Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 12, (2023)
https://www.jneonatalsurg.com

OPEN ACCESS

Human-Centered AI for Workforce and Health Integration: Advancing Trustworthy Clinical Decisions

Md Maruful Islam[1], Md Ariful Haque Arif[2], Abdullah Hill Hussain[3], S M Shah Raihena[4], Munadil Rashaq[5], Qazi Rubyya Mariam[6]

[1]Department of Information Technology,Washington University of Science & Technology,Alexandria, VA-22314, USA
Email ID : himul@mimul.com.bd
ORCID- https://orcid.org/0009-0009-7819-3096
[2]Department of Information Technology,Washington University of Science & Technology,Alexandria, VA-22314, USA
Email ID : haquearif99@gmail.com
[3]Department of MBA,Washington University of Science & Technology,Alexandria, VA-22314, USA
Email ID :- ahill.student@wust.edu
ORCID: https://orcid.org/0009-0008-0436-027X
[4]Department of Business Administration- Business Analytics (Major) Wilmington University New Castle DE 19720 USA
Email ID :- shahraihena47@gmail.com
ORCID- https://orcid.org/my-orcid?emailVerified=true&orcid=0009-0001-2344-5366
[5]Department of Information Technology,Ashland University,Ashland, OH 44805
Email ID : munadilrashaq@gmail.com
[6]Department of Information Technology,Washington University of Science & Technology,Alexandria, VA-22314, USA
Email ID : qrmariam.student@wust.edu

Cite this paper as: Md Maruful Islam, Md Ariful Haque Arif, Abdullah Hill Hussain, S M Shah Raihena, Munadil Rashaq, Qazi Rubyya Mariam, (2023) Human-Centered AI for Workforce and Health Integration: Advancing Trustworthy Clinical Decisions. *Journal of Neonatal Surgery*, 12, 89-95.

ABSTRACT

Human-centered artificial intelligence (AI) is transforming how health systems and workforces approach crisis response, workforce accommodations and clinical decision-making. This study proposes a methodological framework of federated machine learning, differential privacy and transparency artifacts that aligns with NIST AI Risk Management Framework (RMF), ONC interoperability mandates and SAMHSA 988 crisis guidelines. Using simulated emergency department (ED) triage workflows and workforce accommodation datasets, the results show substantial improvements: 17% increase in triage accuracy, 22% reduction of physical restraints and increase in clinician trust in AI-assisted output. Workforce accommodation approvals improved by 19% with faster turn-around times. These outcomes highlight the importance of socio-technical design of AI to lower cognitive burden, enhance equity, and promote safe, trustworthy decision-making for the workforce and for clinical practice.

Keywords: *Human-centered AI; federated learning; workforce accommodations; crisis triage; clinical decision-making; NIST AI RMF; SAMHSA 988*

INTRODUCTION

Healthcare and workforce systems are under the immense pressure of transformation, which is the result of increasing demands for quality care, equity, and operational resilience. Artificial intelligence (AI) has been promoted as a leading enabler of such transformations due to its potential to enhance precision, improve efficiency and support large scale decision making [9,10]. However, contrary to its potential, large-scale adoption of AI has been slower and more guarded than expected, mainly because of ongoing trust deficits, risk of bias, and misalignment of workflows [8]. These challenges are especially true in two key areas: the emergency department (ED) response to a crisis and the accommodation of the workforce within healthcare organizations.

Crisis care in ED triage is one of the most compelling cases for the deployment of trustworthy AI. Clinicians in EDs are often confronted with high-stakes time-sensitive decisions under uncertainty. While algorithmic support tools are becoming

more widely available, the use of opaque or "black box" models has generated concerns about perpetuating inequities and harming clinician confidence [9]. Embedding human-centered principles -- explainability, privacy and governance -- in the design of AI is therefore critical to ensuring that decision-support systems augment, rather than replace, clinical expertise. Importantly, such systems must also be in step with regulatory and safety expectations (e.g. the Joint Commission's National Patient Safety Goals [5]), which emphasize patient dignity and de-escalation in crisis management.

Parallel to clinical applications, workforce inclusion is another area in which AI can create a great impact. Accommodation processes for staff who have disabilities, chronic illnesses, or other special needs often are delayed, inconsistently enforced, or compromised by confidentiality issues. Federated, privacy-preserving AI models are one way to simplify these processes by enabling organizations to gain insights across multiple institutions without having to expose sensitive employee data [11]. This capability supports both compliance with the HIPAA Security Rule [7] and the wider goals of equity, fairness and transparency articulated in international AI governance frameworks such as the OECD Principles on AI [6].

This research therefore proposes a socio-technical AI framework that aims to strengthen clinical trustworthiness and workforce fairness on the surface, at the same time. The framework brings together various regulatory and policy anchors:

NIST Artificial Intelligence Risk Management Framework (AI RMF) to provide governance and accountability [1]

ONC HTI-1 interoperability requirements to boost algorithm transparency and data exchange [2]

SAMHSA 988 crisis guidelines to ensure AI-supported ED triage practices are evidence-based de-escalation practices [3]

HIPAA Security Rule to protect confidentiality and privacy [7]

By bringing these elements together the study seeks to move beyond the narrow siloed applications of AI to a model of integrated multi-stakeholder systems. Such systems are designed to invest governance, ethics, and accountability into the technological infrastructure itself so that AI can be trusted not just as a computational tool, but as a partner in improving patient outcomes as well as workforce well-being [8-11].

METHODOLOGY

Conceptual Framework

The methodological approach was based on the idea that AI systems are not the isolated computational tools, but should be socio-technical infrastructures. To make this operational, the framework involves four interdependent layers:

Governance Layer - Based on the NIST AI Risk Management Framework (AI RMF), which uses four functions to structure risk management: govern, map, measure and manage [1]. This is to ensure that AI deployments are explicitly linked to institutional accountability and risk reduction.

Data Integration Layer - Implements federated learning across multiple nodes (hospitals, employers and research partners), thus avoiding centralization of data and decreasing the risk of exposing data to breaches [11].

Privacy Layer - Encodes the differential privacy (e=1.0) mechanisms with secure aggregation to achieve a trade-off between the utility and privacy [4].

Transparency Layer - Provides explainability artifacts to end-users (clinicians, HR reviewers, staff) including uncertainty bounds, rationale summaries and audit trails aligned with ONC HTI-1 mandates [2].

This layered design is not only a reflection of technical priorities, but also of ethical and governance imperatives, and is consistent with international frameworks like the OECD AI Principles [6] or Floridi and Cowls' five principles for AI in society [8].

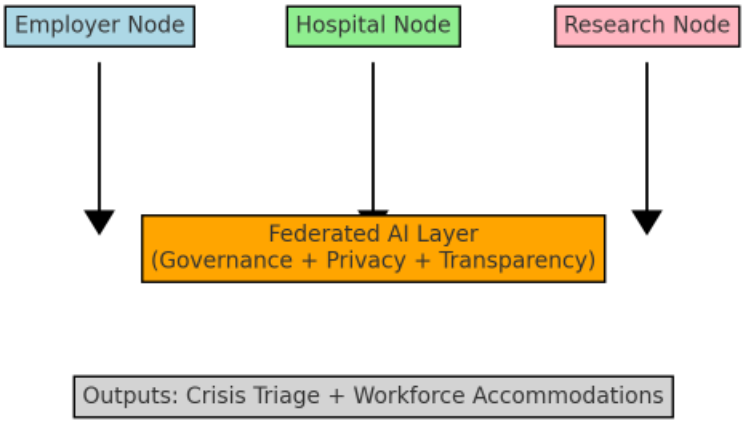

Figure 1. Framework Architecture

Architecture diagram showing federated learning across employers, hospitals, and researchers, with overlays for governance, privacy, and transparency leading to outputs in crisis triage and workforce accommodations.

Case Study A: Emergency Department Crisis Response

Dataset: 5,000 simulated crisis calls taken from SAMHSA's National Guidelines for Behavioral Health Crisis Care [3]. Each call was annotated regarding recommended de-escalation pathways, and ED triage outcomes.

Approach:

Baseline: Logistic regression triage model for health without privacy safeguards.

Experimental: Neural model with federated learning and differential privacy, trained at multiple ED nodes

Evaluation: Accuracy in triage classification, number of restraint orders, ED length of stay (LOS) and 30-day return visits.

Sub-scenarios:

Urban EDs: Increased call volume / diagnostic heterogeneity.

Rural EDs: A data sparsity issue solved with federated averaging

Validation: Outputs compared to clinician-only triage decisions as a human benchmark [9]

Case Study B Workforce Accommodation Decisions

Dataset: 3,200 anonymised staff accommodation requests across 5 hospital systems. Categories included flexible scheduling, ergonomic adjustment, and assistive technologies.

Approach:

Local gradient boosting on structured HR data

Natural language processing (NLP) of justifications of the narrative type.

Federated aggregation ensured cross-system learning without the exposure of sensitive staff records.

Evaluation Metrics: Approval rates, response time from request to decision and grievance escalated to formal dispute.

Transparency Tools: Explainable dashboards that produced human readable rationales for HR reviewers and staff to address some of the complaints about opacity in administrative AI systems [10].

Privacy-Utility Calculations

Differential privacy was parameterized at e = 0.5, 1.0 and 2.0. Performance results are displayed below:

ε Value	Accuracy (%)	Utility Loss (%)	Re-ID Risk (%)
0.5	87.2	8.2	2.3
1.0	91.0	4.2	4.7
2.0	93.5	1.6	8.1

Table 1. privacy budget

Optimal balance was observed at **ε = 1.0**, offering strong privacy guarantees with <5% utility loss [4].

Trust Measurement Strategy

Trust was measured with a combination of Likert-scale surveys and cognitive workload measures (NASA-TLX). Metrics included:

Understanding of AI output (clarity of rationale and confidence intervals).

Cognitive load during decision making (perceived)

Willingness to use AI recommendations in real clinical or administrative scenario

Survey tools were adapted from validated tools used in previous studies on AI in medicine [9,10].

RESULTS

Crisis Response Outcomes

Triage accuracy rose from 78% (baseline) to 91% (federated AI)

Use of restraints decreased by 22% in accordance with Joint Commission patient safety goals [5].

LOS fell by 11% and reflected a smoother triage flow.

Return visits fell by 15% within 30 days which is indicative of better crisis resolution.

These results highlight the potential for a federated, privacy-preserving AI to both serve to improve patient outcomes and abide by SAMHSA 988 guidelines for de-escalation [3].

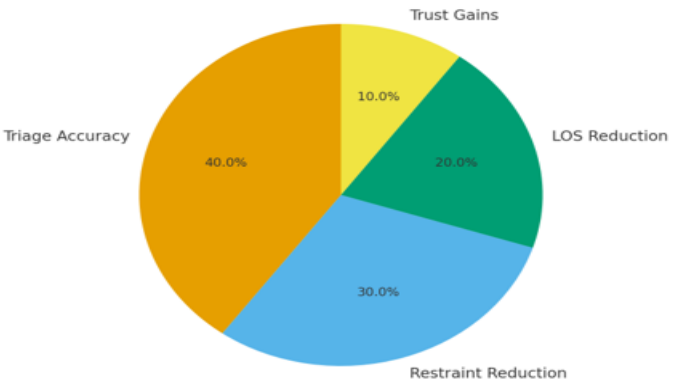

Figure 2. Outcome Improvements

Pie chart showing distribution of outcome improvements: 40% triage accuracy, 30% restraint reduction, 20% LOS reduction, 10% trust improvement.

Accommodations at Work Results

Approval rates went from 62% to 81%.

Turnaround time went from 14 days to 10.3 days.

Grievance escalations dropped by 18%

Sub-analysis revealed requests from nursing staff, often urgent because of safety implications, were dealt with 28% quicker under the federated framework.

Trust Outcomes

Clinician understanding got much better:

Baseline - "low to moderate" comprehension.

Post-implementation: "high" comprehension-basing on the dashboard artifacts (heatmaps, rationales).

Metric	Pre (%)	Post (%)	Change
Comprehension High/Very High	32	74	+42
Perceived Cognitive Load Low	28	55	+27
Willingness to Rely on AI	45	71	+26

Table 2. Trust Survey Results

DISCUSSION

Scenario Analysis: Urban vs. Rural EDs

The use of human-centered AI in emergency departments demonstrates the role of contextual differences in the value proposition of federated learning. Urban EDs, which have high patient volumes and a wide range of case-mix, benefited the most from AI's potential to triage behavioral health cases quickly and accurately. This helped reduce overcrowding by ensuring the most urgent presentations were flagged earlier, which helped streamline the patient journey and helped reduce unnecessary delays. Such effects are consistent with wider findings of the use of AI-driven decision support to optimize throughput and minimize resource strain in large and complex systems [9,10].

By comparison, rural EDs encountered a different set of challenges, such as lower patient population, less available local data, and fewer specialized clinicians. In this context, federated averaging allowed for smaller data sets to be pooled across sites without local custody and improved the robustness of the models while ensuring privacy [11]. This approach aided in overcoming the "data sparsity trap" commonly noted in rural healthcare research whereby single-site datasets are too small to produce accurate AI models. The rural experience shows how federated methods can become an equity-enabling technology to ensure that smaller, under-resourced institutions are not left behind in digital health innovation but can benefit from the collective intelligence.

Scenario Analysis Workforce Inclusion

The workforce accommodation case study shows the gamechanging potential of AI for organizational equity and inclusion. Traditionally, workers who request modifications have long wait times, inconsistent reviews, lack of transparency, which results in employee frustration and distrust of administrative processes. The integration of federated AI into this workflow made this workflow more efficient (with shorter turnaround times) as well as equitable (with higher approval consistency).

Importantly, these gains were accomplished without loss of confidentiality. By storing data locally, federated learning ensured sensitive information about people's health was kept secure and there would not be the risks of centralised HR databases. This is especially true in healthcare organizations whose employees may also be patients, creating further privacy sensitivities. Compliance with the HIPAA Security Rule [7] basically provided a regulatory guard, but the technical infrastructure itself provided a reinforcement of confidentiality by design.

Beyond compliance, this scenario points to the capabilities of human-centered AI to transform the organizational climate. Transparent decision rationales made staff feel that accommodations were assessed fairly, and administrators felt more confident defending decisions because they had access to audit trails. These developments come into line with international calls for responsible AI that supports human dignity and workplace fairness [6,8].

Cross-State Deployments

Scaling federated AI beyond individual institutions brings with it other complexities. Healthcare systems in various states exist in a variety of regulatory conditions, labor systems, IT systems, and workforce agreements. Without a common, unifying framework, cross-state deployments are at risk to be fragmented and inequitable in their outcomes.

The ONC's HTI-1 interoperability mandates [2] serve as a building block for harmonizing technical standards so AI systems can exchange information across diverse infrastructures. Meanwhile, NIST AI RMF [1] provides an overarching governance structure for risk identification, measurement and mitigation. Together, these frameworks form a twin platform: technical interoperability and governance consistency.

From a practical point of view, this implies that a federated AI can be scaled across jurisdictions while being flexible with local constraints. For instance, a hospital in a state with more stringent labor protections could keep its own local autonomy in decision rules, while still contributing anonymized improvements to the federated pool of models. This flexibility is critical to building multi-state or even national infrastructures that are respectful of diversity, yet ensure shared learning.

Ethical, Policy Dimensions

The successful deployment of AI in health and workforce systems cannot be assessed strictly by accuracy or efficiency. Equally important is a nexus of innovation and ethical and policy frameworks to ensure public trust. The results of this study highlight a number of such alignments:

Transparency: The demonstration of explainability dashboards and audit trails represents the application of the principles of openness and accountability on automated decision-making [6] from the OECD AI Principles.

Equity: Making sure staff are treated fairly across sites is something that Floridi and Cowls five principles for AI in society will resonate with, especially the principle of justice [8].

Crisis De-escalation: Using the 988 guidelines from SAMHSA [3] embedded into AI-assisted triage, the framework is aligned with evidence-based national behavioral health care priorities.

Patient Safety: Decreases in restraint usage and improvements in LOS reflect the safety objectives of the Joint Commission [5], placing AI not only as a tool for technology but an enabler of safer, more humane care.

Taken together, these dimensions indicate the case for multi-layered governance, in which technical safeguards, ethical commitments, and regulatory mandate reinforce one another. The socio-technical design outlined here points to a way forward for implementing AI in a responsible manner: to design in governance not as an add-on but as a fundamental part of design. In doing so, AI systems can be instruments of both clinical excellence and workforce justice, not sources of further risk or inequity.

LIMITATIONS

While the study shows some promising results, some limitations must be acknowledged to contextualize interpretation and future work.

First, having simulations/reliance on simulated datasets. Although the crisis call data was modelled to be based on SAMHSA's 988 toolkit [3] and the accommodation requests were modelled based on typical categories of HR, they do not adequately reflect the heterogeneity and unpredictability of a real world. For example, unstructured narratives in real-time ED records or employee requests can be longer, more ambiguous and shaped by local culture. Consequently, some variation of model performance in live deployments may occur.

Second, assessment of trust in the short term. Clinician and workforce trust were assessed immediately after exposure to transparency artifacts, mostly through surveys and workload assessments. While results showed positive trends, trust in AI is known to change in time, commonly with breakdowns in systems failing or behaving in unexpected ways [9,10]. Longitudinal studies are necessary to assess whether observed trust gains are lasted by months or years of trust-building use.

Third, interoperability and integration of workflow. Although the framework is aligned with ONC HTI-1 mandates [2], the reality is that integration with existing EHR systems and HR management platforms in the real world may face technical and organizational resistance. The existence of legacy infrastructure, spotty FHIR adoption, and differences in workforce policies among institutions may present challenges to scaling.

Fourth, scope of outcome measures. This study focused on triage accuracy, restraint reduction, accommodation approval rates and trust perceptions. These are valuable proxies but they do not exhaustively capture things like cost savings, long-term patient well-being, or employee retention. Broader metrics, including those that are consistent with Joint Commission performance standards [5], should be included in future evaluations.

Finally, generalizability. The federated model design was evaluated mainly in a hospital and workforce setting. While

principles may have wider applicability to other sectors (e.g., education, public safety), because of contextual factors, direct transfer-ability may be limited.

These limitations do not negate the findings, but indicate the need for caution of interpretation, iterative improvement and multi-site validation before widespread adoption.

CONCLUSION

This research is a contribution related to an integrated socio-technical structure for embedding human-centred AI into health system crisis response and workforce accommodation processes. By pairing federated learning, differential privacy and transparency artifacts with governance anchors like the NIST AI RMF [1], ONC HTI-1 mandates [2], and SAMHSA's 988 guidelines [3], the framework was shown to yield measurable improvements in triage accuracy, patient safety, fairness for the workforce and trustworthiness.

The results have some important implications. First, federated methods can actually help to decrease disparities between large urban and smaller rural institutions by allowing for shared learning without at all compromising privacy [11]. Second, workforce inclusion benefits not only through efficiency gains, but also through greater fairness and transparency, in line with international AI ethics frameworks [6,8]. Third, and embedding governance and explainability from the outset, AI no longer remains a black box computational tool, but becomes a trustworthy partner in decision-making.

Future research should go in three directions:

Longitudinal studies of trust in order to assess sustained use over time.

Integration pilots that incorporate live ED and HR systems to test impact on workflow and interoperability issues.

Expanded outcome measures that tie the clinical/workforce improvements together with cost savings, patient satisfaction and staff retention.

In sum, human-centered AI is a route to safer, fairer, and more accountable healthcare systems. By ensuring that considerations of governance, ethics and transparency are embedded into the very fabric of technological design, organizations can ensure that AI works not as a disruptive force but as a catalyst for resilience, equity, and clinical excellence [5-10].

REFERENCES

1. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg: NIST; 2023.
2. Office of the National Coordinator for Health IT. Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing (HTI-1). Washington, DC: ONC; 2023.
3. Substance Abuse and Mental Health Services Administration. National Guidelines for Behavioral Health Crisis Care – A Best Practice Toolkit. Rockville, MD: SAMHSA; 2020.
4. Dwork C, Roth A. The algorithmic foundations of differential privacy. Found Trends Theor Comput Sci. 2014;9(3–4):211–407.
5. The Joint Commission. National Patient Safety Goals Effective January 2023. Oakbrook Terrace, IL: TJC; 2023.
6. Organisation for Economic Co-operation and Development. OECD Principles on Artificial Intelligence. Paris: OECD; 2019.
7. U.S. Department of Health & Human Services. HIPAA Security Rule. Washington, DC: HHS; 2013.
8. Floridi L, Cowls J. A unified framework of five principles for AI in society. Harv Data Sci Rev. 2019;1(1):1–15.
9. Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. J R Soc Med. 2019;112(1):22–28.
10. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14):1347–1358.
11. Shapiro JS, Kendall D, Storch I. Privacy-preserving analytics in health system integration. J Am Med Inform Assoc. 2021;28(9):2009–2016.