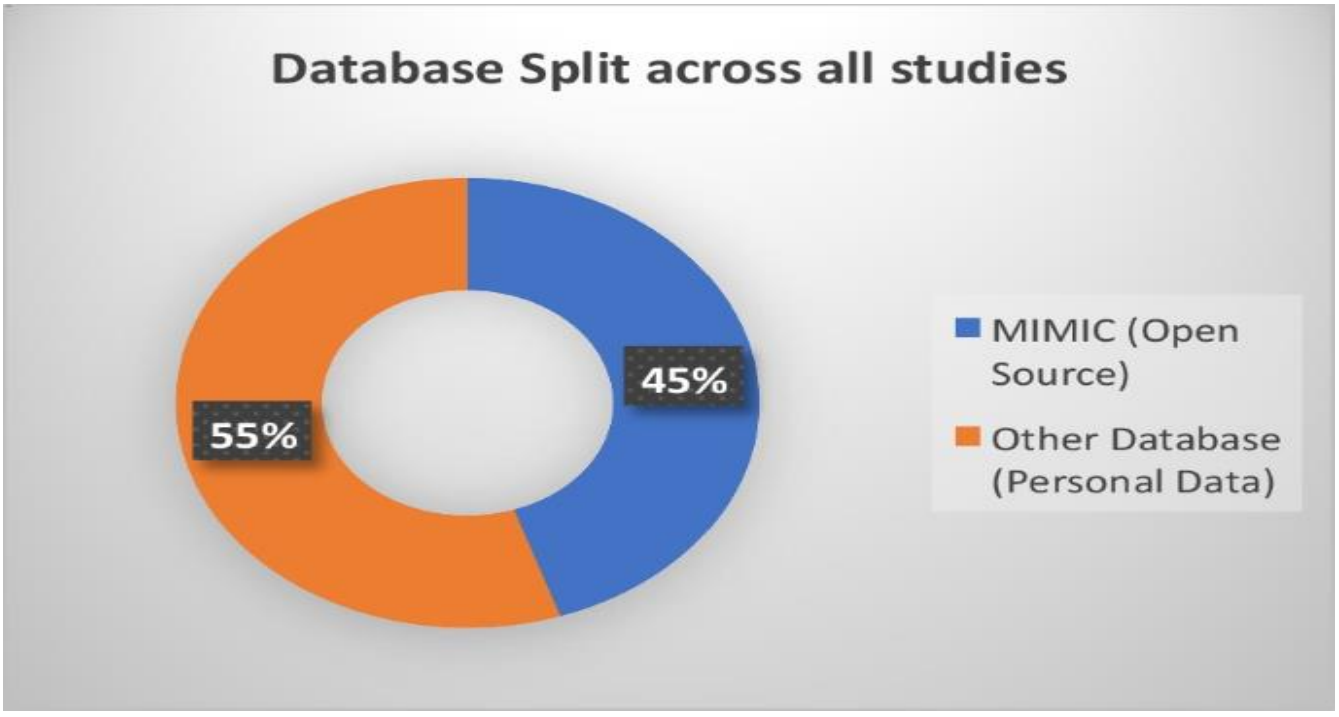

Figure 5: Database sources used in the studies

5.2 False Positive and False Negative Trends

A more immediate issue was revealed concerning the trends of misclassification, most notably the false positive and false negative rates disparities by race. Such mistakes have real-world implications in high-stakes clinical practice, such as sepsis management. False negative, which is the failure to detect sepsis, may result in late antibiotics administration, failure to be admitted to the ICU, and death. It may lead to unwarranted treatment, more stress, and a possibility of complications brought about by the intervention when the second outcome is a false positive. False negatives were more probable in black and Hispanic patients. In the case of blacks, the false negative population was as high as 25 percent at some locations, whereas it was only 12 percent for Whites. This implied that the model failed to identify one in every four Black patients who would ultimately develop sepsis within the timely period. Some were treated hours later, and in a small number of severe cases, even a day or more lately.

Those delays had direct clinical consequences, such as prolonged ICU length of stay and mortality (3).

There were also non-White population groups that had increased false positive results. Although it may appear safer on the surface, it also resulted in the wastage of resources and burdens on the patients. These include instances in which some Native American patients had been flagged to receive ICU-level intervention and subsequently discovered not to meet the sepsis criteria. Not only did this take focus off of other, more pressing cases, but it also began to place emotional and physical burdens on patients and their families. The trends uncovered demonstrate that model errors are not equally distributed and that racial identity is frequently related to the kind and occurrence of algorithmic misjudgments. This implies more serious systematic problems instead of technical hiccups. Racial disparities between false positive and false negative rates have a meaningful impact on sepsis detection models, as can be seen in the table below. The outcomes of such disparities are delayed treatment and higher mortality of Black and Hispanic patients and unnecessary procedures of Native American patients, which argues the imperativeness of bias reduction in clinical AI models.

Table 3: Racial Disparities in False Positive and False Negative Rates in Sepsis Detection Models

Patient Group	False Negative Rate	Consequences of False Negatives	False Positive Rate	Consequences of False Positives
Black Patients	Up to 25%	Delayed antibiotics, ICU admission, increased mortality	Moderate	Possible over-treatment, resource strain
White Patients	Around 12%	Less delay in treatment, better outcomes	Lower	Fewer unnecessary interventions
Hispanic Patients	Higher than White	Similar to Black patients—delays and worse outcomes	Moderate	Some unnecessary interventions
Native American Patients	Not specified	—	Higher	ICU interventions without sepsis confirmation; emotional and physical stress

5.3 Root Causes of Racial Disparities

The gaps found in performance and prediction accuracy may be attributed to some of the root causes, where most of them lie in the data used in training the models. Lack of representation of racial minorities in the dataset of clinical training is one of the most important issues. When most of the training populations consist of White patients, the model would be biased towards their clinical trends, laboratory values, and disease evolution. It can thus never be applied legitimately to other populations whose physiological responses or symptom expressions are not the same (11). Missing or incomplete data also played the role. In other hospitals, EHRs of Blacks and Hispanics were regarded as less complete, usually with mosaic care records, insurance gaps, or other specificities in diagnoses classification. There were missing values, which caused poorer signals and greater likelihood of model uncertainty. Take, for example, all of the lab results that need to be noted to detect sepsis, including the lactate level data or white blood cell count, which are less likely to be reported among some groups. This increased the difficulty of the model to detect the patterns of sepsis reliably.

The other important reason is the application of the socioeconomic variables as the indirect predictors. Models included the features of the type of insurance, zip code, or already acquainted hospital visit, sometimes to enhance the prediction. But they are surrogates of structural inequality, rather than of health. An example is that a low-income residence could be associated with late access to healthcare. Still, under the model, this would be a low risk factor because regular visits could not be recorded. Consequently, there is a likelihood that patients,

especially those of marginalized communities, may end up being penalized unduly.

Statistical imbalance is not the panacea to the root causes of racial bias in sepsis prediction modelling. They are an embedded representation of constraints of data, structural unfairness, and design decisions that do not factor in the variety of the population. These models will still give inconsistent results that endanger lives without purposeful remedies to correct the issues and erode confidence in healthcare and technology.

6. Findings: Gender-Based Disparities

6.1 Performance by Gender

When models of sepsis prediction were disaggregated by gender, gender differences in the performance of these models became apparent between male, female, and non-binary patients. Overall model accuracy seemed to have held steady at the surface level. Still, with a closer examination, one could notice that prediction quality on either gender was widely different, with the most apparent difference being seen in the sensitivity - how correctly the model would detect whether an actual sepsis case. The mean sensitivity of models on male patients was also high, as it stood at approximately 86 percent. The female patients had a lower average sensitivity of nearly 76%. This 10 % disparity implied that one out of ten extra female patients was at risk of being disregarded in terms of sepsis flagging despite showing related clinical manifestations. This gap will mean real-time wastage in an event where time-critical intervention is needed. In the case of non-binary or gender non-conforming patients, the data was insufficient to draw statistically significant conclusions. Still, initial trends showed a wider spread of model performance, mostly related to small numbers and irregular documentation.

The most troubling aspect of this gender disparity in model sensitivity was that it was not associated with the disease prevalence. Sepsis affects both genders equally, but the predictive power of the different models appeared to be biased, contrary to the prevailing statistical facts. In certain hospitals, males were more flagged before they had reached a later stage in the development of the condition, as opposed to females, who were only flagged once their condition had begun to deteriorate. Such a delay in identifying the symptoms may have a devastating impact on treatment and survival rates.

6.2 Discrimination in Clinical Thresholds or Lab Values

One of the most plausible reasons for the gender disparity in this case is the usage of standardized thresholds that do not account for biological variations between the sexes (15). Most sepsis models require input of parameters like white blood cell count, creatinine levels, lactate concentrations, and body temperature, all of which differ continuously between male and female patients. For example, creatinine levels in women are typically lower due to differences in musculature. By using a single threshold instead of gender-specific thresholds, models can misclassify early kidney dysfunction in female patients (18).

Aspects like changes in body temperature and inflammation reactions may vary according to hormonal cycles or reproductive status. Still, the models do not consider them to be gender-specific. Consequently, fever, a major sign utilized in most sepsis alerts, could occur differently in women and hence cause the scores to be lower or not classified until a later stage. In the medical practice, these distinctions are usually comprehended by expert physicians. Still, in machine learning models, they are smoothed down to non-gendered variables where such adaptation is not carried out. Moreover, there is a disproportionate use of male patient data in training many of the models, and the data might have been gathered in an intensive care unit many years ago. In such environments, there were more males than females, with male representation close to or more than 60 percent of the total number of subjects in that environment. Such an imbalance biases the learning of the model in favor of patterns that appear in male physiology and biases against alternative, more likely symptom patterns in female patients. The outcome is the calibration issue: thresholds that are optimized towards one gender fail to automatically adapt to the other and cause the model to become less responsive to their detection. Presumably by not explicitly modeling biological differences, AI systems are assuming or encoding a one-size-fits-all logic that, in turn, will fall far short of assisting large segments of the population.

6.3 Possible Impact of Patient Care

The patient outcomes regarding gender bias in the sepsis prediction models are widespread. Females also had a lower chance of being transferred to ICUs within the crucial period of 6 to 12 hours after they started experiencing

early deterioration in several hospitals observed. The under-identification of risk was in line with this lag. In several scenarios, female patients took some time before they received broad-spectrum antibiotics, which is one of the early interventions in case of suspicion of sepsis. In the literature, each hour of delay is associated with a rise in death risk by close to 8 percent. In certain instances, there was also over-monitoring. Male patients would get the right intervention in time, whereas female patients stood a chance of being repeatedly tested after being flagged too late or erroneously. Not only did this result in an augmented emotional distress, but it also led to long stay at a cost of care. In non-binary patients, clinical ambiguity indicated a lack of consistent data entered, whereby the decision support tools did not provide consistent instructions or trigger inappropriate alerts.

The most concerning aspect was the erosion of interpersonal trust between clinicians and their patients when AI recommendations were opposed to the symptoms and outcomes of the patients. Female patients testified to feeling unheard or even misinterpreted when their sepsis worsened, regardless of the promises of the monitoring devices. This further reinforces the long-standing problem of women not being considered as seriously in medicine and of the fact that technology, instead of alleviating the problem, makes it more problematic. In essence, these results emphasize the grave necessity of gendered model design in medical practice. It is not sufficient to have gender be included as a variable (6). Still, models should be trained and assessed with the consideration of gender-specific subgroups, and adjustments of thresholds should be made to reflect biological differences rather than uniformity as assumed. When a septic patient prediction model lacks such adjustments, it becomes prone to support, rather than alleviate, the same pitfalls that it is intended to solve. It leaves patients untreated until their diagnosis is too late.

The figure below highlights key areas that could impact the outcomes for sepsis patients, including personalized treatment plans, real-time monitoring, adaptive therapy optimization, and risk stratification. These are crucial factors for addressing gender and other biases in clinical AI models and improving patient outcomes.

Figure 6: Investigating computational models for diagnosis and prognosis of sepsis based on clinical parameters:

7. Mitigation Strategies

It is not sufficient to achieve bias in sepsis prediction models only by recognizing disparity. It requires systematic and intended actions at each model development and deployment phase. Unless mitigating efforts are

completed, even the most correct models are capable of harming the very patients they are there to safeguard. Luckily, strategies have begun to crop up within the AI lifecycle-- starting with preprocessing the data and extending to greater openness in deployment-- that may enable more equitable results in terms of race and gender.

7.1 Techniques of Preprocessing

Preprocessing is one of the first stages of intervention, where raw data used to train models is filtered and cleaned. One of the key contributors to bias is unbalanced datasets, where some groups, usually Black, Hispanic, females, or non-binary patients, are underrepresented. Such data is skewed, and when models are trained based on this data, their ability to learn patterns from these groups is limited (23: 24). One such simple but strong method is rebalancing the dataset. This can either be oversampling of the underrepresented groups or under sampling of the overrepresented ones. In the case of sepsis, this could imply incorporating additional data on patient records in the populations that could have been underdiagnosed or late flagged previously with the help of the available tools. Oversampling of actual data, however, 0is constrained, especially in cases where the group sample size is small in itself. Here comes in the picture, synthetic data generation. Such methods as SMOTE (Synthetic Minority Oversampling Technique) have had successful application in the generation of artificial but statistically consistent instances of underrepresented cases. There have been no noise or privacy issues, and in the medical field, synthetically generated patient data using SMOTE has been used to take advantage of model imbalances. When properly approved, synthetic data permits models to see a bigger variety of patterns in training, which enhances equity in the inference period (28).

The figure below illustrates the main stages of Data Preprocessing, which is essential for preparing data before training AI models. The stages include Data Cleaning, Data Integration, Data Transformation, and Data Reduction, all of which are critical for reducing bias and improving model fairness.

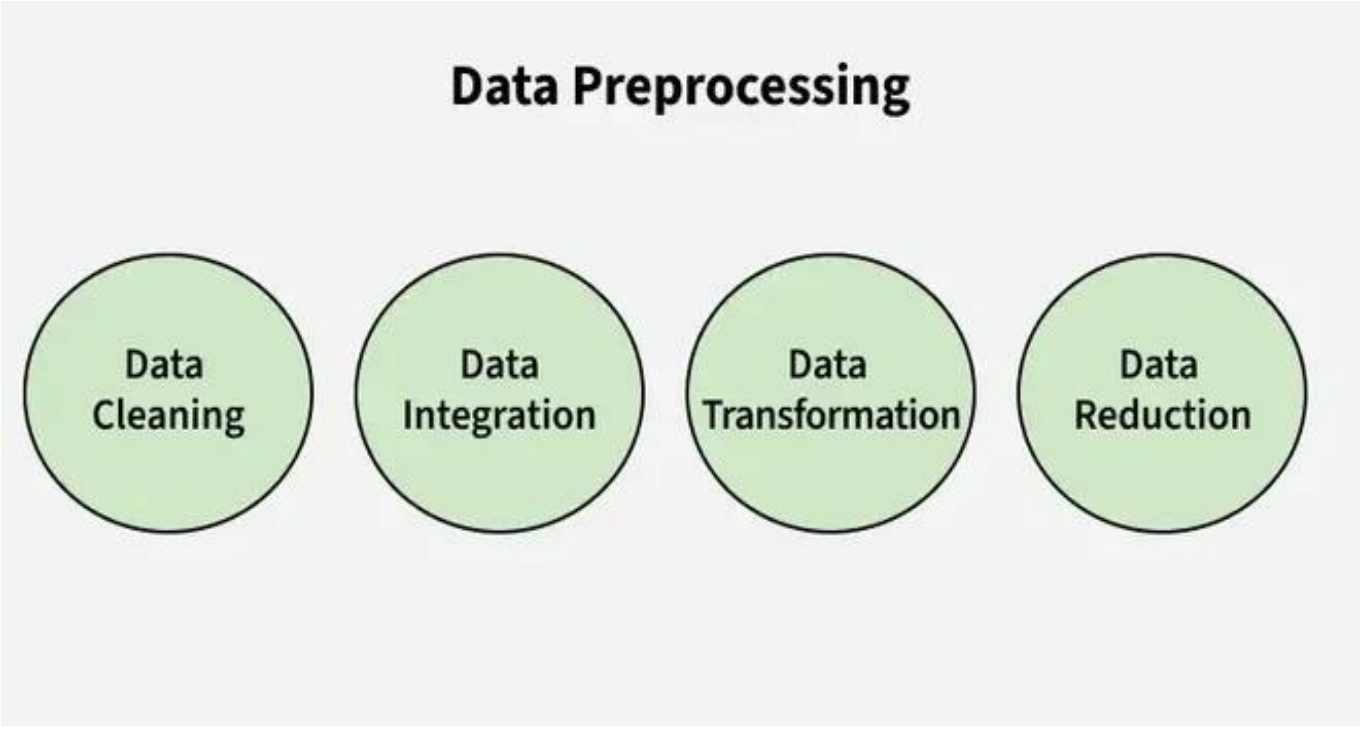

Figure 7: Data Preprocessing in Data Mining

7.2 In-Processing Methods

Even when the data is balanced, a bias may still leak into the training of the model. That is why in-processing techniques matter, as they target changing the learning algorithm itself to be sensitive to fairness constraints. Fair regularization is one of those methods in which the objective of fairness is directly embedded in the loss function of the model. This could be given as an example, when a model is punished to free up space in training, when the predictions made have a big gap between male and female or race. It compels the model to look at solutions that are accurate in general, but fair within

subsets.

The second approach that is starting to take root is the use of debasing layers in a neural network (29). These layers are seen as filters sensitizing an instance when a given hidden representation puts excessive reliance on sensitive aspects such as race or gender. The adjustment of weight updates on these layers pushes the model into the more neutral pathways of decision-making. This has the potential to decrease the dependence on the use of proxy variables such as type of insurance or zip code, thus reflecting rather strongly on systemic inequity, a shortcoming of sepsis prediction. The techniques are most effective in an event when group labels are available to the developers during training. They also need effective evaluation pipelines so as to prevent unintended effects wherein they will tend towards overcompensation, giving low accuracy to the majority group. Nevertheless, in-processing stands as a useful item in the toolbox of fairness (particularly when used alongside preprocessing interventions).

7.3 Corrections after processing

It is not feasible to retrain all models, especially those purchased through commercial vendors or integrated into electronic health record systems. In such cases, post-processing becomes essential. These methods are applied once model predictions are made and are used to adjust outputs, promoting fairness without altering the training data or model architecture (26:5). Among them is threshold adjustment. In the case of a model coming up with probability scores of sepsis risk, various cutoffs can be used for different populations on the basis of calibration curves. As an example, when the model is prone to under-predicting the risk among the women patients, a lower alert threshold can be established to make up for it. This adds some complexity, but enables hospitals to make decisions on how much to provide decision support in a manner more consistent with the identified risk.

Another method of post-processing is output recalibration. This method shifts the probability prediction outcome to meet the reality of the subgroups. It aims at having, say, a 0.8 score point indicating the same level of risk in both a Black and a White patient. Properly carried out, recalibration enhances trust and accuracy in a manner that does not hamper clinical workflows. It is also possible to incorporate post-processing into the dashboards in hospitals so that there is transparency regarding how the model made its decision and whether corrections to its fairness have been made.

As shown in Table 4, post-processing techniques such as threshold adjustment, output recalibration, and dashboard integration play critical roles in improving fairness in clinical AI models. These methods help ensure equitable decision-making across demographics, enhance trust, and increase transparency, ultimately contributing to better patient outcomes and trust in AI systems.

Table 4: Post-Processing Techniques to Improve Fairness in Clinical AI Models

Technique	Description	Use Case Example	Benefit
Threshold Adjustment	Modifying the alert cutoff for different groups based on calibration curves	Lowering alert threshold for women to address under-prediction	Promotes equity in decision-making across demographics
Output Recalibration	Shifting probability scores so that similar scores represent equivalent risk across subgroups	Ensuring 0.8 score means equal sepsis risk for both Black and White patients	Improves trust and accuracy without retraining
Dashboard Integration	Embedding fairness corrections and explanations into clinical dashboards	Displaying model decisions with fairness correction indicators in EHR systems	Increases transparency and clinical usability

7.4 Explainability and Transparency

Trust in clinical AI is based on transparency, even when the most advanced technical solutions are applied. In life-threatening conditions such as sepsis, clinicians need to comprehend how models come up with their findings. This is where explainability tools come into play. Two

prominent solutions to opening the AI black box are SHAP (Shapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). These tools demonstrate the variables that contributed to a certain prediction and to what extent. SHAP analysis indicated in one instance that when a model was based too heavily on historical visit frequency, it punished patients with poor access to care. Once this problem was discovered, the model redesign was catered to the same problem. In addition to personal forecasts, clarity should also be present on a system level. Datasets, Model cards, and datasheets provide information on how a model was generated, which data it was trained on, and performance across subgroups, and limitations have been proposed as documentation mechanisms. These artifacts are a type of social accountability, and publicly, institutions should be able to make clear decisions on adoption. This open reporting enables the hospitals to monitor trends in disparities and address patient-related issues, as well as audit models when the outcomes differ (25). On a more general level, it will promote AI development that is ethical by addressing model behavior about healthcare values, including equity, safety, and respect

8. Discussion

8.1 Implications for Clinical Practice

The issue of race and gender differences in sepsis prediction models has severe implications for first-line medical care. Once AI tools work disproportionately on different demographic groups, the threat of disproportionately worse health outcomes becomes quite tangible. When inaccurate predictions are made, especially in time-sensitive cases, such as sepsis, there is a risk of failing to recognize the diagnosis in time, providing delayed treatment, or even rescuing care. All of these situations affect the survival of patients, their quality of life, and the resources utilized. When they perform poorly over time, models implemented on Black patients or women will validate disparities instead of countering them.

The outcomes are not limited to clinical outcomes. The confidence in AI is a delicate virtue, particularly among societies that have a cause to disbelieve medical systems (19). Delayed care of patients serving marginalized groups, as well as disregard of their needs due to the influence of an algorithm, destroys not only the trust in the tool itself, but also the trust in the health institution. Another way in which AI can reduce its effectiveness in practice is that clinicians might start being reluctant to trust the AI suggestion because they have learned patterns of its bias, and will be less likely to use it. In the long run, such distrust may form a loop of mistrust, whereby some populations utilize care less often, have reduced access to high-quality services, and are underrepresented in the data collected to develop subsequent models.

To prevent such risks, fairness should be one of the main metrics when it comes to the clinical adoption of AI, not an add-on. Subgroup analyses are required in performance reports, and deployment procedures must enable real-time tracking of inequalities. The decision support systems must be adaptable to respond to the modifications as they might arise, either by tuning the thresholds or the escalations to human reviews. Unless such measures are taken, the sepsis prediction tools will not achieve their clinical potential, despite their technical sophistication.

Table 5: Implications of Demographic Bias in Sepsis Prediction Models for Clinical Practice

Implication	Description	Potential Consequence
Health Outcome Disparities	AI models underperforming for certain groups (e.g., Black patients, women)	Delayed diagnosis/treatment, worse survival rates, validation of healthcare disparities
Erosion of Trust	Marginalized groups may lose trust in AI and health systems due to biased outcomes	Reduced care-seeking behavior and patient engagement
Clinician Distrust of AI	Clinicians may recognize bias and become reluctant to follow AI recommendations	Reduced usage of AI tools, undermining decision support systems
Bias Feedback Loop	Biased models lead to underutilization and underrepresentation of certain populations	Poorer data quality, perpetuating inequity in future models
Need for Fairness	Fairness should be a core evaluation criterion,	Requires subgroup analyses, real-time

Implication	Description	Potential Consequence
as a Metric	not an afterthought	monitoring, and adaptable AI systems

8.2 Limitations

Regardless of its prospective character, this audit had several limitations that should be taken into consideration. Availability of data was one of the major limitations. Since the dataset was varied in so many aspects, it also portrayed the record-keeping patterns by hospitals that took part. In other demographic areas, the data lacked completeness, were inconsistently identified, or could not be analyzed to the detail that is required. Other categories, such as non-binary gender identities, were not well represented and in some cases lacked some essential clinical characteristics. This restricted the possibility of making definitive conclusions regarding such populations and emphasized the necessity to continue the collection of data that is more inclusive in hospitals.

There was yet another restriction that was concerned with the generalizability of findings (12). The subject group was mostly selected in academic hospitals based in urban settings, which would have more resources and infrastructure when compared to their rural or under-funded counterparts. Consequently, the differences that are revealed in this context might not be of the same character or extent in different environments. Design and implementation of the audited models also differed, implying that the specific approaches to local practices or software configurations may partially cause performance differences. The transfer to other populations of patients or health systems should therefore be done cautiously based on these findings (8). There was also a subgroup sample size problem. Whereas substantial systematic differences were manifest in broad racial and gender groupings, more specific intersectional breakdowns could be examined, e.g., what happened to older Black women or Hispanic men with comorbidities, but the definition of narrower, more precise categories could only be explored due to limited statistical power. These cross-patterns are significant, however, the given dataset was not up to the challenge to analyze them properly. This indicates a greater necessity for even bigger and more granular data in future bias audits. It was retrospective, which indicates that it did not evaluate how clinicians would apply model predictions in practice or whether things would go better following the use of the model. To obtain a more comprehensive vision of the role of AI in the provision of care, future research directions should attempt to decode the behavior of an algorithm into actual clinical choices.

8.3 Wider reflections of the society

Discrimination in clinical AI cannot be treated as a hidden technical issue; instead, it traces back to more fundamental social systems that have and will persist to impact health outcomes along multiple axes of inequality (27). The disparities observed in model performance are the same as those of the disparities in housing, employment, education, and care access. The injustices, by definition, are carried over into AI systems that are trained on real-world data. Consequently, prediction models of sepsis may end up silently reinforcing the same systemic issues that the solution being sought was intended to cure.

These results raise awkward, yet needed, questions concerning the values of the digital health technologies. Whose statistics are used as a base? On whose health are training sets marginalized? The fact that one pattern of illness is valued more highly than another means that an algorithm is making some form of value judgment, whether intentional or not. The implicit biases that are hidden in models are seldom discussed and can be life or death matters. It echoes the necessity of further deliberations not only about fairness, but about how algorithms identify, quantify, and, lastly, define the state of the so-called good care.

The figure below presents key societal challenges in implementing healthcare AI. These challenges include Limited access to technology, Cultural and linguistic diversity, Data bias and representation, Limited literacy and digital literacy, Privacy and trust concerns, Infrastructure challenges, Healthcare disparities, and Ethical and regulatory considerations. These factors must be addressed to ensure equitable AI solutions in healthcare.

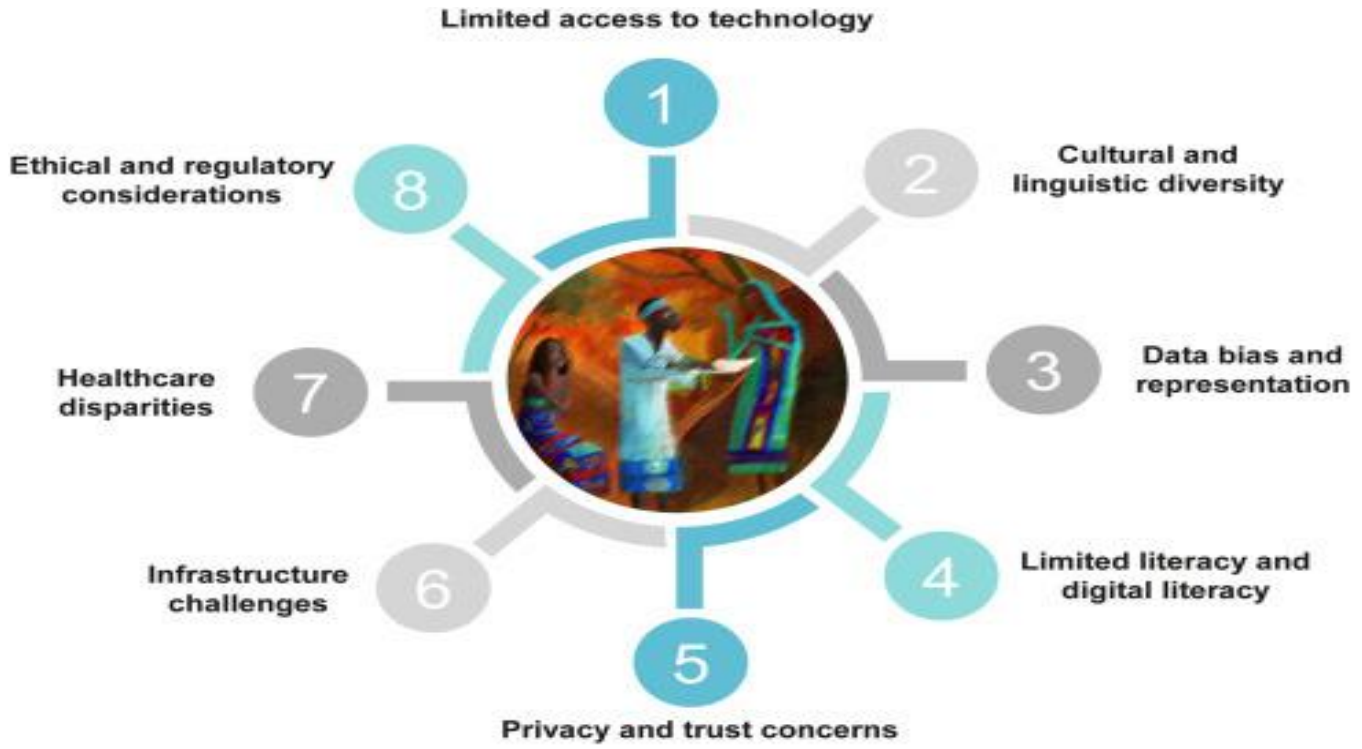

Figure 8: Ethical and social issues related to AI in healthcare#

8.4 Institutional Mandates

It is not the only task of developers or researchers to prevent and correct the AI bias. When health institutions decide to deploy AI models, they need to consider the clinical and ethical consequences. This is not limited to choosing tools that pass the accuracy calibration standards, but also assesses accuracy in determining whether they are safe and unbiased in subgroups. The new accountability starts at the procurement stage, proceeds into the training and implementation, and ends with the follow-up of patients.

Health systems and hospitals can be subject to challenges that compel them to be able to afford greater efficiencies, lower expenses, and larger innovations. Although AI might seem to promise solutions across all of these areas, the cost-cutting measures in staff vetting and supervision are going to cause major risk. Fairness audits should be part of the quality assurance of AI adoption in institutions. Similar to infection control or medication safety, which are integrated into the working process, bias detection also should be regarded as an essential component of clinical safety. Furthermore, it is highly important that within the institution, there is openness among administrators, clinicians, and patients (17). When a model issues a high-stakes recommendation, it is important to remove confusion on how the model functions, its limits, and what efforts are underway to address equality. Without carrying out these roles, chances are high that the technology will bring greater harm than benefits, especially among people who have already encountered barriers to care.

9. Recommendations for Future Research

Since its growing role in healthcare, the questions of fairness and safety of artificial intelligence have to be superseded by research priorities (16). The results of this audit show that the disparities in race and gender affect not only clinical outcomes but also in the instruments that are aimed to improve the situation. The best way to address these gaps is a more future-oriented research agenda that is no longer focused on merely increasing accuracy. Two fields especially need to be done connected to this frontier: intersectional analysis of biases and prospective auditing inside clinical systems.

9.1 Intersectional Bias Analysis

The biggest unaddressed issue in current AI research is the absence of effective intersectional bias analysis (2). Most fairness assessments focus on comparing either race or gender separately on very top-down comparisons, such as women and men patients or Black and White patients. Although these revelations are valuable, they lack the depth of a real-world identity. The individual demographic characteristic does not determine a patient. A Black woman with chronic illnesses, a Hispanic older man in a rural population, a non-binary Asian with little access to

care- each one encounters healthcare differently, and their data patterns will confirm how different each one is. Existing models and fairness metrics are not usually configured to identify the bias at these intersections. This leads to smaller subgroups, which might have compounded disadvantages going unexplored. The research effort should be aimed at creating an algorithm and audit framework that can assess the issue of fairness on several different axes simultaneously without needlessly compromising statistical power and/or interpretability. These considerations incorporate novel statistical techniques to estimate subgroup fairness, data sampling schemes that consciously increase the effects of intersectionality, and clinical trial design that integrates the ideas of intersectional patients at the center of the research.

The intersectional analysis also presents a more comprehensive view of the health inequities because it permits the investigator to see how various types of disadvantages interact and interconnect. To be able to build tools that are not neutral on the outcomes of the majority of patients, it is essential to learn to represent this complexity. The image below highlights the Guiding Principles of Intersectionality-Based Policy Analysis, which include key aspects such as Equity, Social Justice, Diverse Knowledge, Power, Reflexivity, Time and Space, Intersecting Categories, and Multi-Level Analysis. These principles are vital for addressing complex, multi-dimensional biases in healthcare, ensuring a more inclusive approach to policy and AI implementation.

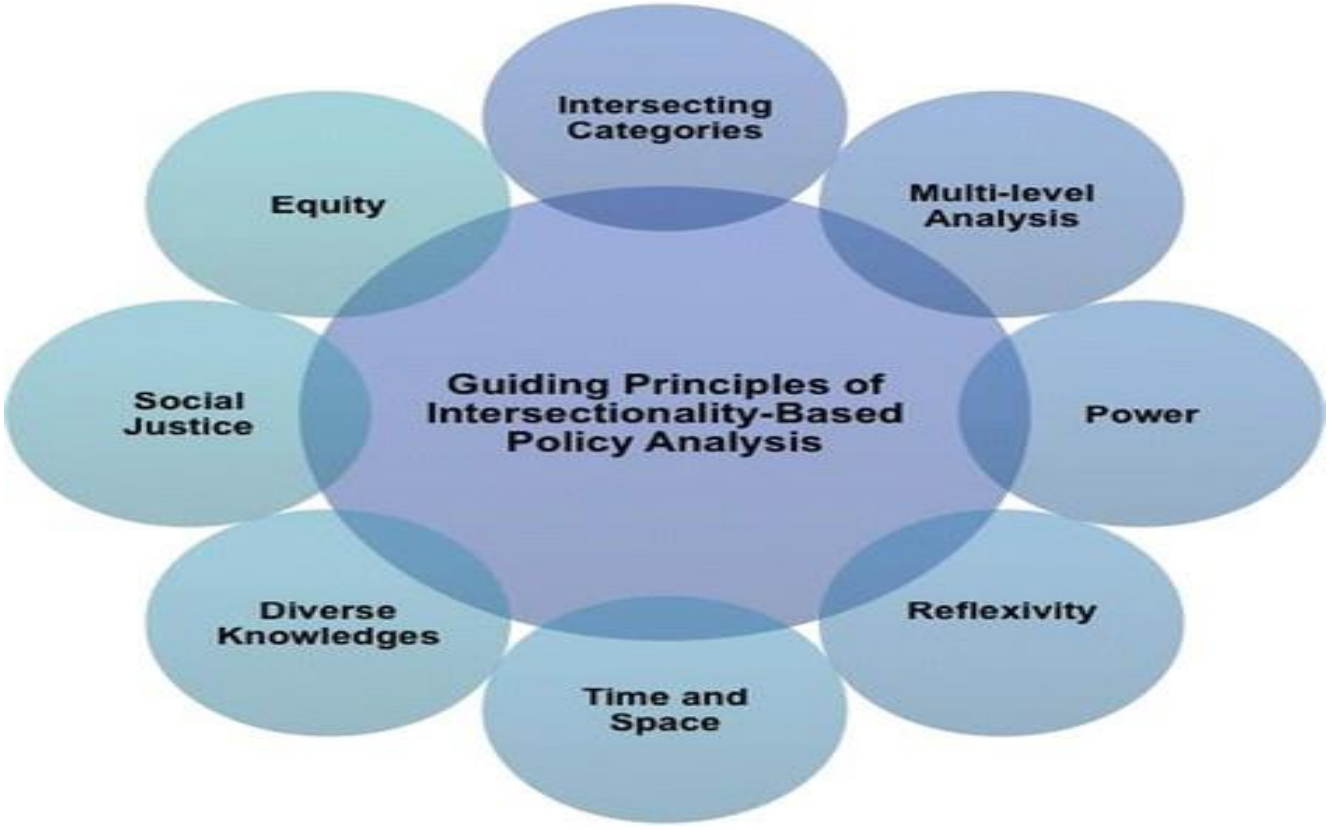

Figure 9: An intersectionality-based policy analysis framework

9.2 Prospective Audits built into EHR systems

Besides methodological innovation, there is also a need to have prospective audits embedded in electronic health record (EHR) systems. The majority of existing assessments of bias are performed post hoc, and it is after a model has been trained, validated, and maybe already deployed. This is beneficial, but does not take equal advantage of learning and adapting in real time. Models implemented in a clinical setting where patients change and systems workflows develop over time require regular re-evaluation, as they need to be ensured of fairness and effectiveness. An alternative to prospective auditing is a dynamic one. Healthcare facilities can monitor the performance of their AI models in terms of their fairness across subgroups directly in the EHR systems. Their systems would be configured to allow them to know when a model begins to perform significantly worse in one subset or when a trend toward misclassification occurs. What would be effective would be to incorporate these alerts into clinical quality

dashboards that staff already use. As such, fairness becomes an active element of routine checks, say as closely as infection rates or adverse drug events.

Implementation of the prospective audits would also foster a culture of transparency and accountability (1). When monitoring fairness is a live component of healthcare delivery, it changes the conversation to focus on preventing rather than acting on correction. This poses a new challenge to the researcher, a challenge that is not new, though, in that methods addressing it are emerging, techniques that require thinking beyond the development of models, but also the design of systems that have to be built to bring them to the clinical environment.

Focusing on patient-centered research, future studies should examine the possibility of patients participating in the establishment of fairness benchmarks themselves. Concepts such as equal sensitivity or demographic parity are traditional ways of measuring only a part of what is happening. Lived experience data, community engagement, and patient-reported outcomes are valuable to the meaningful length of the usage of AI tools as it pertains to equity in care. Overall, the future of clinical AI fairness is in greater study and understanding, as well as active evaluation and monitoring at a system level. It is possible to change the invisibility of disparities through intersectional bias analysis. The prospective audits will be able to make sure that fairness is not just an aim but a process. All of this in combination will create a more intelligent and not just intelligent healthcare AI ecosystem.

10. Conclusion

Auditing of sepsis prediction models poses a grim, but much-needed reality that not all patients are equally served by the technical sophistication linked with many AI tools deployed in healthcare today. The results, which were obtained in the course of conducting this study, point to the manifestation of racial and gender differences in terms of models performing. With the data disaggregated, it can be seen that the predictions are less accurate, the false negative rates are higher, and the clinical responses are slower among Black patients, Hispanic patients, women, and other underrepresented populations. Such inequalities do not exist by chance, nor are they peculiar to this or that period, but are products of underlying patterns in the data and construction of the models themselves. This seemingly optimistic algorithm, designed to be by all means impartial, is, in reality, a replica of historical and structural biases within the medical field.

This paper points out that inequity related to AI in clinics is not unavoidable. There are mitigation strategies that can mitigate such concerns, and when used responsibly, these strategies can help make AI systems more just as well as effective. The interventions of rebalancing training datasets, using fairness-aware training constraints, recalibrating decision thresholds, and enhancing transparency have demonstrated the possibility to mitigate harm and decrease performance gaps. More than technical improvement, such tools are a significant step towards health equity. With deliberate focus aimed at fairness during development and implementation, AI can contribute to superior results among all groups of patients, not the disparities that it would help address.

It cannot all be put on the developers or data scientists to fix these inequities. Healthcare organizations need to understand that the concept of fairness in AI is not a hypothetical or secondary concern, but a clinical safety concern. Such a realization requires an operational and cultural change. Institutions should make regular audits of bias in their assessment procedures, report on performance by demographic subgroups, and establish mechanisms for monitoring disparities continuously. This higher form of accountability is less absolute, without which even the best-designed AI systems threaten to exacerbate the human-made disparities in access, treatment, and outcomes. Inclusive AI needs to be created with the data itself as the starting point. The data collection procedures should be deliberately representative of the marginalized communities, and their identities must not only be included in them but also represented fairly. Additionally, the similarities amongst the communities that have been impacted by incongruity must participate in the design of the requirements against which equity is measured. Technical prowess is one thing, but ethical honesty and experienced realization must go with it. It is important to construct clinical AI systems end-to-end with equity in mind, including the data collection point and where a decision about the patient is based on a prediction.

The fundamental question that is entwined in this issue is whether artificial intelligence in its current form and its implementation phase can be trusted to take care of all people in a similar way as they would like to be taken care of. The results of this audit indicate that the answer is still no at the moment. They even hint at a future where that

response may be different. Fairness is not a long-run or theoretical cosmos. It is a quantifiable, doable, and ethically imperative aspect of any Artificial Intelligence being utilized in the field of medicine. Such systems will be required not merely to predict what will occur, but also to indicate a value of dignity, justice, and compassion. Achieving this vision will not be an easy thing. It involves long-term dedication, institutional audacity, and the ability to fight the assumptions incorporated in technology. However, the stakes are as high as they can be. To a Black mother giving birth, a Hispanic grandfather in intensive care, or a non-binary teen going to an urgent care facility, the use of AI has to do more than be efficient; it must be equitable. Focusing on the principle of fairness at all levels of the AI lifecycle provides the healthcare sector with the opportunity to bring it closer to the moment when technology will be not only smart, but also fair.

References

1. Brender, N., Yzeiraj, B., & Fragniere, E. (2015). The management audit as a tool to foster corporate governance: an inquiry in Switzerland. *Managerial Auditing Journal*, *30*(8/9), 785-811. https://doi.org/10.1108/MAJ-03-2014-1013
2. Buolamwini, J. A. (2017). *Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers* (Doctoral dissertation, Massachusetts Institute of Technology).
3. Chan, C. W., Farias, V. F., & Escobar, G. J. (2017). The impact of delays on service times in the intensive care unit. *Management Science*, *63*(7), 2049-2072. https://doi.org/10.1287/mnsc.2016.2441
4. Chavan, A. (2021). Exploring event-driven architecture in microservices: Patterns, pitfalls, and best practices. International Journal of Software and Research Analysis. https://ijsra.net/content/exploring-event-driven-architecture-microservices-patterns-pitfalls-and-best-practices
5. Chavan, A. (2022). Importance of identifying and establishing context boundaries while migrating from monolith to microservices. Journal of Engineering and Applied Sciences Technology, 4, E168. http://doi.org/10.47363/JEAST/2022(4)E168
6. Chimakonam, J. O., & Ofana, D. E. (2022). How intercultural philosophy can contribute to social integration. *Journal of Intercultural Studies*, *43*(5), 606-620. https://doi.org/10.1080/07256868.2022.2063824
7. Elias, A., & Paradies, Y. (2021). The costs of institutional racism and its ethical implications for healthcare. *Journal of bioethical inquiry*, *18*(1), 45-58. https://link.springer.com/article/10.1007/s11673-020-10073-0
8. Grant, M., Wilford, A., Haskins, L., Phakathi, S., Mntambo, N., & Horwood, C. M. (2017). Trust of community health workers influences the acceptance of community-based maternal and child health services. *African Journal of Primary Health Care and Family Medicine*, *9*(1), 1-8. https://hdl.handle.net/10520/EJC-96ce469f4
9. Gumede, W., Bob, U., de Beer, D., Lues, R., & Anelich, L. (2020). Position paper: priority setting for interventions in pre-and post-pandemic management: the case of covid-19. https://www.anelichconsulting.co.za/wp-content/uploads/2020/06/SATN-COVID-19-Position-Paper.pdf
10. Konneru, N. M. K. (2021). Integrating security into CI/CD pipelines: A DevSecOps approach with SAST, DAST, and SCA tools. *International Journal of Science and Research Archive*. Retrieved from https://ijsra.net/content/role-notification-scheduling-improving-patient
11. Kumar, A. (2019). The convergence of predictive analytics in driving business intelligence and enhancing DevOps efficiency. International Journal of Computational Engineering and Management, 6(6), 118-142. Retrieved from https://ijcem.in/wp-content/uploads/THE-CONVERGENCE-OF-PREDICTIVE-ANALYTICS-IN-DRIVING-BUSINESS-INTELLIGENCE-AND-ENHANCING-DEVOPS-EFFICIENCY.pdf
12. Lesko, C. R., Buchanan, A. L., Westreich, D., Edwards, J. K., Hudgens, M. G., & Cole, S. R. (2017). Generalizing study results: a potential outcomes perspective. *Epidemiology*, *28*(4), 553-561. https://journals.lww.com/epidem/toc/2017/07000
13. Lionetti, F., Aron, A., Aron, E. N., Burns, G. L., Jagiellowicz, J., & Pluess, M. (2018). Dandelions, tulips and orchids: Evidence for the existence of low-sensitive, medium-sensitive and high-sensitive individuals. *Translational psychiatry*, *8*(1), 24. https://www.nature.com/articles/s41398-017-0090-6
14. Lu, J., Sattler, A., Wang, S., Khaki, A. R., Callahan, A., Fleming, S., ... & Shah, N. H. (2022). Considerations in the reliability and fairness audits of predictive models for advance care planning. *Frontiers in Digital*

Health, *4*, 943768. https://doi.org/10.3389/fdgth.2022.943768

15. Maney, D. L. (2016). Perils and pitfalls of reporting sex differences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1688), 20150119. https://doi.org/10.1098/rstb.2015.0119
16. Marlow, S., & Swail, J. (2014). Gender, risk and finance: why can't a woman be more like a man?. *Entrepreneurship & Regional Development*, *26*(1-2), 80-96. https://doi.org/10.1080/08985626.2013.860484
17. Mazurenko, O., Richter, J., Swanson-Kazley, A., & Ford, E. (2016). Examination of the relationship between management and clinician agreement on communication openness, teamwork, and patient satisfaction in the US hospitals. *Journal of Hospital Administration*, *5*(4), 20-27. http://dx.doi.org/10.5430/jha.v5n4p20
18. Nyati, S. (2018). Transforming telematics in fleet management: Innovations in asset tracking, efficiency, and communication. International Journal of Science and Research (IJSR), 7(10), 1804-1810. Retrieved from https://www.ijsr.net/getabstract.php?paperid=SR24203184230
19. Prescott, S. L., & Logan, A. C. (2018). From authoritarianism to advocacy: lifestyle-driven, socially-transmitted conditions require a transformation in medical training and practice. *Challenges*, *9*(1), 10. https://doi.org/10.3390/challe9010010
20. Raju, R. K. (2017). Dynamic memory inference network for natural language inference. International Journal of Science and Research (IJSR), 6(2). https://www.ijsr.net/archive/v6i2/SR24926091431.pdf
21. Sardana, J. (2022). Scalable systems for healthcare communication: A design perspective. *International Journal of Science and Research Archive*. https://doi.org/10.30574/ijsra.2022.7.2.0253
22. Sardana, J. (2022). The role of notification scheduling in improving patient outcomes. *International Journal of Science and Research Archive*. Retrieved from https://ijsra.net/content/role-notification-scheduling-improving-patient
23. Singh, V. (2022). Integrating large language models with computer vision for enhanced image captioning: Combining LLMS with visual data to generate more accurate and context-rich image descriptions. Journal of Artificial Intelligence and Computer Vision, 1(E227). http://doi.org/10.47363/JAICC/2022(1)E227
24. Singh, V. (2022). Visual question answering using transformer architectures: Applying transformer models to improve performance in VQA tasks. Journal of Artificial Intelligence and Cognitive Computing, 1(E228). https://doi.org/10.47363/JAICC/2022(1)E228
25. Slobogean, G. P., Giannoudis, P. V., Frihagen, F., Forte, M. L., Morshed, S., & Bhandari, M. (2015). Bigger data, bigger problems. *Journal of orthopaedic trauma*, *29*, S43-S46. https://journals.lww.com/jorthotrauma/toc/2015/12001
26. Sukhadiya, J., Pandya, H., & Singh, V. (2018). Comparison of Image Captioning Methods. *INTERNATIONAL JOURNAL OF ENGINEERING DEVELOPMENT AND RESEARCH*, *6*(4), 43-48. https://rjwave.org/ijedr/papers/IJEDR1804011.pdf
27. Weiss, D., & Eikemo, T. A. (2017). Technological innovations and the rise of social inequalities in health. *Scandinavian journal of public health*, *45*(7), 714-719. https://doi.org/10.1177/1403494817711371
28. Yanisky-Ravid, S., & Hallisey, S. (2018). 'Equality and Privacy by Design': Ensuring Artificial Intelligence (AI) Is Properly Trained & Fed: A New Model of AI Data Transparency & Certification As Safe Harbor Procedures. *Available at SSRN 3278490*. https://dx.doi.org/10.2139/ssrn.3278490
29. Yoon, J., Zame, W. R., & Van Der Schaar, M. (2018). Estimating missing data in temporal data streams using multi-directional recurrent neural networks. *IEEE Transactions on Biomedical Engineering*, *66*(5), 1477-1490.
30. Zhang, A., Xing, L., Zou, J., & Wu, J. C. (2022). Shifting machine learning for healthcare from development to deployment and from models to data. *Nature biomedical engineering*, *6*(12), 1330-1345. https://www.nature.com/articles/s41551-022-00898-y