ETHICAL GUIDELINES
FOR
HUMANITARIAN AI SYSTEMS

Principles for Ethical Design, Development, and Deployment
in Health Supply Chain Optimization

Version 3.0

November 2025

International Foundation for Recovery and Development (IFRAD)

EXECUTIVE SUMMARY

Artificial Intelligence is reshaping how humanitarian health supply chains anticipate, procure, and deliver medicines in Uganda and across East Africa. This transformation introduces both technological promise and deep ethical responsibility.

These Ethical Guidelines establish the framework for all design, development, and deployment of AI tools for supply chain optimization in humanitarian and fragile settings. Developed by IFRAD under the Elrha Humanitarian Innovation Fund, they govern the AI Framework for Health Supply Chain Optimization currently being implemented across ten facilities in Karamoja and Southwestern Uganda.
Grounded in baseline assessment data from August-September 2025 and validated through comprehensive stakeholder consultation in November 2025, these guidelines define how AI systems must protect human dignity, address equity gaps, and uphold the principle of do no harm. They align with Uganda's National Development Plan IV, the Sphere Humanitarian Charter, Uganda's Data Protection and Privacy Act (2019), and WHO Ethics and Governance of AI for Health (2021).

These guidelines emerge from operational reality rather than abstract principles. Every requirement reflects actual constraints documented in baseline assessment and stakeholder validation: 89% facility connectivity failure rates, storage capacity correlation with stockouts ($r = -0.695$), frontline worker burden concerns, and the prevalence of system fragmentation. The guidelines specify how ethical principles translate into algorithmic parameters, interface design decisions, and override protocols.

This is the first comprehensive ethical framework for humanitarian AI supply chain systems developed in Africa for African contexts. It provides practical guidance for governments, NGOs, software developers, donors, and researchers creating or funding AI systems that influence humanitarian supply chains.

PURPOSE AND SCOPE

These guidelines accompany the AI Framework for Health Supply Chain Optimization, forming its ethical foundation. They define minimum ethical conditions under which AI may be used to inform decisions affecting medicine availability, nutrition commodities, and lifesaving supplies.
Their purpose is not abstract moral commentary but practical direction on embedding humanitarian ethics within algorithms, databases, predictive analytics, and logistics dashboards. They apply to governments, non-governmental organizations, software developers, donors, and research institutions creating or funding AI systems for humanitarian supply chains.

The framework governs the three-tier forecasting system currently developed. Tier 1 rule-based forecasting operates offline at facility level. Tier 2 hierarchical statistical forecasting runs on district servers. Tier 3 machine learning forecasting operates at central level for district planning. Each tier has distinct ethical requirements based on operational context.

HUMANITARIAN ETHICAL MANDATE

Humanitarian technology must begin from humanity, not efficiency.
Every AI system deployed under humanitarian mandate must uphold the four humanitarian protection principles from the Sphere Handbook (2018):

- Avoid exposing people to further harm
- Provide access to impartial assistance
- Protect people from physical and psychological abuse
- Assist people to claim rights and recover dignity

For AI systems, these principles translate into design obligations: transparency in automated decisions, equity in data representation, and the right to human oversight. No algorithmic efficiency justifies harm, exclusion, or exploitation.

GUIDING PRINCIPLES FOR AI IN HUMANITARIAN CONTEXTS

Do No Harm

All AI activities must undergo pre-deployment ethical risk assessments evaluating potential harm to individuals, communities, or institutions. The system must never autonomously execute a recommendation without validated human approval at decision points appropriate to the tier and context.

Tier-specific safeguards:

- Tier 1 (facility-level): Recommendations are advisory only. Facility staff must manually approve all orders through weekly inventory review process.
- Tier 2 (district-level): Hierarchical statistical forecasts supplement local data but cannot override facility decisions.
- Tier 3 (central-level): Machine learning predictions inform district procurement planning but require human validation before redistribution decisions.

Equity and Inclusion

Models must be trained and tested on data reflecting facility types, refugee settlements, and low-connectivity regions. Baseline evidence demonstrates exclusion of rural or refugee facilities perpetuates inequity. Algorithmic design must counter bias and guarantee proportional benefits across all facility tiers.

Concrete equity mechanisms implemented:

- Storage capacity equity adjustment: Facilities with very inadequate storage receive order recommendations capped at 50% of calculated demand. Facilities with inadequate storage receive recommendations capped at 70%. This prevents recommendations exceeding physical infrastructure capacity (addresses $r = -0.695$ correlation between storage and stockouts).
- Refugee-serving facility prioritization: Facilities in refugee settlements receive 1.2x weighting in allocation algorithms. Baseline data shows Nakivale facilities experience longer stockout durations despite serving large populations.
- Nutrition program integration: System integrates with Nutrition Appointment Platform (WHO/WFP/UNICEF/UNHCR) to proactively adjust nutrition commodity forecasts based on enrollment changes in refugee and climate-affected populations. This addresses WFP-identified persistent stockouts of nutrition supplies.
- Inventory frequency recommendations: System automatically triggers weekly inventory count recommendations for facilities experiencing more than four stockouts annually. Baseline shows weekly counts correlate with reduced stockouts compared to monthly counting.
- Cold start equity protocols: New facilities without historical data use stratified baseline averages matched by facility type, region, and refugee-serving status. System never initializes rural facilities using urban facility averages.

Transparency and Explainability

Every AI recommendation must be intelligible to human operators. Developers must provide clear explanations of predictions, uncertainty margins, and data sources.

Tier-specific transparency requirements:

- Tier 1 displays calculation logic in plain language: "Based on 3-month average (800 tablets) + Malaria Season adjustment (+200) = 1,000 calculated. Capped at 700 due to Inadequate storage rating."
- Tier 2 shows model confidence intervals and flags data quality issues when historical records are incomplete.

- Tier 3 provides feature importance rankings to district officers showing which variables most influenced predictions (storage capacity, past stockout patterns, seasonal factors, delivery lead times).

Accountability and Oversight
Responsibility for outcomes rests with human actors and institutions, not algorithms. Oversight mechanisms include governance committees, audit trails, and incident-reporting protocols consistent with Elrha's Incident Prevention and Management Policy.

Foundational Principle: Local Edits Always Win
When conflicts arise between system recommendations and user adjustments, the facility worker's local edit is preserved. This 'local edits always win' policy reflects technical validation findings that health workers on the ground have contextual knowledge algorithms cannot capture. The system logs conflicts for district review but cannot retroactively override facility decisions without explicit facility consent. This design builds trust and acknowledges that frontline workers have ground truth about their facility's needs.

Human-in-the-loop requirements by context:
- Routine inventory replenishment: System executes automatically when facility staff manually approves during weekly inventory review.
- Emergency redistribution: Requires explicit district-level approval regardless of model confidence. No automated emergency orders.
- Outbreak response: Facility clinical officer must activate outbreak mode and adjust multipliers based on disease severity.
- Delivery delays: District supply officer reviews redistribution recommendations when facilities report road closures.

Override logging: All human overrides are logged with user ID, timestamp, original recommendation, final decision, and mandatory reason selection. High override rates flag systematic forecasting errors requiring model refinement, not user penalties. The system operates on a no-blame principle.

Privacy and Data Protection
Respect for privacy is both humanitarian duty and legal requirement. Data collection and processing must comply with national and international data protection laws, include informed consent procedures, and restrict access through multi-tier authentication.

Offline data protection**:** Baseline assessment shows 89% of facilities experience unreliable connectivity, with typical offline periods of 2-3 days. During offline operation, data stored locally on devices must be encrypted using AES-256 with Android Keystore System. Keys are generated within hardware-backed secure enclaves and never appear unencrypted in memory or file storage. Synchronization protocols verify data integrity before upload. Data is archived after 12 months active retention and purged after upload confirmation to district servers.

Fair Benefit and Knowledge Sharing
Outputs and insights generated from AI systems are public goods within humanitarian contexts for research, evaluation, and policy development. Technical specifications, algorithms, and implementation guides must be published openly to enable replication and adaptation.

Commercial implementation acknowledgment: While humanitarian implementations remain open and freely replicable, commercial adaptations for private sector supply chains may be developed by implementers to sustain operations. Commercial licenses do not restrict humanitarian use, adaptation, or knowledge dissemination for public good purposes.

Frontline Worker Protection

Stakeholder validation identified frontline worker burden as a critical ethical concern. Health workers, nurses, and district health officers are already overburdened and digital systems must reduce, not increase, their workload. This principle directly addresses the "do no harm" obligation as it applies to system users.

Worker protection mechanisms implemented:

- Administrative burden reduction: Auto-save functionality eliminates explicit save actions. Single-tap order adjustments replace multi-step approval workflows. System generates paper backup forms matching HMIS 105 registers, eliminating duplicate manual entry.
- Battery optimization: Automatic synchronization only occurs when device battery exceeds 20% AND device is not in active use. This prevents device depletion that would render workers unable to use phones for essential communication. Manual sync option bypasses battery checks for urgent situations.
- Cognitive load reduction: Single-column layouts for all mobile interfaces reduce visual complexity. Minimum 48dp touch targets accommodate use in field conditions. Adaptive sizing ensures usability on budget 5-inch devices (480x800px) common in Uganda.
- Training burden minimization: Embedded video guides and interactive tutorials eliminate need for external training sessions. Progressive interface complexity reveals advanced features only when users demonstrate basic proficiency.
- Device charging accommodation: System requirements acknowledge real-world constraints including lack of reliable electricity. Offline-first architecture ensures functionality without power-dependent connectivity. Low computational requirements enable use on older, battery-constrained devices.

DATA GOVERNANCE AND PROTECTION

Ethical management of data lies at the core of responsible AI. Data gathered from health facilities, refugee settlements, and national systems must be treated as sensitive humanitarian information.

Developers and implementing partners must:

- Collect only data necessary for model training and validation. No patient-level health data is collected.
- Obtain explicit informed consent for primary data collection involving human subjects (facility staff interviews, system testing).
- Use anonymization for all storage and transmission. Facility identifiers are hashed. Geographic coordinates are rounded to district level.
- Maintain data retention schedules aligned with seven-year record-keeping requirements in Elrha grant conditions. After seven years, all identifiable data is permanently deleted.
- Conduct annual data protection audits with results shared with Ministry of Health.
- Never transfer data to third parties without written authorization from IFRAD and Ministry of Health.

ALGORITHMIC INTEGRITY AND HUMAN OVERSIGHT

Algorithms influence life-critical decisions. Their ethical integrity is non-negotiable.

Documentation Requirements

Developers must document each model's design, inputs, limitations, and validation results. Documentation includes:

- Algorithm type and rationale for selection
- Feature engineering decisions and data preprocessing steps

- Known limitations and failure modes
- Validation methodology and performance metrics
- Override patterns and model refinement history

Offline Decision-Making Protocols
Human oversight operates differently under intermittent connectivity. Standard practices assuming continuous internet access fail in humanitarian contexts.

- Offline override mechanism: Facility staff can modify or reject any Tier 1 recommendation locally without internet connection. Overrides are logged with mandatory reason selection from dropdown menu (Outbreak, Delivery Delay, Budget Constraint, Forecasting Error, Clinical Judgment, Other). Logs sync automatically during next connection.
- Emergency decision authority: When facilities are offline during emergencies requiring urgent redistribution, district officers can approve emergency orders based on last-known facility data. System flags these decisions for post-hoc facility confirmation during next sync.
- Credential management: Offline credentials remain valid for 90 days. System issues warning at 75 days (15 days before expiry). Credential refresh occurs during each successful sync. This accommodates extended offline periods in remote facilities.
- Sync failure handling: If a facility remains offline beyond 7 days, district office receives automated alert. District supply officer contacts facility via phone to verify supply status and manually process critical orders if needed.

Algorithmic Audits
Quarterly algorithmic audits evaluate bias, error propagation, and fairness across regions and population groups. Audits assess:

- Forecast accuracy by facility type (refugee-serving vs. non-refugee, HC II vs. HC III vs. RRH)
- Stockout reduction patterns across regions
- Override rates and reasons by facility
- Emergency procurement frequency changes
- Storage capacity constraint effectiveness

Audit results must be transparent to donors and government partners. Kyambogo University Department of Data Science, Networks & Artificial Intelligence conducts independent technical validation.

RISK MANAGEMENT AND ACCOUNTABILITY

Risk management is an ethical process, not only technical safeguard.

Ethical Risk Register
Implementers must maintain an Ethical Risk Register identifying potential harms such as:

- Data leakage through compromised devices or unsecured transmission
- Biased forecasts disadvantaging specific facility types or regions
- Misuse of predictive analytics for political manipulation of supply allocation
- Over-reliance on automation reducing human judgment capacity
- System failure during critical supply decisions
- Increased frontline worker burden from poorly designed digital tools

Each identified risk must have assigned mitigation actions, responsible persons, and reporting intervals.

Incident Response
When harm occurs, implementers must report incidents within 24 hours to the Ministry of Health. The 24-hour clock starts from when the incident is discovered by any project personnel, not when it occurred.

Offline context adjustment: For facilities experiencing connectivity outages, the reporting clock starts when the facility next syncs and the incident is detected by system monitoring, or when facility staff report via phone, whichever occurs first.
Incident reports must include root cause analysis, affected facilities or populations, immediate mitigation steps, and corrective measures. Serious incidents (data breaches, systematic bias affecting multiple facilities, algorithmic errors causing stockouts) require public disclosure to affected communities.

Trade-offs and Ethical Tensions
Ethical oversight has costs. These guidelines acknowledge that ethics is not free and that principles sometimes conflict.

- Speed vs. oversight: Requiring human review for all predictions would delay urgent orders. The framework prioritizes timely delivery for routine replenishment while requiring oversight for emergency redistribution and outbreak response.
- Transparency vs. complexity: Full algorithmic transparency would overwhelm facility staff with technical details. The framework provides appropriate transparency for each tier: simple explanations for Tier 1, confidence intervals for Tier 2, feature importance for Tier 3.
- Equity vs. efficiency: Storage capacity caps reduce forecasting errors but may perpetuate infrastructure inequity. The framework addresses this by pairing algorithmic constraints with explicit infrastructure improvement advocacy to Ministry of Health through governance committee recommendations.
- Privacy vs. functionality: Strong encryption protects data but increases computational requirements on low-power devices. The framework uses AES-256 with Android Keystore for sensitive data while accepting lighter encryption for temporary local storage during offline operation.

Principle hierarchy when conflicts occur: Do No Harm takes precedence. If providing timely medicines requires bypassing standard oversight, the system allows facility staff offline override authority while logging decisions for post-hoc review.

IMPLEMENTATION AND COMPLIANCE MECHANISMS

Compliance is monitored through multi-layered review:

- Institutional oversight: Ministry of Health Ethics Committee reviews quarterly compliance reports covering override patterns, stockout trends, system performance, and incident logs.
- Independent review: Kyambogo University validates ethical conformity of algorithms through technical audits assessing bias, fairness, and transparency.
- Government supervision: Ministry of Health receives semi-annual ethical performance updates and approves any major system modifications.
- User engagement: Facility managers complete brief monthly feedback surveys during routine inventory cycles asking: (1) Did the system recommend anything that seemed wrong? (2) Did you override any recommendations this month? Survey responses inform model refinements.

All implementing partners must sign an Ethical Compliance Declaration before receiving or deploying the framework.

INSTITUTIONAL RESPONSIBILITIES AND GOVERNANCE

Ethical governance relies on shared accountability:

- Ministry of Health serves as custodian of these guidelines and maintains the Ethics and Data Governance Office.
- Kyambogo University provides technical validation and ongoing ethical training for AI developers.
- Ministry of Health acts as regulatory authority confirming consistency with national health data and AI policies.

- Implementing Partners and NGOs bear operational responsibility for adherence in field applications.

Forecasting Governance Committee

A Forecasting Governance Committee convenes quarterly to review compliance status, approve updates, and manage grievances.

Membership:

- Ministry of Health (Pharmacy Division)
- District Health Officers (rotating representatives from pilot districts)
- Facility users (nominated by peers from baseline facilities)
- Technical partner (IFRAD or successor organization)
- Academic validator (Kyambogo University Department of Data Science)

Responsibilities:

- Quarterly review of model performance metrics and stockout trends
- Analysis of override patterns and reasons to identify systematic issues
- Approval of algorithm changes and model refinements
- Recommendations for framework refinement and infrastructure improvements
- Advocacy for addressing identified inequities in storage capacity and connectivity

ALIGNMENT WITH NATIONAL, REGIONAL, AND GLOBAL FRAMEWORKS

These guidelines align with:

- Uganda's National Development Plan IV (NDP IV)
- Uganda's Data Protection and Privacy Act (2019)
- National eHealth Policy Framework (2013)
- African Union Continental AI Strategy (2024)
- WHO Ethics and Governance of Artificial Intelligence for Health (2021)
- OECD AI Principles (2019)
- Sphere Humanitarian Charter and Minimum Standards in Humanitarian Response (2018)

Alignment confirms that African-developed ethical guidance contributes to global standards rather than merely adopting them. These guidelines offer concrete implementation specifications that other frameworks often leave abstract.

CONCLUSION AND CONTINUING COMMITMENT

Ethical innovation is not constraint but foundation for sustainable technological progress. The AI Framework for Health Supply Chain Optimization demonstrates that advanced analytics can merge with humanitarian values.

These guidelines are living instruments. As new technologies and contexts emerge, they will evolve through dialogue among governments, researchers, developers, and communities. Version 3.0 incorporates lessons from baseline assessment, technical specification development, and comprehensive stakeholder validation conducted in November 2025. Future versions will integrate operational experience from deployment across the ten pilot facilities.

What this document achieves: It translates ethical principles into operational parameters. Storage capacity becomes an algorithmic constraint. Human oversight becomes tier-specific approval workflows. Equity becomes concrete weighting factors. Transparency becomes plain-language explanations visible in user interfaces. Frontline worker protection becomes battery optimization and auto-save functionality.

By grounding AI in humanitarian ethics informed by African operational realities, these guidelines assert both technical capability and moral leadership in defining how intelligent systems serve human need.

APPENDIX A: BASELINE EVIDENCE SUMMARY

These ethical guidelines emerge from empirical findings documented in the IFRAD Baseline Assessment (August-September 2025) across ten facilities in Karamoja and Southwestern Uganda.

Key findings informing ethical requirements:

- 89% of facilities experience unreliable connectivity with typical offline periods of 2-3 days, requiring offline-first ethical frameworks
- Storage capacity shows strong negative correlation with stockout frequency ($r = -0.695$), justifying algorithmic constraints
- Refugee-serving facilities experience longer stockout durations, supporting prioritization weighting
- Facilities conducting weekly inventory counts show lower stockout incidents than monthly counters
- Half of facilities cannot report budget utilization rates, demonstrating information asymmetry requiring transparency mechanisms
- System fragmentation across DHIS2, eAFYA, CSSP, and paper records creates data quality challenges requiring validation protocols
- Staff express strong support for automation provided systems remain offline-capable and allow human override during outbreaks and delivery disruptions

APPENDIX B: TECHNICAL ARCHITECTURE ALIGNMENT

Ethical requirements align with three-tier forecasting architecture:

Tier 1 (Rule-Based, Offline, Facility-Level):

- Uses 3-month moving average with seasonal adjustments and storage capacity caps
- Provides plain-language explanations visible in offline interface
- Allows full offline override with mandatory reason logging
- Implements cold start protocols using stratified facility averages

Tier 2 (Statistical, District-Level):

- Hierarchical Exponential Smoothing (HES) using bottom-up reconciliation runs on district servers
- Provides confidence intervals and data quality flags
- Supplements but cannot override facility decisions (stored as read-only JSON, not in operational SQLite)

Tier 3 (Machine Learning, Central-Level):

- Random Forest or XGBoost for district-level demand prediction and stockout risk scoring
- Provides feature importance rankings to district officers
- Requires human validation before redistribution decisions
- Undergoes quarterly bias audits across facility types and regions

APPENDIX C: VALIDATION SUMMARY

These guidelines incorporate feedback from comprehensive validation process conducted November 2025:

Technical Validation (Kyambogo University):

- Validator: Ojok Ivan, Department of Data Science, Networks & Artificial Intelligence, Kyambogo University
- 8 specifications assessed: offline-first architecture, local data security, database performance, data archiving, system interoperability, UI/mobile layout, data governance, predictive analytics

- Outcome: 7 fully validated, 1 modified (Tier 2 forecasting changed from ETS to HES for scalability)

Stakeholder Validation (November 12, 2025, Golden Tulip Hotel):

- 17 participants representing Ministry of Health, District Health Officers, civil society, WFP, Elrha, academia
- Key concerns raised: frontline worker burden, device charging constraints, equity safeguards, district-level ownership
- Framework positioning: align with NDP IV, emphasize district ownership model, integrate nutrition program platforms

APPENDIX D: REVISION HISTORY

Version	Date	Changes
1.0	September 2025	Initial draft based on preliminary framework design.
2.0	October 2025	Replaced abstract confidence thresholds with tier-specific oversight protocols aligned to actual architecture. Added concrete equity mechanisms with quantified parameters from baseline assessment. Introduced offline decision-making protocols addressing 89% connectivity failure rate. Added cold start equity requirements for new facility initialization. Included trade-offs and ethical tensions section acknowledging implementation constraints. Revised community engagement mechanism to realistic monthly facility surveys. Clarified commercial adaptation stance while maintaining open humanitarian use.
3.0	November 2025	Post-validation revision incorporating stakeholder and technical validation feedback: - Corrected validator attribution from Makerere to Kyambogo University (Ojok Ivan, Department of Data Science) - Added explicit 'Local Edits Always Win' principle in Section 3.4 reflecting validated conflict resolution policy - Added new Section 3.7: Frontline Worker Protection addressing burden reduction, battery optimization, cognitive load, training minimization, device charging accommodation - Updated Tier 2 model description in Appendix B from 'Exponential smoothing' to 'Hierarchical Exponential Smoothing (HES) using bottom-up reconciliation' - Added Nutrition Appointment Platform integration in equity mechanisms (Section 3.2) responding to WFP stakeholder emphasis - Completed governance structure (Section 8) with specific Forecasting Governance Committee composition matching validated framework - Enhanced trade-offs section (6.3) to include infrastructure improvement advocacy paired with algorithmic constraints - Added credential management details (90-day validity) in offline protocols (Section 5.2) - Added new Appendix C documenting validation outcomes: 7 of 8 specifications validated, 1 modified - Updated executive summary to reference NDP IV alignment and November 2025 stakeholder validation - Enhanced data protection section with Android Keystore System specification and 12-month archiving policy