LEX LOCALIS-JOURNAL OF LOCAL SELF-GOVERNMENT
ISSN:1581-5374 E-ISSN:1855-363X
VOL. 21, NO. S2 (2023)

ARTIFICIAL INTELLIGENCE ETHICS AND PRINCIPLES IN PUBLIC SECTOR HEALTHCARE: A GOVERNANCE FRAMEWORK FOR RESPONSIBLE AI ADOPTION IN LOCAL AND SUB-NATIONAL HEALTH SERVICES

Kalyana Krishna Kondapalli[1], Chaitanya Gunupudi[2]

[1]Technical Architect, India
[2]IT Platform Senior Cloud Engineer, India

kalyanakondapalli@gmail.com[1]
chaitanya.gunupudi970@gmail.com[2]

Abstract

The proliferation of artificial intelligence (AI) and machine learning (ML) in public health systems presents profound ethical, regulatory, and governance challenges for local and sub-national governments. AI-enabled clinical decision support systems (CDSS) leveraging ensemble methods including Random Forest (RF) and Gradient Boosting Machines (GBM), unsupervised patient stratification techniques, and SHAP-based explainability offer significant potential for improving diagnostic accuracy, risk stratification, and service efficiency. However, deployment within publicly accountable healthcare institutions raises fundamental questions of transparency, fairness, accountability, human oversight, privacy, and robustness that performance metrics alone cannot resolve. This paper provides a comprehensive analysis of the ethical principles and governance frameworks required for responsible AI adoption in local public health services. Drawing on systematic literature review, comparative analysis of international regulatory instruments OECD AI Principles (2019), UNESCO Ethics Recommendation (2021), European Commission AI Regulation Proposal (2021), Council of Europe Convention 108+ (2018), GDPR Article 22 (2018), and the US AI Bill of Rights (2022) and a technical evaluation of ML approaches to CDSS, the paper proposes a five-component governance framework encompassing pre-deployment ethics assessment, transparency and explainability requirements, ongoing monitoring and accountability, public participation, and institutional capacity development. Implementation challenges including technical barriers, institutional capacity constraints, legal complexity, and structural equity asymmetries are examined in depth. Six policy recommendations are derived for statutory reform, procurement redesign, governance capacity, citizen rights, equity benchmarking, and evaluation research. The paper argues that trustworthy, equitable AI in local public health requires governance frameworks embedded in democratic institutions and informed by the communities whose health services are at stake.

Keywords

Artificial intelligence ethics; AI governance; public sector AI; clinical decision support; explainable AI; algorithmic accountability; local self-government; health equity; SHAP; machine learning; data protection; responsible innovation.

I. INTRODUCTION

Artificial intelligence and machine learning are transforming the operational infrastructure of public health service delivery at an unprecedented pace. Clinical decision support systems (CDSS) powered by supervised and unsupervised learning are being deployed within local and sub-national governments the primary institutional layer responsible for delivering publicly funded healthcare to citizens across diagnostic imaging, patient risk stratification, treatment recommendation, and administrative resource

allocation. The scale and speed of this transformation create both extraordinary promise and substantial governance risk for public health institutions.

The technical promise is well documented. Supervised ensemble learning methods particularly Random Forest (RF) and Gradient Boosting Machines (GBM) have consistently demonstrated strong predictive performance across clinical prediction tasks including hospital readmission, mortality risk, disease progression, and treatment response, achieving accuracy rates exceeding 90 percent on benchmark clinical datasets (Antoniadi et al., 2021; Sutton et al., 2020). Support Vector Machines (SVM) offer utility in high-dimensional, smaller-sample clinical settings; logistic regression provides a transparent, interpretable baseline whose coefficients can be directly reviewed by clinicians. Unsupervised techniques k-means clustering and Principal Component Analysis (PCA) enable patient stratification revealing clinically coherent subgroups not apparent through conventional diagnostic categorization, facilitating personalized care planning and resource allocation (Damaraji et al., 2020).

The integration of SHAP (Shapley Additive Explanations) into explainable AI (XAI) pipelines has introduced a degree of post-hoc interpretability previously unavailable, generating both global feature importance rankings across patient cohorts and granular patient-level explanations reviewable by clinicians at the point of care (Antoniadi et al., 2021). Applied to gestational diabetes prediction, SHAP correctly surfaces clinically recognized risk factors as dominant predictors, providing clinicians with evidence of clinical coherence (Du et al., 2022). Synthetic data generation further expands ML applicability in privacy-sensitive clinical settings where real patient data is insufficient or too sensitive for direct use in model training (Rajotte et al., 2022).

Yet AI deployment in public health is, at its core, a governance challenge rather than a technical one. A model achieving 94 percent aggregate accuracy may perform substantially worse—and inequitably for demographic groups underrepresented in training data (Obermeyer et al., 2019). A SHAP interface providing technically accurate explanations may overwhelm clinicians and contribute to automation bias rather than supporting substantive professional oversight (Liu et al., 2020). A procurement process prioritizing vendor capability and cost over ethical due diligence may embed systems that erode institutional accountability and citizen rights in ways that are difficult to reverse (Meijer & Thaens, 2021).

These concerns are neither hypothetical nor peripheral. The widely cited case of a commercial health risk stratification tool assigning systematically lower risk scores to Black patients with equivalent clinical needs to white patients demonstrated that poorly governed clinical AI can systematically reproduce and amplify structural health inequities (Obermeyer et al., 2019). The growing international regulatory landscape OECD AI Principles (2019), UNESCO Ethics Recommendation (2021), European Commission AI Regulation Proposal (2021), Council of Europe Convention 108+ (2018), and the US AI Bill of Rights (2022) reflects global consensus that principled governance, not technical optimization alone, is required for AI in high-stakes public health domains.

A critical gap nonetheless persists between this growing normative consensus and the operational governance realities of local and sub-national government. International frameworks establish principles but do not design the institutional arrangements, procurement frameworks, monitoring systems, or capacity development strategies required to operationalize those principles within the specific constraints of local health authorities. This paper addresses that gap. It proposes a comprehensive governance framework grounded in six internationally recognized ethical principles transparency, accountability, fairness and non-discrimination, privacy and data protection, human oversight, and robustness and safety

and designed with explicit attention to the institutional realities of sub-national government. The framework is informed by comparative regulatory analysis, systematic literature review, and technical assessment of ML approaches to CDSS.

The paper is organized as follows. Section II presents a comprehensive literature review covering AI ethics, ML in CDSS, public sector governance, XAI and clinical trust, and data equity. Section III describes the methodology. Section IV presents the governance framework, including technical foundations, core principles, regulatory landscape, and five governance components. Section V examines implementation challenges. Section VI presents policy implications. Section VII concludes.

II. LITERATURE REVIEW

A. The Ethics of AI in Healthcare

The ethical analysis of AI in healthcare has evolved from broader bioethical traditions beneficence, non-maleficence, autonomy, and justice to address the specific moral challenges of algorithmic systems operating at scale in complex clinical environments. Mittelstadt et al. (2016) provide an influential mapping of AI ethics across six concern clusters: inconclusive evidence, inscrutable evidence, misguided evidence, unfair outcomes, transformative effects, and traceability. Each cluster has direct governance implications: opacity demands explainability obligations; unfair outcomes demand bias assessment and equity monitoring; traceability demands audit infrastructure. Floridi et al. (2018) articulate five core AI ethics principles beneficence, non-maleficence, autonomy, justice, and explicability noting that explicability is distinctively novel to AI ethics, reflecting the challenge of algorithmic opacity absent from prior technological ethics frameworks.

The dominant integrating concept in both academic and policy discourse is "trustworthy AI." The European Commission's High-Level Expert Group on AI (HLEG AI, 2019) articulates seven requirements: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental wellbeing; and accountability. Critically, the HLEG framework distinguishes between lawful AI, ethical AI, and robust AI, insisting all three dimensions are necessary for genuine trustworthiness. These requirements have since been operationalized in the European Commission's 2021 Proposal for an AI Regulation, which classifies clinical decision support systems as high-risk AI requiring mandatory conformity assessment, technical documentation, logging, transparency obligations, accuracy and robustness requirements, and mandatory human oversight mechanisms prior to market placement.

A central and persistent tension in healthcare AI ethics is the trade-off between predictive performance and interpretability. Conventional supervised models logistic regression, shallow decision trees offer high interpretability but typically underperform on complex clinical prediction tasks relative to ensemble and deep learning methods. Deep learning architectures achieve superior discriminative performance but produce outputs clinicians and patients cannot readily interrogate or contest (Arrieta et al., 2020). This opacity is ethically problematic because it forecloses meaningful contestation of AI-informed decisions affecting patient care a right protected under GDPR Article 22 and Convention 108+ and because it impedes the professional accountability that clinicians owe to their patients.

Recent scholarship cautions against conflating explainability with accountability. Wachter et al. (2017) argue that post-hoc XAI methods including SHAP and LIME describe what a model did rather cannot

support meaningful contestation of algorithmic decisions without counterfactual information. An explanation identifying which features drove a prediction does not reveal whether the model is fair, whether the features are ethically appropriate proxies for clinical risk, or whether the patient has any practical mechanism to challenge the outcome. Genuine accountability requires institutional structures governance bodies, complaint mechanisms, audit processes not merely technical explanations.

B. Machine Learning in Clinical Decision Support

The clinical AI literature encompasses a substantial body of evidence on the performance of supervised and unsupervised ML approaches across diverse prediction tasks. Random Forest has emerged as particularly widely used for clinical prediction given its resistance to overfitting, capacity for high-dimensional mixed-type data, and natural generation of feature importance scores (Sutton et al., 2020). Its ensemble approach combining predictions from multiple decision trees trained on bootstrap samples with random feature subsets reduces variance while maintaining predictive flexibility. Gradient Boosting Machines construct trees sequentially, each correcting the residual errors of its predecessors, enabling capture of complex non-linear clinical variable interactions with typically competitive or superior performance relative to Random Forest on structured clinical data.

Support Vector Machines offer particular utility in high-dimensional, smaller-sample settings conditions common in specialized clinical domains with limited labelled training data. SVM identifies the optimal separating hyperplane by maximizing the margin between support vectors, providing competitive generalization under appropriate kernel selection. Logistic regression, while typically outperformed by ensemble methods on complex tasks, provides interpretable coefficient-level insights directly usable by clinicians and regulatory bodies, and remains valuable as a transparent baseline for governance comparison.

Unsupervised methods play a complementary role. K-means clustering partitions patients into clinically coherent groups by minimizing within-cluster variance, revealing phenotypic subgroups for example, distinct COPD exacerbation patterns or sepsis trajectories that may not align with conventional diagnostic categories (Damaraji et al., 2020). PCA reduces high-dimensional clinical feature spaces to lower-dimensional representations that preserve maximal variance, removing collinearity, supporting visualization, and stabilizing downstream supervised learning. The governance implication of unsupervised methods is that their outputs patient clusters, reduced feature spaces lack ground truth labels against which bias can be directly assessed, creating particular challenges for fairness governance.

Deep learning methods convolutional neural networks for medical imaging, recurrent and transformer architectures for sequential EHR and physiological data achieve state-of-the-art performance on complex clinical tasks including lesion detection, CT/MRI classification, ECG interpretation, and dynamic mortality prediction. However, their computational requirements, interpretability limitations, and sensitivity to distributional shift create significant governance challenges, particularly in resource-constrained local health settings where monitoring capacity and regulatory oversight are often limited (Arrieta et al., 2020).

C. SHAP-Based Explainability and Clinical Trust

SHAP (Shapley Additive Explanations) has emerged as the dominant XAI approach for structured clinical data, computing feature contributions through a game-theoretic Shapley value framework that satisfies desirable axiomatic properties of consistency, local accuracy, and missingness (Antoniadi et al., 2021).

Global SHAP analysis generates feature importance rankings across an entire patient cohort, allowing governance bodies to verify that models weight recognized clinical risk factors appropriately rather than relying on potentially problematic proxies. Local SHAP analysis generates patient-specific attribution values visualized as waterfall or force plots providing clinicians with reviewable explanations of individual predictions at the point of care.

Applied across clinical prediction tasks, SHAP has demonstrated clinical coherence: pre-pregnancy BMI, maternal age, and family history correctly surfaced as dominant predictors of gestational diabetes risk (Du et al., 2022); established neurological biomarkers correctly weighted in brain imaging models (Dhombres et al., 2022). These applications illustrate the dual governance function of SHAP: validating clinical coherence for oversight bodies and supporting clinician comprehension for meaningful human oversight. In ALS quality-of-life prediction, SHAP-based explanations were shown to directly inform clinical conversations and personalized care planning (Antoniadi et al., 2022), demonstrating integration potential beyond technical validation.

Nonetheless, important limitations require governance acknowledgement. SHAP explanations characterize model behavior rather than causal mechanisms high SHAP attribution to a feature indicates its contribution to prediction, not causal clinical relevance. Under distributional shift, SHAP attributions may become misleading, identifying features as important in ways that reflect dataset artefacts rather than genuine clinical relationships. The cognitive demand of SHAP interfaces creates human-factors risks automation bias when clinicians over-defer to algorithmic outputs, and explanation fatigue when high-volume alert generation impedes substantive review (Liu et al., 2020). Governance frameworks cannot treat XAI capability as sufficient accountability; institutional structures must complement technical explainability.

D. AI Governance in the Public Sector

AI governance in public institutions raises challenges qualitatively distinct from those in commercial settings. Public institutions operate under legal obligations of non-discrimination, due process, administrative accountability, and democratic legitimacy that create unique governance requirements (Danaher et al., 2017). Algorithmic decision-making in government engages constitutional principles and human rights frameworks including equality rights, the right to a fair hearing, and the right to an effective remedy whose protection requires AI governance structures that go beyond commercial compliance frameworks (Veale & Binns, 2017).

Research on AI governance at the local government level has consistently identified significant institutional barriers. Meijer & Thaens (2021) document the "algorithmizing" of local bureaucratic decision-making across European jurisdictions, finding that algorithmic systems are frequently deployed without the governance structures required to manage them, with procurement decisions driven by vendor presentations and cost pressures rather than structured ethical assessment. The resulting information asymmetry between public-sector deployers and commercial AI vendors who hold substantial advantages in technical expertise, proprietary data, and institutional resources creates structural vulnerability in public sector AI governance.

The concept of algorithmic accountability has been proposed as a unifying framework for public sector AI governance, encompassing transparency obligations (duty to disclose the existence and nature of algorithmic systems affecting citizens), auditability (capacity for independent scrutiny of system

performance and decision processes), and contestability (right of affected persons to challenge and seek review of algorithmic outcomes) (Diakopoulos, 2016). These three dimensions directly correspond to the governance requirements of GDPR Article 22 which provides rights to human intervention, explanation, and contestation for automated decisions with significant effects and Convention 108+ which requires that individuals subject to automated decisions can express their view, obtain human review, and receive a decision explanation (Council of Europe, 2018).

The vendor relationship itself represents a structural governance challenge in public sector AI procurement. Commercial contracts typically protect proprietary model details through intellectual property rights, making independent algorithmic audit impossible without specific contractual provisions. Service level agreements focus on system uptime and throughput rather than ethical performance, and typically do not include the ongoing monitoring, bias reporting, or right-to-audit provisions that responsible governance requires. Addressing these structural deficits requires procurement frameworks specifically designed for AI governance imposing documentation, transparency, audit access, ongoing monitoring obligations, and contractual liability for governance failures as standard contract conditions (Meijer & Thaens, 2021).

E. Data Governance, Privacy, and Health Equity

The ethical deployment of AI in public health is inseparable from robust health data governance. GDPR imposes strict requirements on the processing of health data classified as a special category requiring explicit consent or specific legal authority including mandatory Data Protection Impact Assessments (DPIAs) for high-risk processing, data minimisation and purpose limitation obligations, and Article 22 safeguards against automated individual decision-making with significant effects. Local health authorities deploying AI systems must navigate these requirements while managing practical complexities of cross-institutional data sharing, data quality assurance, and interoperability across heterogeneous clinical information systems.

Synthetic data generation has emerged as a governance-relevant approach to the tension between ML training data utility and privacy protection (Rajotte et al., 2022). By generating statistically representative synthetic datasets that do not correspond to real patients, models can be trained and validated in settings where real patient data is too sensitive to share or where training data volume is insufficient. However, synthetic data introduces its own governance considerations: statistical fidelity may be imperfect, particularly for rare conditions or underrepresented demographic groups, potentially introducing subtle biases absent from real data governance. Governance frameworks must address these limitations explicitly rather than treating synthetic data as a governance-neutral alternative to real patient data.

Health equity concerns constitute a critical dimension of AI data governance that extends beyond technical debiasing. Historical clinical datasets encode patterns of structural inequality in healthcare access, diagnosis, and treatment inequalities associated with race, gender, socioeconomic status, disability, and geography. AI systems trained on such data do not merely reflect historical inequity; they operationalize it at scale within future clinical decision-making, potentially creating self-reinforcing cycles in which algorithmically reinforced access disparities generate training data that perpetuates those disparities in subsequent model iterations (Obermeyer et al., 2019). Addressing this risk requires deliberate governance choices about data collection, population representation, and the metrics used to evaluate model equity choices that are fundamentally political and institutional, not merely technical.

F. Identified Governance Gaps

Despite the expanding regulatory landscape, significant governance gaps persist at the sub-national level. International instruments—OECD Principles, UNESCO Recommendation, EU AI Regulation Proposal establish normative frameworks but provide limited operational guidance for local government implementation. National AI strategies address the public sector in broad terms without specifying the institutional arrangements required of local health authorities. Academic research on AI governance has concentrated at the national and supranational levels, leaving the institutional context of sub-national government where front-line AI-enabled health service delivery occurred. This paper addresses these gaps by proposing a governance framework specifically designed for local government, bridging normative principles and operational practice.

III. METHODOLOGY

This paper employs a multi-method conceptual and normative methodology structured across three analytical stages, each contributing a distinct evidential stream to the governance framework proposed in Section IV.

A. Systematic Literature Review

The first stage involved systematic review of academic literature on AI ethics in healthcare, ML in CDSS, XAI, public sector AI governance, data governance, and health equity. Databases searched included PubMed, Scopus, Web of Science, and Google Scholar using controlled vocabulary and free-text strings related to artificial intelligence, machine learning, clinical decision support, ethics, governance, explainability, fairness, accountability, and local government. Studies published up to and including 2022 were included. Grey literature from the OECD, UNESCO, the European Commission, and the Council of Europe was reviewed alongside academic sources. Technical documentation from the NIST AI Risk Management Framework and IEEE Ethically Aligned Design initiative informed governance design methodology.

B. Comparative Regulatory Analysis

The second stage comprised structured comparative analysis of six international and regional regulatory instruments: OECD AI Principles (2019); UNESCO Recommendation on the Ethics of AI (2021); European Commission AI Regulation Proposal, COM/2021/206 (2021); US Blueprint for an AI Bill of Rights (2022); Council of Europe Convention 108+ (2018); and GDPR Article 22 (2018). Each instrument was analysed against a common framework addressing ethical principles articulated specific obligations imposed on AI deployers in public health contexts enforcement mechanisms, accountability structures and operational guidance for local authorities. The comparative analysis identifies areas of normative convergence and gaps in sub-national operational guidance.

C. Technical Assessment and Framework Synthesis

The third stage comprised technical assessment of supervised and unsupervised ML approaches to CDSS characterizing performance characteristics, interpretability properties, and governance implications of RF, GBM, SVM, logistic regression, k-means, and PCA drawing on the experimental literature. SHAP-based explainability properties and limitations were assessed against governance accountability requirements. The governance framework was then synthesized by integrating outputs of the three analytical stages: normative principles from regulatory analysis, implementation requirements from literature review, and

technical governance implications from ML assessment. The methodology is conceptual rather than empirical; empirical validation through comparative case studies of governance framework implementation across diverse local authority contexts is identified as a priority for future research.

IV. ETHICAL PRINCIPLES AND GOVERNANCE FRAMEWORK

A. ML Performance Evaluation and Governance Implications

A comprehensive evaluation of ML approaches to clinical decision support provides the technical foundation for governance framework design. Table I presents indicative performance metrics for the principal ML approaches assessed in the literature across clinical prediction tasks using multimodal clinical datasets incorporating structured EHR data, laboratory results, and imaging-derived features. These figures represent representative performance profiles rather than universal benchmarks.

TABLE I. Indicative Performance Metrics for ML Approaches in Clinical Decision Support

Model	Acc.%	Prec.%	Rec.%	F1	AUC
Random Forest	92	90	88	0.89	0.95
Gradient Boosting	91	89	87	0.88	0.94
Support Vector Machine	87	85	80	0.82	0.92
Logistic Regression	83	80	78	0.79	0.88
Ensemble (RF+GBM)	94	92	91	0.91	0.97

Several governance-relevant observations emerge. First, ensemble methods particularly the RF+GBM combination consistently achieve superior performance across all standard metrics (accuracy 94%, AUC-ROC 0.97), supporting their prioritization for public health CDSS where computational resources permit. Second, the performance gap between logistic regression and ensemble methods is narrower than often assumed; in settings where interpretability, regulatory simplicity, or computational constraints outweigh marginal accuracy gains, simpler interpretable models may be preferable on governance grounds. Third, and most critically, aggregate performance metrics do not capture distributional performance variation across demographic subgroups. A system achieving 94% overall accuracy may perform at 80% or below for groups underrepresented in training data a disparity with direct clinical safety and equity implications. This observation makes subgroup performance evaluation an indispensable governance requirement, not an optional analytical supplement.

The SHAP-based explainability layer complements quantitative metrics by enabling governance validation of clinical coherence verifying that models weight recognized clinical risk factors appropriately and supporting clinician-level oversight at the point of care. However, SHAP explains model behavior rather than causal mechanisms, and requires careful human-factors design to avoid automation bias and

explanation fatigue. Governance frameworks must therefore combine XAI capability with institutional accountability structures that do not depend solely on technical explanation adequacy.

B. Core Ethical Principles

Drawing on the comparative regulatory analysis and literature review, six core ethical principles are identified as foundational to responsible AI adoption in local public health. These principles reflect genuine normative convergence across the major international instruments reviewed. Table II presents each principle with its governance requirement and specific local government implementation challenge.

TABLE II. Core AI Ethics Principles and Governance Requirements for Local Health Authorities

Principle	Governance Requirement	Local Govt Challenge
Transparency	Open algorithmic audits; XAI disclosure of decision logic to clinicians and citizens	Vendor opacity; proprietary model protection; limited public-sector AI expertise
Accountability	Clear institutional liability; audit trails; designated AI governance roles	Diffuse responsibility across functions; absence of AI accountability infrastructure
Fairness	Mandatory bias audits; subgroup performance monitoring across protected characteristics	Inequitable historical training data; demographic underrepresentation
Privacy	GDPR/DPIA compliance; data minimisation; purpose limitation;	Cross-agency data sharing; fragmented regulatory oversight

Principle	Governance Requirement	Local Govt Challenge
	Article 22 safeguards	
Human Oversight	Mandatory human-in-the-loop for high-stakes decisions; genuine override authority	Automation bias; alert fatigue; insufficient AI literacy among clinicians
Robustness	Post-deployment drift monitoring; validated fail-safes; lifecycle governance	No post-deployment monitoring resources; siloed IT governance structures

Transparency requires that local health authorities disclose the existence, purpose, data inputs, and decision logic of AI systems affecting health decisions to both clinicians and citizens. It is a precondition for democratic accountability and public trust (Diakopoulos, 2016). Full algorithmic disclosure is not required proprietary model weights need not be published but meaningful transparency demands sufficient information for affected persons to understand how decisions are reached and to exercise the rights to explanation and contestation protected by GDPR Article 22 and Convention 108+.

Accountability requires clear institutional responsibility for the consequences of AI-enabled decisions. This means identifying responsible officers and governance bodies, establishing audit mechanisms with genuine investigative capacity, and ensuring clinicians, administrators, and elected representatives retain meaningful oversight rather than delegating accountability to automated processes (Liu et al., 2020). The accountability framework must extend through the vendor relationship, with contracts imposing specific obligations and preserving local health authority rights to audit, override, and withdraw AI systems.

Fairness and non-discrimination require that AI systems do not produce systematically worse outcomes for population groups defined by protected characteristics. This demands both prospective governance representative training data and subgroup performance evaluation at deployment and reactive governance continuous monitoring for emergent disparities and prompt remedial action (Obermeyer et al., 2019; Veale & Binns, 2017). In public health, where equitable access to care is a legal and ethical obligation, algorithmic inequity constitutes a potential violation of equality law and human rights, not merely a technical performance deficiency.

Privacy and data protection require full compliance with GDPR, including DPIAs before AI deployment, data minimisation and purpose limitation, and operationalization of Article 22 rights for patients subject to AI-informed clinical decisions (Kondapalli & Gunupudi, 2018). Health data among the most sensitive personal data requires heightened procedural protections and meaningful governance oversight of cross-institutional data flows and sharing arrangements.

Human oversight requires substantive human review of algorithmic outputs before consequential clinical decisions, not nominal human presence. This is a governance commitment to the irreducible professional accountability of clinicians in patient care (European Commission HLEG AI, 2019). Meaningful oversight requires clinician AI literacy, adequate review time, genuine override authority, and override documentation. Governance frameworks must actively design for substantive oversight rather than assuming that the presence of a human decision-maker satisfies accountability requirements.

Robustness and safety require reliable system performance across real-world clinical conditions including data quality variability, distributional shift, rare conditions, and edge cases with clear protocols for monitoring, retraining, and system retirement (Antoniadi et al., 2021). The dynamic nature of clinical practice new treatments, evolving disease patterns, changing coding conventions means that AI systems performing adequately at deployment may degrade without active lifecycle governance, creating patient safety risks that static governance assessments cannot anticipate.

C. Comparative Regulatory Landscape

Table III presents a comparative overview of the major international and regional frameworks relevant to AI governance in local public health. The analysis reveals both the scope of emerging international consensus and the significant implementation gap facing sub-national governments.

TABLE III. Comparative Overview of International AI Governance Frameworks for Local Public Health

Framework	Year	Core Ethical Focus	Local Health Relevance
OECD AI Principles	2019	Human-centred values; transparency; accountability	Normative basis for national and sub-national AI policy
UNESCO AI Ethics Recommendation	2021	Human dignity; fairness; data governance	Global normative framework for public health AI governance
EU AI Regulation Proposal	2021	Risk-based categorization;	CDSS classified high-risk;

Framework	Year	Core Ethical Focus	Local Health Relevance
		conformity assessment	mandatory human oversight
US AI Bill of Rights	2022	Anti-discrimination; human alternatives; notice	Applicable to automated systems in public health programmed
Council of Europe Conv. 108+	2018	Data protection; automated decision safeguards	Binding obligations on AI-enabled health data processing
GDPR Article 22	2018	Right to human review; explanation; contest automated decisions	Directly applicable to clinical risk scoring and triage AI

The European Commission's 2021 AI Regulation Proposal provides the most operationally detailed regulatory framework for high-risk clinical AI, establishing specific requirements for technical documentation, logging, transparency information to deployers, accuracy and robustness validation, and mandatory human oversight prior to and during deployment. Critically, it applies not only to AI system developers but to deployers including public health authorities procuring commercial AI creating direct regulatory obligations for local governments previously unsubject to technology-specific regulation. GDPR Article 22 provides directly applicable rights for patients subject to automated clinical decision-making including the right to human intervention, explanation, and contestation that local health authorities must operationalize within governance arrangements. Convention 108+ provides binding data protection safeguards for automated processing across signatory states, directly applicable to AI-enabled health data systems.

The comparative analysis reveals a consistent gap: normative completeness at the international level contrasts with limited operational guidance at the sub-national level. None of the instruments reviewed provides specific guidance on the institutional arrangements, governance procedures, or capacity

development strategies required by local health authorities. This gap is the central governance design challenge the following framework addresses.

D. The Five-Component Governance Framework

The governance framework proposed comprises five interconnected components, each addressing a distinct but mutually reinforcing dimension of responsible AI governance in local public health. Table IV presents an overview of components, key activities, responsible actors, and success indicators.

TABLE IV. Governance Framework: Components, Activities, Actors, and Success Indicators

Component	Key Activities	Responsible Actors	Success Indicators
Pre-Deployment Ethics Assessment	Risk classification; DPIA; bias evaluation; oversight design	AI Ethics Officer; Clinical Lead; Legal; Community Rep.	Assessment completed and published before deployment
Transparency & Explainability	SHAP explanations; public AI register; citizen disclosure	Data Science Team; Clinical End-Users; Comms.	Explanation quality ratings; public register live within 30 days
Ongoing Monitoring	Subgroup monitoring; drift detection; audit trails; retraining	AI Ethics Officer; IT Governance; Clinical Audit	Monthly performance reports: disparity alerts actioned
Public Participation	Patient consultation; community governance representation	Patient Experience Lead; Elected Members; Community Orgs.	Consultation events per deployment; feedback incorporated
Capacity Development	AI literacy training; ethical procurement;	HR/Training; Procurement; Regional AI Governance Body	Training completion rates: procurement

Component	Key Activities	Responsible Actors	Success Indicators
	learning networks		framework adopted

1) Pre-Deployment Ethics Assessment

Before deploying any AI system in a clinical context, local health authorities must conduct a structured ethics assessment addressing: the intended clinical purpose and risk classification under applicable regulatory frameworks; an evaluation of training data representativeness and demographic characteristics; a full DPIA as required under GDPR; assessment of potential discriminatory outcomes including examination of model performance across relevant demographic subgroups; and an analysis of proposed human oversight arrangements, including override mechanisms and documentation requirements. The assessment must be reviewed by a multidisciplinary committee with clinical, legal, data protection, and community representation. It should be documented, publicly available in an appropriate format, and updated at defined intervals throughout the system lifecycle (Liu et al., 2020). Systems classified as high-risk including CDSS informing diagnosis, treatment planning, or patient risk stratification require full assessment prior to deployment and periodic reassessment thereafter.

2) Transparency and Explainability

AI systems in clinical decision support roles must incorporate explainability mechanisms capable of generating comprehensible explanations for predictions at both cohort and individual patient levels. SHAP-based feature attribution is recommended for structured clinical data given its theoretical properties and integration with mainstream ML tooling (Antoniadi et al., 2021). Explanation interfaces must be co-designed with clinical end-users physicians, nurses, pharmacists ensuring explanations are presented in formats appropriate to clinical context, time constraints, and cognitive demands of the clinical role. Local health authorities must maintain a public register of AI systems in clinical use, specifying purpose, data inputs, output types, and governance arrangements for each system. Patients subject to AI-informed decisions must have accessible rights to explanation under GDPR Article 22 and Convention 108+ (Council of Europe, 2018), with explanations provided in plain, non-technical language.

3) Ongoing Monitoring and Accountability

AI governance requires continuous post-deployment monitoring across the full range of metrics relevant to both clinical efficacy and ethical compliance. Monitoring must include: regular evaluation of standard performance metrics accuracy, precision, recall, F1-score, AUC-ROC against pre-specified benchmarks; systematic subgroup performance assessment across demographic characteristics with pre-specified disparity thresholds triggering escalated review; data quality and distributional monitoring to detect shift; and analysis of clinician override patterns to identify systematic disagreement between clinical judgment and algorithmic recommendation. Audit trails of algorithmic outputs and clinician responses including overrides and stated reasons must be maintained for governance review, regulatory inspection, and subject access requests. A designated AI Ethics Officer or equivalent with cross-functional authority, adequate resources, and direct reporting lines to senior leadership must be responsible for oversight, with authority to escalate concerns, mandate retraining, and initiate system withdrawal (Meijer & Thaens, 2021).

4) Public Participation and Civic Engagement

The deployment of AI in publicly funded health services involves value choices that cannot be legitimately resolved by technical experts or administrators alone. Trade-offs between efficiency and equity, algorithmic automation and human judgment, innovation and precaution are fundamentally political choices that belong to democratic communities (Danaher et al., 2017). Local health authorities must establish mechanisms for meaningful not tokenistic public participation in AI governance, including: structured consultation with patient representative groups, community health advocates, and underrepresented populations before major deployments; representation of community voices on AI governance committees; accessible public information about AI systems in use and the governance arrangements governing them; and formal processes for receiving and responding to citizen feedback, complaints, and requests for independent review of AI-informed decisions. Evidence from AI governance in obstetric and maternal health demonstrates that substantive community engagement identifies clinically relevant concerns and value priorities that technical teams do not independently surface, improving both governance quality and democratic legitimacy (Dhombres et al., 2022).

5) Capacity Development and Institutional Learning

The governance framework presupposes institutional capacity in AI ethics, data governance, clinical informatics, regulatory compliance, and community engagement that many local health authorities do not currently possess. The capacity gap between governance requirements and available expertise is among the most significant barriers to ethical AI in local public health and will not be addressed without deliberate investment. Required capacity development includes: formal AI literacy and governance training for clinical, administrative, and elected leadership; development of ethical procurement frameworks imposing specific ethics requirements on AI vendors and supporting structured ethics assessment at procurement stage; creation of regional or national shared governance resources ethics assessment templates, audit protocols, governance guidance reducing the burden on individual authorities; and establishment of cross-authority learning networks for governance experience sharing, including lessons from AI failures. AI governance capacity must be understood as an ongoing institutional learning process continuously adapting to technological evolution, regulatory development, and emerging evidence on governance effectiveness (Meijer & Thaens, 2021).

V. IMPLEMENTATION CHALLENGES

A. Technical Barriers

Clinical data in most public health systems is characterized by incompleteness, inconsistency, and heterogeneity creating fundamental challenges for AI system development and deployment. Missing data non-randomly distributed across demographic groups creates systematic bias risk. Inconsistent coding practices across institutions introduce noise differentially affecting model performance across deployment sites. Distributional shift arising from changes in patient populations, treatment protocols, or data collection standards degrades model performance over time without active monitoring and recalibration (Sutton et al., 2020).

Model interpretability limitations require honest governance acknowledgement. Post-hoc SHAP explanations characterize model behavior rather than causal mechanisms, and their reliability under distributional shift or adversarial conditions has been questioned (Arrieta et al., 2020). Deep learning architectures achieving state-of-the-art performance on imaging and sequential clinical data present

significant interpretability challenges that SHAP addresses only partially. Governance frameworks relying on XAI as a primary accountability mechanism must be complemented by institutional structures human oversight requirements, contestation mechanisms, audit processes that do not depend solely on explanation adequacy. Synthetic data generation offers partial mitigation of training data scarcity and privacy constraints but requires governance of statistical fidelity, particularly for rare conditions and underrepresented populations (Rajotte et al., 2022).

B. Institutional and Organizational Barriers

AI governance in local health authorities requires expertise spanning data science, clinical medicine, regulatory compliance, ethics, procurement, and community engagement a combination rarely available within a single organization. The resulting capacity gap creates significant vulnerability to commercial AI sector representatives with far greater technical expertise and institutional resources than public-sector counterparts. Organizational cultures within local health systems may further resist governance requirements: clinical cultures valuing professional autonomy may resist documentation and oversight obligations; administrative cultures focused on efficiency may deprioritize rigorous ethics assessment; political cultures may resist transparency requirements that attract public scrutiny of AI failures.

Vendor relationships create structural governance challenges. Commercial contracts protect proprietary model details, making independent algorithmic audit impossible without specific contractual provisions. SLAs focus on uptime and throughput rather than ethical performance. The pace of commercial AI development may outstrip public procurement capacity to evaluate new capabilities and risks. Addressing these challenges requires procurement frameworks specifically designed for AI governance imposing documentation, transparency, audit access, monitoring reporting, and contractual liability for governance failures (Meijer & Thaens, 2021).

C. Legal and Regulatory Complexity

Local health authorities operate within overlapping and sometimes conflicting legal frameworks: GDPR data protection obligations, health-specific information governance, procurement law, equality legislation, and the emerging AI-specific regulatory corpus interact in ways creating significant compliance uncertainty. GDPR Article 22 rights human intervention, explanation, contestation for automated decisions with significant effects must be operationalized within local governance structures and communicated to patients accessibly. Convention 108+ binding obligations must be translated into institutional procedures. The 2021 EU AI Regulation Proposal's risk-based classification must be applied to each AI system in the authority's portfolio to determine applicable requirements. Cross-border dimensions commercially procured AI systems trained on data from multiple jurisdictions, hosted on infrastructure under foreign law create jurisdictional governance complexity requiring sophisticated contractual arrangements beyond current local health authority legal capacity.

D. Equity, Power Asymmetries, and Democratic Legitimacy

The most fundamental governance challenge is structural: AI adoption in local public health risks replicating and intensifying existing inequities in health service delivery and democratic participation. Communities most likely to be harmed by algorithmic bias racially marginalized groups, people with disabilities, those in poverty, residents of historically underserved areas are typically those with least political influence, least technical capacity, and least representation in governance processes (Obermeyer et al., 2019). Large technology companies hold substantial advantages in data, computational resources,

and AI expertise over public health authorities. The benefits of AI-enabled efficiency gains may concentrate among well-represented populations while costs of algorithmic errors and inequities fall disproportionately on those least able to contest them.

Addressing these structural asymmetries requires governance choices going beyond technical compliance: deliberate outreach to underrepresented communities in governance consultation; specific equity impact assessment requirements in pre-deployment ethics assessments; inclusion of equity metrics subgroup performance disparities as core performance indicators in monitoring frameworks; and political leadership explicitly committing to equitable AI as a governance priority rather than a technical aspiration.

VI. POLICY IMPLICATIONS AND RECOMMENDATIONS

Six policy recommendations are derived from the governance framework and implementation analysis, addressed to local health authorities, national oversight bodies, and international standards organizations.

First, embed AI ethics governance in the statutory and regulatory frameworks governing local health service delivery. High-level normative principles are insufficient without binding requirements, enforcement mechanisms, and resourcing. Legislative instruments should mandate pre-deployment ethics assessment for high-risk clinical AI; require ongoing subgroup performance monitoring and public reporting; establish rights to explanation, contestation, and human review for patients subject to AI-informed decisions; and create enforcement mechanisms regulatory inspection, administrative sanctions, independent review access ensuring governance requirements are substantively met.

Second, fundamentally reform AI procurement frameworks to incorporate ethical requirements as standard, enforceable contract conditions. Procurement documentation should require training data provenance and demographic characteristics; subgroup validation performance evidence; XAI capability adequate for the intended clinical use case; commitment to ongoing transparency, audit access, and monitoring reporting obligations; clear liability arrangements for AI-related clinical errors; and contractual acceptance of the authority's right to independent audit and system withdrawal. The EU AI Regulation Proposal's risk-based classification provides a useful reference for calibrating procurement requirements to system risk level. Standardized ethical procurement frameworks developed at national or regional level should be made available to reduce compliance burden on individual local authorities.

Third, establish dedicated AI governance capacity commensurate with the scale and risk profile of AI deployments. This includes AI Ethics Officers with cross-functional authority and direct reporting lines to senior leadership; formal AI literacy training for clinical, administrative, and elected leadership; and engagement with national or regional governance support bodies providing guidance, template frameworks, audit support, and learning exchange. Ensuring smaller and less-resourced authorities have access to shared governance infrastructure is critical to preventing two-tier AI ethics standards within local government.

Fourth, operationalize citizen rights to transparency, contestation, and participation through concrete accessible mechanisms. Rights existing only on paper provide no governance protection. Local health authorities should publish plain-language information about AI systems in use; establish simple, accessible procedures for patients to request explanations of AI-informed decisions and challenge outcomes; create formal complaint and review mechanisms with designated response timelines; and establish substantive community engagement programmed bringing diverse citizen voices into AI

governance at the design stage. Evidence demonstrates that genuine community engagement improves governance quality and democratic legitimacy, not merely its appearance (Dhombres et al., 2022).

Fifth, mandate subgroup performance reporting and equity benchmarks as standard components of AI performance evaluation. Aggregate metrics accuracy, precision, recall, F1-score, AUC-ROC provide an incomplete and potentially misleading picture of performance in diverse populations. Procurement documentation must require subgroup performance disaggregated by relevant demographic characteristics; ongoing monitoring must include pre-specified disparity thresholds triggering remedial action; and public performance reporting must include equity metrics alongside aggregate measures, enabling democratic accountability for AI equity outcomes (Liu et al., 2020).

Sixth, develop research and evaluation capacity to build an evidence base for effective AI governance in local public health. The current evidence on which governance arrangements reduce bias, improve equity, enhance accountability, and build public trust is thin. Rigorous evaluation including quasi-experimental designs assessing real-world impact of governance interventions on health outcomes, equity, and institutional accountability is urgently needed. International comparative research examining governance approaches across diverse local government contexts, legal systems, and institutional cultures would identify governance strategies with broad applicability and inform context-specific adaptation.

VII. CONCLUSION

Artificial intelligence and machine learning hold genuine potential to improve the quality, efficiency, and equity of public health services delivered by local and sub-national governments. Ensemble learning approaches Random Forest, Gradient Boosting Machines, and their combination achieve high predictive accuracy across clinical decision support tasks. SHAP-based explainability provides interpretable, clinician-reviewable insights into model predictions. Unsupervised patient stratification enables personalized care planning beyond conventional diagnostic categorization. Synthetic data generation addresses privacy-sensitive training data constraints. These are real and significant advances.

But technical performance is not a sufficient basis for deploying AI in publicly accountable health institutions. Transparency, fairness, accountability, human oversight, privacy, and robustness are foundational governance requirements grounded in internationally recognized ethical principles and enforceable legal obligations not optional qualities of responsible AI. The five-component governance framework proposed in this paper pre-deployment ethics assessment, transparency and explainability, ongoing monitoring and accountability, public participation, and institutional capacity development provides a structured, operationally realistic approach to meeting these requirements within the institutional constraints of local government, grounded in comparative regulatory analysis and the technical realities of ML deployment.

The framework acknowledges significant implementation challenges: technical barriers arising from data quality limitations and model interpretability constraints; institutional barriers from governance capacity gaps and vendor power asymmetries; legal complexity from overlapping regulatory obligations; and structural equity challenges from the inequitable distribution of AI governance costs and benefits across communities. These challenges are real but not insurmountable; the governance literature provides growing evidence base for interventions capable of addressing them.

Several limitations merit acknowledgement. The framework is conceptual rather than empirically validated. Its applicability across diverse local government contexts requires careful adaptation; the

governance implications of large language models, multimodal imaging AI, and federated learning systems deserve deeper analysis than the scope of this paper permits. Future research should address these gaps through comparative case studies, empirical governance intervention evaluation, and sustained engagement with the communities most directly affected by AI-enabled health service decisions.

Ultimately, decisions about the values embedded in public health AI the balance between efficiency and equity, algorithmic automation and human judgment, innovation and precaution are political and ethical choices belonging to the communities whose health services are at stake. Local governments, as the institutions closest to those communities and most directly democratically accountable for health service delivery, have both the institutional responsibility and the democratic mandate to lead in building governance frameworks that trustworthy, equitable, and accountable AI in public health requires.

REFERENCES

[1] Antoniadi, A. M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B. A., & Mooney, C. (2021). Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: A systematic review. Applied Sciences, 11(11), 5088. https://doi.org/10.3390/app11115088

[2] Antoniadi, A. M., Galvin, M., Heverin, M., Wei, L., Hardiman, O., & Mooney, C. (2022). A clinical decision support system for the prediction of quality of life in ALS. Journal of Personalized Medicine, 12(3), 75. https://doi.org/10.3390/jpm12030075

[3] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.002

[4] Cömert, Z., Yang, Z., Velappan, S., Boopathi, A. M., & Kocamaz, A. F. (2018). Performance evaluation of empirical mode decomposition and discrete wavelet transform for computerized hypoxia detection. 2018 26th Signal Processing and Communications Applications Conference (SIU), 1–4. IEEE. https://doi.org/10.1109/SIU.2018.8404193

[5] Council of Europe. (2018). Convention 108+: Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Modernised). CETS No. 108. https://www.coe.int/en/web/data-protection/convention108/modernised

[6] Cutler, D. M., Nikpay, S., & Huckman, R. S. (2020). The business of medicine in the era of COVID-19. JAMA, 323(20), 2003–2004. https://doi.org/10.1001/jama.2020.6573

[7] Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., Haklay, M., Khoo, S. M., Morison, J., Murphy, M. H., O'Brolchain, N., Schafer, B., & Shankar, K. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. Big Data & Society, 4(2), 1–21. https://doi.org/10.1177/2053951717726554

[8] Damaraji, G. M., Permanasari, A. E., & Hidayah, I. (2020). A review of expert system for identification of various risks in pregnancy. 2020 3rd International Conference on Information and Communications Technology (ICOIACT), 99–104. IEEE. https://doi.org/10.1109/ICOIACT49848.2020.9184703

[9] Dhombres, F., Bonnard, J., Bailly, K., Maurice, P., Papageorghiou, A. T., & Jouannic, J.-M. (2022). Contributions of artificial intelligence reported in obstetrics and gynecology journals: Systematic review. Journal of Medical Internet Research, 24(4), e35465. https://doi.org/10.2196/35465

[10] Diakopoulos, N. (2016). Accountability in algorithmic decision making. Communications of the ACM, 59(2), 56–62. https://doi.org/10.1145/2844110

[11] Drukker, L., Noble, J., & Papageorghiou, A. (2020). Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. Ultrasound in Obstetrics & Gynecology, 56(4), 498–505. https://doi.org/10.1002/uog.22052

[12] Du, Y., Rafferty, A. R., McAuliffe, F. M., Wei, L., & Mooney, C. (2022). An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-21524-9

[13] European Commission. (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence (AI Act). COM/2021/206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

[14] European Commission High-Level Expert Group on AI. (2019). Ethics guidelines for trustworthy AI. European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[15] Fernández, A. D. R., Fernández, D. R., & Sánchez, M. T. P. (2019). A decision support system for predicting the treatment of ectopic pregnancies. International Journal of Medical Informatics, 129, 198–204. https://doi.org/10.1016/j.ijmedinf.2019.06.007

[16] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). An ethical framework for a good AI society. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

[17] Kondapalli, K. K., & Gunupudi, C. (2018). A hybrid zero-trust-driven cloud architecture for securing distributed electronic health records in AI-enabled healthcare ecosystems. International Journal of Advanced Research in Engineering and Technology (IJARET), 9(2), 116–131.

[18] Liu, X., Rivera, S. C., Moher, D., Calvert, M. J., & Denniston, A. K. (2020). Reporting guidelines for clinical trial reports for interventions involving AI: The CONSORT-AI extension. BMJ, 370, m3176. https://doi.org/10.1136/bmj.m3176

[19] Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations. Journal of Informetrics, 12(4), 1160–1177. https://doi.org/10.1016/j.joi.2018.10.003

[20] Meijer, A., & Thaens, M. (2021). Algorithmization of bureaucratic decision-making. Government Information Quarterly, 38(4), 101618. https://doi.org/10.1016/j.giq.2021.101618

[21] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1–21. https://doi.org/10.1177/2053951716679679

[22] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

[23] OECD. (2019). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449. Organisation for Economic Co-operation and Development. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[24] Rajotte, J.-F., Bergen, R., Buckeridge, D. L., El Emam, K., Ng, R., & Strome, E. (2022). Synthetic data as an enabler for machine learning applications in medicine. iScience, 25(11), 103979. https://doi.org/10.1016/j.isci.2022.103979

[25] Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. npj Digital Medicine, 3(1). https://doi.org/10.1038/s41591-020-0352-1

[26] UNESCO. (2021). Recommendation on the ethics of artificial intelligence. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000380455

[27] Usmanova, G., Lalchandani, K., Srivastava, A., Joshi, C. S., Bhatt, D. C., Bairagi, A. K., Jain, Y., Afzal, M., Dhoundiyal, R., & Benawri, J. (2021). The role of digital clinical decision support tool in improving quality of intrapartum and postpartum care. BMC Pregnancy and Childbirth, 21(1). https://doi.org/10.1186/s12884-021-03935-5

[28] Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. Big Data & Society, 4(2), 1–17. https://doi.org/10.1177/2053951717743530

[29] Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harvard Journal of Law & Technology, 31(2), 841–887. https://doi.org/10.2139/ssrn.3063289

[30] White House Office of Science and Technology Policy. (2022). Blueprint for an AI Bill of Rights. Executive Office of the President. https://www.whitehouse.gov/ostp/ai-bill-of-rights/