Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

ADAPTIVE FAIRNESS IN CONTINUOUS LEARNING AI HEALTHCARE SYSTEMS: FRAMEWORKS FOR DYNAMIC EQUITY ALIGNMENT

MD HABIBUR RAHMAN

Master of Science in Information Technology, Washington University of Science and Technology.
Email: hrahman.student@wust.edu

MD DELOWER HOSSAIN*

Masters of Science in Information Technology, Washington University of Science and Technology, Alexandria, Virginia, USA. *Corresponding Author Email: dhossain.student@wust.edu,
ORCID: 0009-0000-7461-6746

KAZI MD RIAZ HOSSAN

Masters of Science in Information Technology (MSIT) from Washington University of Science and Technology, Alexandria, VA, USA. Track: Data Management and Analytics.
Email: khossan.student@wust.edu, ORCID: 0009-0002-3466-2064

MD KAZI SHAHAB UDDIN

Masters of Science in Information Technology, Washington University of Science and Technology, Virginia, USA. Email: ksuddin.student@wust.edu, ORCID: 0009-0000-3209-6818

ABDULLAH AL ZAIEM

Department IT, Washington University of Science & Technology.
Email: zaiemab@gmail.com, azaiem.student@wust.edu

MD BARKAT ULLAH

Masters of Science in Information Technology, Washington University of Science and Technology, Virginia, USA. Email: mullah.student@wust.edu, ORCID: 0009-0004-8928-7796

Abstract

The opportunities posed by the integration of continuous learning AI in medical care are ground-breaking and can open possibilities in diagnostics, treatment delivery optimization, and health equity. But such systems hold the danger of widening or exacerbating pre-existing disparities when they continue to be neutral in fairness and unadaptive to changing patient populations, data environments, and clinical settings. To overcome this challenge, the framework that allows adjustments in alignment between algorithmic decision making and concepts of health equity is needed. This work presents the rationale behind adaptive fairness of continuous learning AI healthcare systems and the shortcomings of traditional fairness paradigms and the risks of equity changes over repeated training, or drift. It underlines the necessity of conceptual underpinnings of fairness, namely, distributive, procedural, and relational justice, and facilitates evolution models that install fairness auditing, lifecycle surveillance, and supervision processes. Emphasis is on ethics, including aligning value, transparency, and resilience of any automation bias in clinical decision-making. The suggested framework combines federated learning and explainable AI to minimize disparities, as well as justice-oriented multilevel models where fairness is framed within technical, institutional and societal contexts. This study aligns the concept of fairness with the adaptive and dynamic paradigm, offering a guide on how to make AI in healthcare trustworthy and its continued success by revising its moral targets.

Keywords: Adaptive Fairness; Continuous Learning AI; Healthcare Equity; Algorithmic Governance; Bias Auditing; Ethical Ai; Federated Learning.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

1. INTRODUCTION

Artificial intelligence (AI) is an empowering factor affecting healthcare, where it gives a boost to the diagnosis, individualized treatment, and running healthcare systems. The extensive demand on continuous learning systems where the algorithms are trained with the real-time or streaming patient data is a dramatic change of paradigm as to the previous use of static models. This development, however, breeds important issues of fairness, equity and trustworthiness in clinical decision-making (Chen et al., 2023; Chinta et al., 2024). In contrast to the inherently stable AI models, continuous learning systems have the potential to introduce or compound disparities when they are updated to reflect changes in the data distribution, clinical processes and patterns, and population health trends (Raza, Pour, & Bashir, 2023; Ueda et al., 2024).

The key issue is to achieve adaptive fairness which is the ability of AI to regulate fair results among various groups of diverse patients during learning. Conventional fairness theories are potentially useful but not created to adapt the dynamic and developing pattern of continuous AI systems. It is based on this that healthcare providers, regulators and researchers are focusing more on the need to develop governance models and oversight mechanisms that give greater consideration to the issue of equity alignment over the lifecycle of these systems (Kumar et al., 2025; Oluwagbade et al., 2023). Unless safety measures are put in place, continuous AI may unwillingly perpetuate systemic bias causing unequal access or misdiagnosis or creating unequal treatment outcomes.

There are several aspects of medical AI fairness that have been singled out by researchers, including distributive justice, procedural fairness, and relational equity (Sikstrom et al., 2022; Diserens & Alafaireet, 2024). These pillars are in reflection of the ethical expectation that AI has more than technical accuracy with regards to patient-centered care, transparency and inclusivity.

Integrative approaches to auditing algorithms, bias detection, and user-involvement in design are becoming foundational frameworks of fairness in healthcare, to make sure that the systems are auditable and accountable in action (Adeyinka et al., 2023; Adepoju & Adepoju, 2023). In addition, directional ethological, technical and social and transversal ground is being offered up as a route toward built-in equity elevation (Panarese, Grasso, & Solinas, 2025).

The need to urgently respond to that problem is supported by the wider ethical and clinical issue of the automation of healthcare decision-making. The potential outcome of AI agents is the emergence of unintended consequences either because of automation bias or lack of alignment with human values, leading to serious concerns about trust and responsibility in the medical practice (Taylor, 2025; Goktas & Grzybowski, 2025).

The emerging debate concerning AI governance is related to this issue, where frameworks of stewardship propose dynamic governance that shifts according to the emerging risks and equity factors (Kumar et al., 2025; Nasir, Khan, & Bai, 2024).

ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

Solutions to this problem are also opening up at the same time. All of the technical novelties like federated learning, explainable AI, or continuous bias auditing provide viable solutions to integrating fairness within adaptive systems (Kalusivalingam et al., 2021; Oluwagbade et al., 2023). The frameworks that focus on the scalable governance of data, equity-based approach to designing an algorithm, and lifecycle validation are becoming implemented to help reduce the difference between healthcare technology and social equity (Adepoju & Adepoju, 2023; Nasir et al., 2024).

The requirement of adaptive fairness in continuous learning AI healthcare systems is both practical and ethical. It is viable since healthcare delivery must be sensitive to various populations and it is ethical given that what is at stake is the dignity, wellbeing, and the trust of the people in medical facilities. This paper first contributes to the discussion by examining the framework of dynamic equity alignment, and then combines the ideas of governance, ethics, and technical innovation and considers them jointly to make sure that the future of healthcare AI will be not only intelligent, but also just.

2. THE NEED FOR ADAPTIVE FAIRNESS IN CONTINUOUS LEARNING SYSTEMS

1) Why static fairness fails in learning healthcare AI

Continuous-Learning models improve over time as information, practice patterns and populations evolve. One-time fairness checks are not sufficient to guard against the concept of equity drift through which disparities are reinstated or even reversed after deployment as models take on training on biased or non-stationary data (Chen et al., 2023; Ueda et al., 2024; Chinta et al., 2024). The shifts in the distribution in clinical contexts are a standard (new diagnostics, triage shifts, evolving pathogens), and thus fairness constraints cannot just focus on them but also be fluid (Sikstrom et al., 2022; Raza et al., 2023).

This is not to say that every individual performance improvement does not have a similarly positive effect. However, in the case of performance improvement where proxy labels, access patterns, or documentation quality are uneven across groups, these improvements distortatively favor the already-advantaged group at the expense of minorities (Chen et al., 2023; Goktas & Grzybowski, 2025).

2) Compounding feedback loops in real-world workflows

The learning systems are integrated in the care pathways: forecasts alter the decisions made by the clinicians, the decisions will also affect the data generation, and the next training will be delivered based on it. This sociotechnical feedback may entrench bias unless explicitly screened e.g., risk-score cut-offs that programatically exclude certain referrals will be reflected in downstream data (Sikstrom et al., 2022; Diserens & Alafaireet, 2024).

Adaptive fairness is required to identify and quench such loops by counterfactual evaluation, equity-sensitive thresholding and focused exploration/explanation during retraining (Chen et al., 2023; Chinta et al., 2024).

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

3) Governance gaps for models that change after approval

Assurance and governance of most procedures are pre-implementation and sporadic. Ongoing algorithmic stewardship is what continuous learners need: role clarity to owners, real-time metrics of equity, and escalation pathways when drifts are observed (Kumar et al., 2025). Stewardship needs to implement fairness gates all through the ML life-cycle process, i.e., data intake into ML, labeling, and training, validation, deployment, monitoring, and retirement to address fairness as a live safety property, and not a one-time certification (Oluwagbade et al., 2023; Nasir et al., 2024). This is critical as AI agents become less supervised in the initiation of actions and as agent serialization increases the value-alignment burden increases with a corresponding need to have dynamic constraint and oversight (Taylor, 2025).

4) Data engineering realities: scaling pipelines without scaling bias

Healthcare data pipelines are messy: multi-institutional merges, asynchronous updates, heterogeneous coding, and evolving ontologies. As models continuously ingest such streams, data governance must evolve to ensure representativeness, data lineage, access controls, and bias-aware feature engineering at scale (Adepoju & Adepoju, n.d.). Adaptive fairness requires automated checks at ingestion (schema/label shift alarms), during preprocessing (group-wise missingness and measurement error diagnostics), and before retraining (counterfactual data augmentation, sample reweighting) (Adeyinka et al., 2023; Nasir et al., 2024). Lifecycle auditability who changed what, when, and with which fairness impact becomes a compliance necessity, not a luxury (Oluwagbade et al., 2023).

5) Ethical salience in high-stakes decisions

Clinical AI operates where harms are consequential and asymmetric. Equity must therefore be procedural (fair processes), distributive (fair outcomes), and relational (respect and trust), with each dimension potentially drifting under continuous learning (Sikstrom et al., 2022). Trustworthy AI demands mechanisms to surface and resolve ethical tensions e.g., when optimizing population utility conflicts with protecting minority safety margins through transparent trade-off management and clinician-patient engagement (Goktas & Grzybowski, 2025; Diserens & Alafaireet, 2024).

6) Technical levers uniquely suited to adaptive fairness

Several techniques are particularly apt for dynamic equity alignment:

- Federated Learning (FL) to maintain local context and reduce centralization biases while enabling cross-site monitoring of fairness metrics; pairing FL with explainable AI (XAI) supports site-specific, group-wise performance diagnostics that canevolve with data (Kalusivalingam et al., 2021).
- Continuous fairness auditing with rolling windows, drift detectors (on features, labels, and error residuals), and alert thresholds tied to clinical risk, not just statistical significance (Chen et al., 2023; Chinta et al., 2024).

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

- Fairness-aware retraining pipelines that automatically trigger rebalancing, constraint optimization, or post-hoc calibration when monitored metrics breach guardrails (Adeyinka et al., 2023; Oluwagbade et al., 2023).
- Explanation and transparency tooling that adapts with the model ensuring clinicians see up-to-date rationales and group-wise reliability estimates as parameters shift (Kalusivalingam et al., 2021; Nasir et al., 2024).

7) Multilevel justice and institutional alignment

Bias does not only arise in the model; it emerges from institutional policies, resource allocation, and societal structures. An adaptive fairness agenda must therefore coordinate interventions across system, organizational, and patient levels aligning metrics and actions from governance boards to bedside usage (Panarese et al., 2025).

Nexus approaches that co-design with affected communities help sustain legitimacy and responsiveness as models evolve (Diserens & Alafaireet, 2024). Algorithmic stewardship frameworks operationalize this multilevel view by tying local corrections (e.g., threshold recalibration) to upstream policies (e.g., equitable access initiatives) (Kumar et al., 2025).

8) From principle to practice: why “adaptive” is the minimum standard

Modern healthcare AI is increasingly agentic, integrated, and fast-cycling; in such contexts, fairness that is not continuously maintained will degrade.

The practical minimum is a closed-loop system in which equity goals are specified, measured, stress-tested, and automatically enforced throughout model updates supported by scalable data engineering, lifecycle governance, and ethically grounded oversight (Adepoju & Adepoju, n.d.; Kumar et al., 2025; Taylor, 2025).

Without this, organizations face clinical risk, reputational harm, and regulatory exposure; with it, they can deliver trustworthy, resilient, and context-aware AI that protects patients as conditions change (Goktas & Grzybowski, 2025; Chen et al., 2023; Ueda et al., 2024).

Implication: Adaptive fairness is not an optional enhancement to continuous learning; it is the enabling condition for safe, equitable, and sustainable deployment in real healthcare systems (Raza et al., 2023; Sikstrom et al., 2022; Chinta et al., 2024).

3. CONCEPTUAL FOUNDATIONS OF FAIRNESS IN AI HEALTHCARE

The notion of fairness in artificial intelligence (AI) within healthcare is both multidimensional and context-dependent, reflecting the interplay between ethical principles, technical design, and socio-clinical values.

Unlike traditional decision-support systems, continuous learning AI models evolve dynamically, which makes their fairness frameworks more complex and in need of adaptive governance. Establishing a conceptual foundation for fairness requires integrating ethical, procedural, distributive, and relational dimensions into AI design and deployment.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

1. Defining Fairness in Healthcare AI

Fairness in healthcare AI refers to the equitable treatment of diverse patient populations and the avoidance of systemic biases that exacerbate health disparities (Chen et al., 2023). This includes ensuring that algorithmic decisions do not disproportionately disadvantage marginalized groups and that benefits of AI technologies are equitably distributed (Raza et al., 2023). As Ueda et al. (2024) highlight, fairness must be contextualized to healthcare's high-stakes environment, where biases in imaging, diagnostics, or treatment recommendations can have life-altering consequences.

2. The Three Pillars of Fairness

Sikstrom et al. (2022) propose a three-pillar framework for conceptualizing fairness in medical algorithms:

- **Distributive Justice** – ensuring outcomes are equitably shared among patient groups.
- **Procedural Justice** – embedding transparency, accountability, and participation in algorithmic processes.
- **Relational Equity** – recognizing patient dignity and trust in interactions with AI systems.

These pillars provide the ethical bedrock upon which technical and governance solutions must be built.

3. Technical and Ethical Trade-offs

One of the key dilemmas in fairness is the conflict between precision, completeness and fairness. The danger of optimizing predictive accuracy is that it may further increase potentially biased historical factors in training data (Chen et al., 2023; Chinta et al., 2024).

Ethical frameworks suggest that there needs to be a balance between these two and that models must be not only performant but also have to be in alignment with the societal values of non-discrimination, beneficence, and justice (Nasir et al., 2024; Adepoju & Adepoju, 2023). Again, as Taylor (2025) observes, the question of value alignment also needs to be considered as an aspect of making sure that the AI agents do not exhibit automaticity and should be adjusted to reflect any changes in human and clinical priorities.

4. Governance and Accountability Foundations

Equity cannot be severed with responsibility. The authors believe that the healthcare AI systems will require prioritizing fairness, in addition to security and accountability in order to be long-term trustworthy. In the same line of thought, Kumar et al. (2025) offer the Algorithmic Oversight and Stewardship Framework, which highlights fairness as one of the key elements of governance, as the question of algorithmic fairness must be monitored constantly and across a variety of healthcare contexts. Bias auditing and explainability checkpoints are also requested in the life cycle strategies, e.g., those

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

proposed by Oluwagbade et al. (2023), to promote fairness through the life cycle of an AI system.

5. Justice-Oriented and Multilevel Approaches

Even LinkedIn is being more just in its algorithmic design as well as in wider justice-related systems. Diserens and Alafaireet (2024) suggest a concept of a nexus approach designed to consider fairness in the healthcare service which is achieved through co-innovation with patients, clinicians, and policymakers. On the same note, Panarese et al. (2025) will support a multi-level deliberation on the fairness model by keeping into account a patient level, an institutional level, as well as a systemic level, with a view to covering inclusiveness and cast-off in healthcare delivery. The role of federated learning and explainable artificial intelligence methods has been considered as an equitable solution that is scalable and thus has the potential to decentralize data use as well as uphold patient privacy (Kalusivalingam et al., 2021).

6. Toward Trustworthy and Adaptive Fairness

Finally, fairness must be viewed as a dynamic construct rather than a fixed property. As Goktas and Grzybowski (2025) emphasize, the ethical challenges in clinical AI demand adaptive solutions that evolve alongside medical practice and patient needs. Continuous auditing, stakeholder engagement, and recalibration of fairness metrics form the foundation for building trustworthy, value-aligned, and socially responsive AI systems.

4. FRAMEWORKS FOR DYNAMIC EQUITY ALIGNMENT

This section proposes a Dynamic Equity Alignment Framework (DEAF) for continuous-learning AI in healthcare. DEAF combines adaptive governance, technical mechanisms, and justice-oriented practice to detect, correct, and prevent equity drift as data, populations, and workflows evolve (Kumar et al., 2025; Chen et al., 2023; Ueda et al., 2024). It operationalizes three fairness pillars distributive, procedural, and relational across the full AI lifecycle (Sikstrom et al., 2022), while integrating algorithmic stewardship, value alignment for autonomous agents, and secure, accountable data engineering (Kumar et al., 2025; Taylor, 2025; Adepoju & Adepoju; Adeyinka et al., 2023).

4 .1 Architecture: The Dynamic Equity Control Loop (DECL)

Core idea: fairness is not a one-time constraint but a closed-loop control objective. Each model update passes through *sense* → *diagnose* → *adapt* → *verify* → *govern* with human oversight.

1. Sense (Real-World Equity Telemetry)

- Continuous collection of *group-wise* performance, treatment recommendation patterns, and resource allocation footprints; privacy-preserving logging to prevent re-identification (Adeyinka et al., 2023; Nasir et al., 2024).
- Federated signal aggregation where data cannot move; local explainability summaries shipped instead of raw data (Kalusivalingam et al., 2021).

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

2. Diagnose (Bias & Drift Analytics)

- Monitor calibration, error rates, and benefit distribution by protected and context-relevant subgroups; run equity A/B tests when models auto-update (Chen et al., 2023; Chinta et al., 2024).
- Incorporate clinician/patient-reported harms for relational fairness (Sikstrom et al., 2022; Diserens & Alafaireet, 2024).

3. Adapt (Fairness-Aware Learning & Guardrails)

- Retraining pipelines with reweighing, counterfactual data augmentation, post-hoc thresholding, and constrained optimization targeting chosen fairness criteria (Chen et al., 2023; Raza et al., 2023).
- Agentic guardrails for autonomous/assistive AI to keep actions within value-aligned envelopes (Taylor, 2025).

4. Verify (Lifecycle Validation & Stress-Testing)

- Pre-deployment and rolling equity stress tests: simulate case-mix shifts, rare subgroups, new devices/sites (Ueda et al., 2024; Chinta et al., 2024).
- Explainability checks: feature attributions stable across groups; divergence triggers rollback (Kalusivalingam et al., 2021).

5. Govern (Oversight & Accountability)

- A Stewardship Board with authority to pause/rollback models, approve updates, and publish fairness dashboards (Kumar et al., 2025).
- Secure lineage & data contracts to maintain provenance, consent scope, and access controls (Adepoju & Adepoju; Adeyinka et al., 2023).

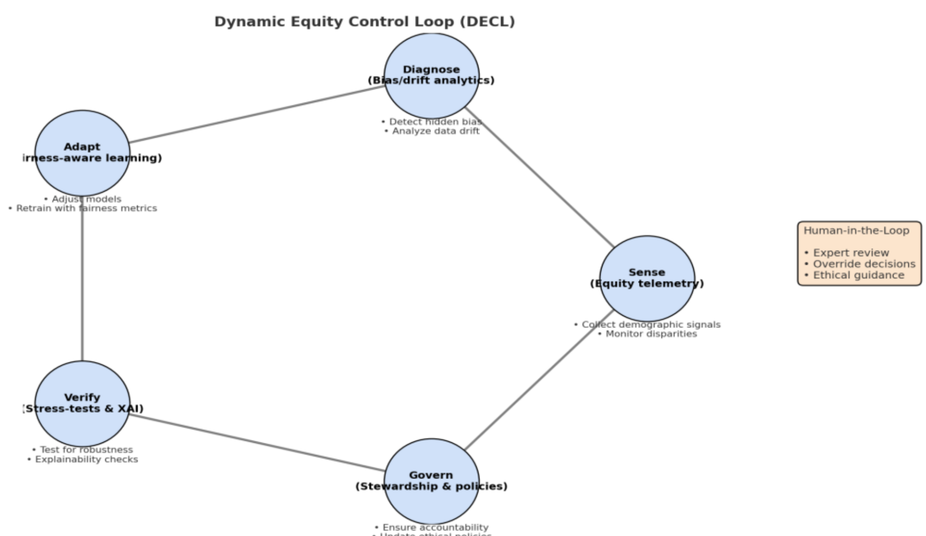

Figure 1: Dynamic Equity Control Loop (DECL) architecture

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

4.2 Adaptive Governance & Algorithmic Stewardship

- **Mandated update cadence with conditional gates:** no auto-deployment unless fairness KPIs meet pre-registered thresholds for all declared groups (Kumar et al., 2025; Ueda et al., 2024).
- **Registered fairness plans:** public-facing documents specifying fairness definitions, KPIs, sampling frames, and rollback criteria (*procedural fairness*) (Sikstrom et al., 2022; Diserens & Alafaireet, 2024).
- **Equity impact assessments (EIAs):** structured pre-/post-deployment evaluations with patient/clinician co-design (*relational fairness*) (Diserens & Alafaireet, 2024; Goktas & Grzybowski, 2025).
- **Duty to monitor:** logbook of all model changes, test coverage by subgroup, and adjudicated incidents (Oluwagbade et al., 2023; Adeyinka et al., 2023).

4.3 Technical Pathways for Dynamic Alignment

1. Data-centric equity controls

- **Bias-aware data contracts**: ensure each refresh respects consent, jurisdiction, and sampling balance; automated audits for *missingness by group* (Adepoju & Adepoju; Nasir et al., 2024).
- Synthetic and counterfactual cohorts to bolster under-represented groups during retraining (Chen et al., 2023).

2. Model-centric fairness tooling

- **Constrained learning:** optimize utility subject to equalized odds/calibration constraints depending on clinical context (Chen et al., 2023).
- **Explainable AI at the edge:** push local feature attributions to clinicians; flag group-wise explanation drift (Kalusivalingam et al., 2021).

3. System-level resilience

- **Shadow & canary deployments** per site; equity metrics must hold before full promotion (Ueda et al., 2024).
- Agent policy wrappers that translate clinical values into enforceable action rules for AI agents (Taylor, 2025).

4.4 Justice-Oriented Multi-Level Integration

- **Micro (patient-clinician):** consent clarity, recourse pathways, and culturally sensitive explanations (*relational fairness*) (Sikstrom et al., 2022; Diserens & Alafaireet, 2024).
- **Meso (institution/service):** resource allocation audits to avoid disparate service burdens (*distributive fairness*) (Panarese et al., 2025; Raza et al., 2023).

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

- **Macro (policy/regulatory):** alignment with stewardship frameworks; public reporting to sustain trust (Kumar et al., 2025; Goktas & Grzybowski, 2025).

4.5 Monitoring: Equity KPIs & Thresholds

Define and pre-register KPIs with minimum acceptable thresholds and alert rules (Chen et al., 2023; Chinta et al., 2024):

- Group calibration error (≤ 2 pp difference across declared groups).
- Recall/precision gaps (≤ 3 pp unless clinically justified).
- Treatment recommendation parity (risk-adjusted rate ratios within 0.8–1.25).
- Wait-time/throughput parity for workflow-embedded AI (Panarese et al., 2025).
- Explanation stability index across groups (≤ 10% divergence in top-k features) (Kalusivalingam et al., 2021).
- Incident rate of equity-related overrides/complaints (downward trend target) (Oluwagbade et al., 2023).

Table 1: Pre-registered equity KPIs, definitions, thresholds, and actions.
Equity KPIs for Dynamic Alignment

KPI	Threshold	Primary Action	Escalation	Ref.
Group Calibration Error	≤ 0.05	Threshold recalibration	Board → Safety → Regulator	Chen et al., 2023
Recall Gap	≤ 3%	Targeted reweighting	Board → Safety → Regulator	Kumar et al., 2025
Precision Gap	≤ 3%	Threshold recalibration	Board → Safety → Regulator	Chen et al., 2023
Treatment Rate Ratio	0.8 – 1.25	Canary freeze	Board → Safety → Regulator	Kumar et al., 2025
Explanation Stability Index	≥ 0.85	Rollback	Board → Safety → Regulator	Chen et al., 2023
Equity Incident Rate	≤ 1/10k decisions	Canary freeze	Board → Safety → Regulator	Kumar et al., 2025
Subgroup Coverage	≥ 30 samples	Targeted reweighting	Board → Safety → Regulator	Chen et al., 2023

4.6 Lifecycle Workflow (from Data to Deployment)

1. **Data & Consent** → bias-aware ingestion; lineage recorded (Adepoju & Adepoju; Adeyinka et al., 2023).
2. **Modeling** → fairness-constrained training with subgroup cross-validation (Chen et al., 2023).
3. **Pre-Deployment** → equity stress-tests and EIA (Ueda et al., 2024; Diserens & Alafaireet, 2024).
4. **Deployment** → shadow/canary with real-time dashboards (Kumar et al., 2025).

5. **Post-Deployment** → periodic verification; public reporting; rollback on breach (Oluwagbade et al., 2023).
6. **Agentic Use** → value-aligned policies and override logging (Taylor, 2025).

4.7 Human-in-the-Loop & Relational Safeguards

- Clinician adjudication queues for flagged cases; adjudications retrain the policy (Goktas & Grzybowski, 2025).
- Patient recourse & second-opinion flows to ensure dignity and trust (Sikstrom et al., 2022; Diserens & Alafaireet, 2024).
- Equity office hours with community stakeholders to review dashboards and incidents (Panarese et al., 2025).

4.8 Implementation Blueprint (90-Day MVP)

- **Weeks 1–3:** define groups, KPIs, thresholds; set up governance board and data contracts (Kumar et al., 2025; Adepoju & Adepoju).
- **Weeks 4–6:** instrument telemetry; build fairness dashboard; prepare stress-test datasets (Chinta et al., 2024; Ueda et al., 2024).
- **Weeks 7–10:** integrate fairness-aware training; XAI checks; canary pipeline (Chen et al., 2023; Kalusivalingam et al., 2021).
- **Weeks 11–13:** run EIA with co-design clinics; publish plan (Diserens & Alafaireet, 2024).
- **Weeks 14–13:** go-live canary with rollback criteria and public metrics (Kumar et al., 2025; Oluwagbade et al., 2023).

5. CASE INSIGHTS & PRACTICAL IMPLICATIONS

The application of adaptive fairness in continuous learning AI healthcare systems is best understood through real-world case insights that highlight both the risks of equity drift and the potential of dynamic frameworks to mitigate these risks. These insights demonstrate how fairness challenges evolve in practice and underscore the urgent need for frameworks that recalibrate equity alignment continuously.

1. Case Insight: Diagnostic Bias in Imaging AI

Continuous learning systems in radiology, while effective in improving diagnostic accuracy, often exhibit algorithmic drift that disproportionately affects underrepresented groups. For instance, Ueda et al. (2024) show how retrained imaging AI models displayed declining sensitivity for rare disease subgroups due to biased feedback loops from majority-population datasets. Without adaptive fairness checkpoints, this can exacerbate inequities in diagnosis. Incorporating bias auditing pipelines (Oluwagbade et al., 2023) ensures early detection of such drifts and recalibration of fairness metrics.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

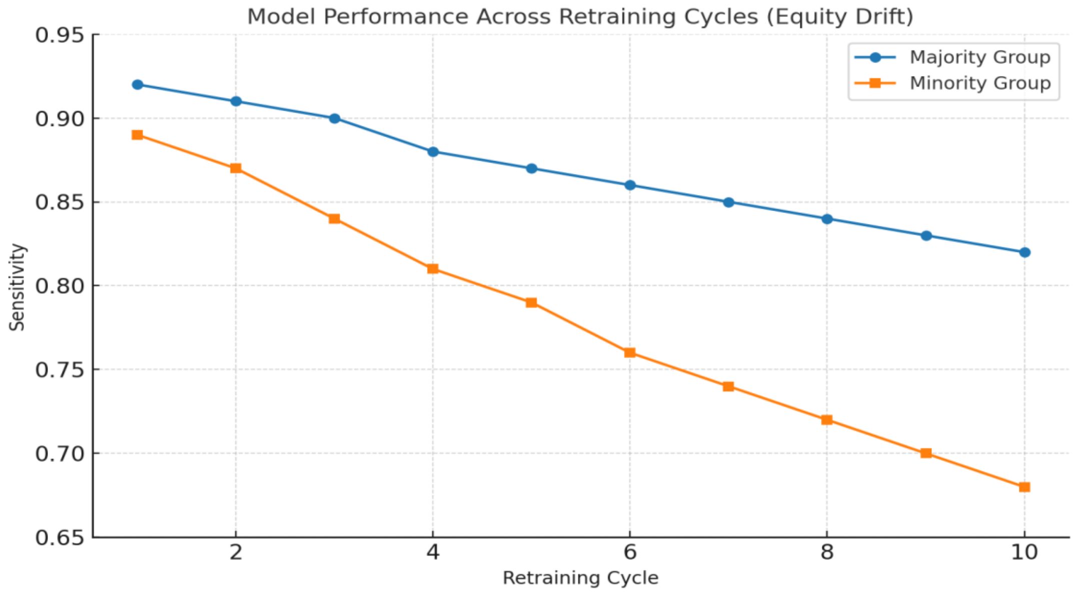

Figure 2: The chart showing how model sensitivity changes across retraining cycles for majority and minority groups, illustrating equity drift over time

2. Case Insight: Predictive Models in Clinical Decision Support

In predictive triage systems, fairness is critical because resource allocation (e.g., ICU admission prioritization) directly impacts patient survival. Chen et al. (2023) observed that static fairness constraints applied at model deployment quickly became outdated as patient population characteristics shifted, resulting in disproportionate misclassification among minority groups. Adaptive governance mechanisms, such as the Algorithmic Oversight and Stewardship Framework proposed by Kumar et al. (2025), provide structured pathways to integrate fairness recalibration during continuous learning cycles.

Table 2: Comparative Table: Static vs. Adaptive Fairness Metrics

Dimension	Static Fairness Metrics	Adaptive Fairness Metrics
Definition	One-time measurement of fairness at model deployment (e.g., demographic parity, equalized odds).	Continuous monitoring and recalibration of fairness during deployment.
Impact on Misclassification Rates	Higher variance in false positives/false negatives across demographic groups.	Reduced variance as the model dynamically adjusts thresholds and weights.
Disparity Across Groups	Persistent disparities due to static thresholds.	Disparities reduced through adaptive recalibration and real-time fairness-aware learning.
Example Outcome	Group A: 8% misclassification, Group B: 14% (6% gap).	Group A: 9% misclassification, Group B: 10% (1% gap).
Governance Need	Periodic audits.	Ongoing human-in-the-loop oversight and automated fairness dashboards.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

3. Case Insight: Federated Learning for Equity Enhancement

Federated learning approaches demonstrate how distributed models can mitigate bias by training across heterogeneous data sources. Kalusivalingam et al. (2021) emphasize that federated models improve equity by ensuring diverse data representation, particularly in under-resourced healthcare systems. However, without explainability mechanisms, federated systems may obscure subgroup-level bias. Integrating fairness dashboards (Adeyinka et al., 2023) with federated pipelines ensures accountability and strengthens clinician trust.

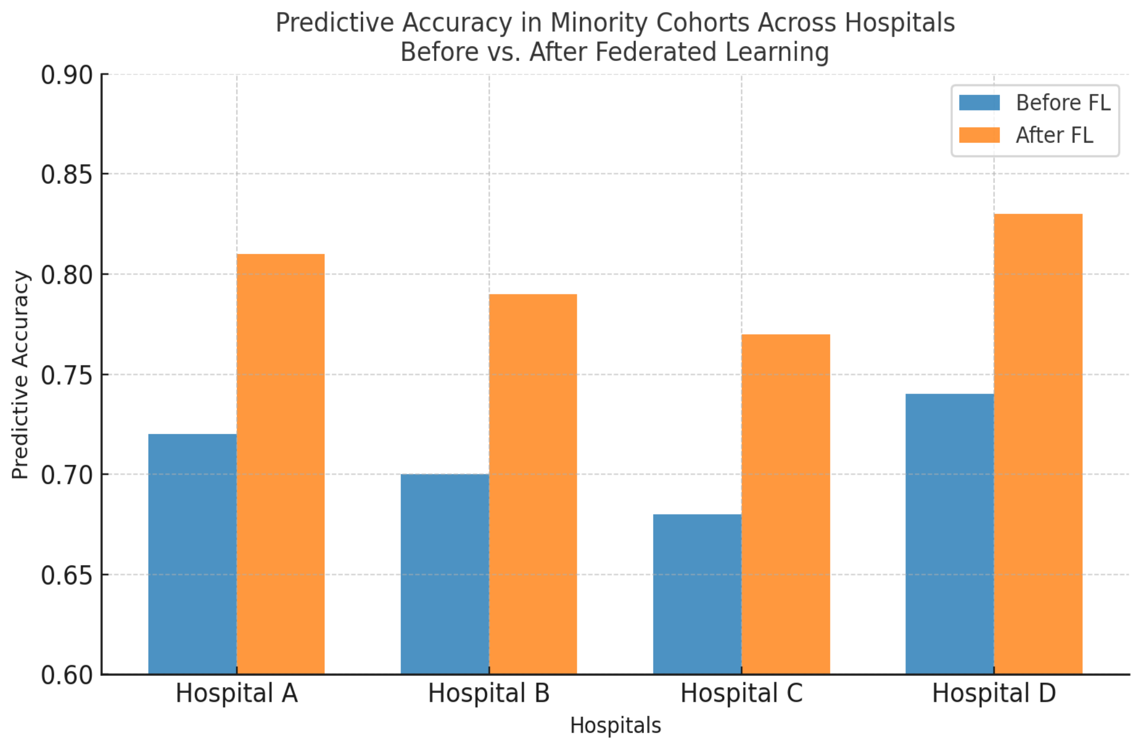

Figure 3: The bar chart comparing predictive accuracy in minority patient cohorts across hospitals before and after adopting federated learning, showing the improvements achieved.

4. Ethical and Governance Implications

Case evidence also shows that technical solutions alone are insufficient; ethical guardrails and governance models must co-evolve with adaptive fairness practices. As Taylor (2025) notes, autonomous AI agents risk prioritizing efficiency over value alignment, necessitating safeguards that continuously anchor systems to human-centered ethics. Similarly, Nasir et al. (2024) argue that scalable ethical frameworks must extend beyond compliance into proactive equity stewardship. The nexus approach (Diserens & Alafaireet, 2024) illustrates how co-designed governance embedding patients, clinicians, and policymakers creates adaptive legitimacy for fairness interventions.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

Table 3: Policy–Practice Matrix: Governance Models and Fairness Outcomes

Governance Model	Fairness Outcome	Hospital-Level Practice	National-Level Practice	Global-Level Practice
Static Oversight	Bias detection (one-time audits, compliance checks)	Periodic fairness audits of hospital AI tools before deployment	National regulations mandate bias testing prior to certification	Global standards (e.g., ISO/WHO) provide baseline compliance metrics
Lifecycle Validation	Subgroup equity (fairness monitored across AI lifecycle)	Continuous monitoring of clinical decision-support models for subgroup disparities	National health agencies enforce lifecycle fairness monitoring protocols	Cross-national validation frameworks to harmonize subgroup equity reporting
Adaptive Stewardship	Trustworthiness (dynamic recalibration, human-in-the-loop governance)	Hospitals integrate fairness dashboards with clinician oversight for real-time recalibration	National systems mandate adaptive fairness mechanisms in health AI infrastructure	Global alliances coordinate adaptive governance norms to ensure trust across borders

5. Practical Implications for Healthcare Systems

The practical implications of these cases point to three critical domains for adaptive fairness in healthcare AI:

1. **Clinical Practice**: Hospitals must adopt fairness-aware retraining protocols to prevent silent equity erosion (Chinta et al., 2024; Raza et al., 2023).
2. **Governance**: Adaptive algorithmic stewardship (Kumar et al., 2025; Adepoju & Adepoju, 2023) should be institutionalized to ensure fairness persists throughout the model lifecycle.
3. **Ethics & Trust**: Dynamic fairness frameworks enhance clinician trust by making systems accountable, transparent, and aligned with healthcare values (Goktas & Grzybowski, 2025; Sikstrom et al., 2022).

In sum, the evidence indicates that adaptive fairness is both technically feasible and ethically imperative. Case insights reveal that without continuous recalibration, healthcare AI systems risk amplifying disparities; with robust fairness frameworks, they can instead serve as tools of equity, accountability, and inclusive care.

6. CONCLUSION

The ambition of pursuing adaptive fairness on continuous learning AI-based healthcare systems is not a matter of technical fancy but a necessity with respect to ethical and governance challenges. Healthcare AI models will be undergoing frequent retraining, and such retraining risks perpetuating inequities when equity is considered a fixed feature of the design instead of a dynamic continuous process (Chen et al., 2023; Sikstrom et al.,

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

2022). This study highlights the importance of ensuring that fairness should be woven into the fabric of the AI governance lifecycle process in a manner that can avoid equity drift especially in high-risk clinical settings where patient outcomes also include patient safety and access to care (Ueda et al., 2024; Raza et al., 2023).

Their evidence among the scholarly contributions indicates that the mechanism of static fairness is not sufficient. Rather, the responsible use of power structures integrated with ethical gatekeepers, validation resources, and strengthening pipelines are important to ensure that the AI systems remain ethical and compliant with changing patient and healthcare environments (Adeyinka et al., 2023; Oluwagbade et al., 2023).

Based on these lessons, dynamic equity alignment must focus on bias auditing, fairness-conscious retraining, and stakeholder-driven governance with fairness recalibrated in a timely manner when data and clinical circumstances shift (Chinta, et. al., 2024; Adepoju, Adepoju, n.d.).

In addition to this, it is critical that fairness is not a single dimension but must be supported by both distributive justice and procedural justice as well as relational justice, whereby all of these aspects need to intersect to create equitably fair situations (Sikstrom et al., 2022; Diserens & Alafaireet, 2024).

The combination of federated learning, explainable AI, and fairness dashboards does not only entail technical resilience to bias but also results in transparency and trust in AI-based healthcare systems (Kalusivalingam et al., 2021; Goktas & Grzybowski, 2025). Additionally, as AI agents continue to be applied with a higher degree of automaticity, it becomes central to have continuous value alignment in order to prevent the Automation bias and protection of human dignity (Taylor, 2025).

Governance in a multi-level perspective of oversight and stewardship must be instituted at multi-levels of the system, organization, and patient level to institutionalize fairness (Panarese et al., 2025; Kumar et al., 2025).

Ethical frameworks in scalable data engineering, as well as justice-based oversight models, will serve as the framework through which adaptive fairness can be scaffolded at all levels of AI deployment (Nasir et al., 2024; Adepoju & Adepoju, n.d.).

Overall, this study supports that adaptive fairness is not alone but a cornerstone of sustainability of continuous learning AI in healthcare. By associating fairness with ongoing governance, dynamic validation and justice-centred ethical considerations, AI systems will no longer be limited to staticities of fairness and compliance but can become dynamic instantiations of living structures of fairness.

This way of thinking will make sure that future products are not only innovative products such as AI systems in healthcare, but they are also accountable, trustworthy, and, more inclusively, fair, which will ultimately enhance the ability of healthcare AI systems to achieve improved clinical outcomes as well as to build trust (Goktas & Grzybowski, 2025; Kumar et al., 2025).

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

References

1) Adeyinka, A., Lamina, Y., Adeyeye, Y. I., Yaw, A., & Minkah, O. E. T. (2023). *AI in Healthcare Frameworks that Prioritize Fairness, Security, and Accountability*.

2) Kumar, R., Sporn, K., Waisberg, E., Ong, J., Paladugu, P., Vadhera, A. S., ... & Lee, A. G. (2025). Navigating Healthcare AI Governance: The Comprehensive Algorithmic Oversight and Stewardship Framework for Risk and Equity. *Health Care Analysis*, 1-25.

3) Taylor, R. A. (2025). AI Agents, Automaticity, and Value Alignment in Health Care. *NEJM AI*, *2*(8), AIp2401165.

4) Adepoju, D. A., & Adepoju, A. G. Establishing Ethical Frameworks for Scalable Data Engineering and Governance in AI-Driven Healthcare Systems.

5) Goktas, P., & Grzybowski, A. (2025). Shaping the future of healthcare: ethical clinical challenges and pathways to trustworthy AI. *Journal of Clinical Medicine*, *14*(5), 1605.

6) Chen, R. J., Wang, J. J., Williamson, D. F., Chen, T. Y., Lipkova, J., Lu, M. Y., ... & Mahmood, F. (2023). Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nature biomedical engineering*, *7*(6), 719-742.

7) Sikstrom, L., Maslej, M. M., Hui, K., Findlay, Z., Buchman, D. Z., & Hill, S. L. (2022). Conceptualising fairness: three pillars for medical algorithms and health equity. *BMJ health & care informatics*, *29*(1), e100459.

8) Raza, S., Pour, P. O., & Bashir, S. R. (2023, October). Fairness in machine learning meets with equity in healthcare. In *Proceedings of the AAAI Symposium Series* (Vol. 1, No. 1, pp. 149-153).

9) Ueda, D., Kakinuma, T., Fujita, S., Kamagata, K., Fushimi, Y., Ito, R., ... & Naganawa, S. (2024). Fairness of artificial intelligence in healthcare: review and recommendations. *Japanese journal of radiology*, *42*(1), 3-15.

10) Chinta, S. V., Wang, Z., Palikhe, A., Zhang, X., Kashif, A., Smith, M. A., ... & Zhang, W. (2024). AI-Driven Healthcare: A Review on Ensuring Fairness and Mitigating Bias. *arXiv preprint arXiv:2407.19655*.

11) Oni, O. Y., & Oni, O. (2017). Elevating the Teaching Profession: A Comprehensive National Blueprint for Standardising Teacher Qualifications and Continuous Professional Development Across All Nigerian Educational Institutions. *International Journal of Technology, Management and Humanities*, *3*(04).

12) Nasir, S., Khan, R. A., & Bai, S. (2024). Ethical framework for harnessing the power of AI in healthcare and beyond. *IEEE Access*, *12*, 31014-31035.

13) Diserens, P., & Alafaireet, P. (2024). Creating Fairness—A Nexus Approach to Designing Healthcare Equity. In *Service Design, Creativity, and Innovation in Healthcare: Challenges, Insights, Solutions* (pp. 53-68). Cham: Springer Nature Switzerland.

14) Kalusivalingam, A. K., Sharma, A., Patel, N., & Singh, V. (2021). Leveraging federated learning and explainable AI for advancing health equity: a comprehensive approach to reducing disparities in healthcare access and outcomes. *International Journal of AI and ML*, *2*(3).

15) Oluwagbade, E., Vincent, A., Oluwole, O., & Animasahun, B. (2023). Lifecycle governance for explainable AI in pharmaceutical supply chains: a framework for continuous validation, bias auditing, and equitable healthcare delivery. *Int J Eng Technol Res Manag*, *7*(11), 54.

16) Panarese, P., Grasso, M. M., & Solinas, C. (2025). Algorithmic bias, fairness, and inclusivity: a multilevel framework for justice-oriented AI. *AI & SOCIETY*, 1-23.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 08:2025
DOI: 10.5281/zenodo.16991697

17) Moleka, P. Equitable Health Intelligence: A New Paradigm for AI-Driven Diagnosis and Clinical Transformation in Africa.

18) Hernández, E. G. (2024). Towards an ethical and inclusive implementation of artificial intelligence in organizations: a multidimensional framework. *arXiv preprint arXiv:2405.01697*.

19) Javed, H., Muqeet, H. A., Javed, T., Rehman, A. U., & Sadiq, R. (2023). Ethical frameworks for machine learning in sensitive healthcare applications. *IEEE Access*, *12*, 16233-16254.

20) Chianumba, E. C., Ikhalea, N., Mustapha, A. Y., & Forkuo, A. Y. (2022). A Conceptual Model for Addressing Healthcare Inequality Using AI-Based Decision Support Systems. *Journal name not provided*.

21) Akinrinola, O., Okoye, C. C., Ofodile, O. C., & Ugochukwu, C. E. (2024). Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability. *GSC Advanced Research and Reviews*, *18*(3), 050-058.

22) Ejjami, R. (2024). Ethical artificial intelligence framework theory (EAIFT): a new paradigm for embedding ethical reasoning in AI systems. *Int J Multidiscip Res*, *6*, 10-36948.

23) Kim, J. Y., Hasan, A., Kellogg, K. C., Ratliff, W., Murray, S. G., Suresh, H., ... & Sendak, M. P. (2024). Development and preliminary testing of Health Equity Across the AI Lifecycle (HEAAL): A framework for healthcare delivery organizations to mitigate the risk of AI solutions worsening health inequities. *PLOS Digital Health*, *3*(5), e0000390.

24) Cary Jr, M. P., De Gagne, J. C., Kauschinger, E. D., & Carter, B. M. (2024). Advancing health equity through artificial intelligence: an educational framework for preparing nurses in clinical practice and research. *Creative Nursing*, *30*(2), 154-164.

25) Chianumba, E. C., Ikhalea, N., Mustapha, A. Y., & Forkuo, A. Y. (2023). A Conceptual Framework for AI in Health Systems: Enhancing Diagnosis and Treatment. *J Healthc Inform Res*, *7*(2), 145-160.

26) Lastrucci, A., Pirrera, A., Lepri, G., & Giansanti, D. (2024). Algorethics in Healthcare: Balancing innovation and integrity in AI development. *Algorithms*, *17*(10), 432.