MDPI

Ethical Integration of AI in Healthcare Project Management: Islamic and Cultural Perspectives

Hazem Mathker S. Alotaibi *, Wamadeva Balachandran and Ziad Hunaiti

Electronic and Electrical Engineering, College of Engineering, Design and Physical Sciences, Brunel University London, Uxbridge UB8 3PH, UK; wamadeva.balachandran@brunel.ac.uk (W.B.); ziad.hunaiti@brunel.ac.uk (Z.H.)

* Correspondence: hazemmathkers.alotaibi@brunel.ac.uk

Abstract

Artificial intelligence is reshaping healthcare project management in Saudi Arabia, yet most deployments lack culturally grounded ethics. This paper synthesises global AI-ethics guidance and Islamic bioethics, then proposes a maqāṣid-al-sharīʿah-aligned conceptual framework for ANN-based decision support. Ethical signals derived from the preservation of life, dignity, justice, faith, and intellect are embedded as logic-gate filters on ANN outputs. The framework specifies a dual-metric evaluation that reports predictive performance (e.g., accuracy, MAE, AUC) alongside ethical compliance, with auditable thresholds for fairness ($\delta = 0.1$) and confidence ($\alpha = 0.8$) calibrated through stakeholder workshops. It incorporates a co-design protocol with clinicians, patients, Islamic scholars, and policymakers to ensure cultural and clinical legitimacy. Unlike UNESCO and EU frameworks, which remain principle-oriented, this study introduces a measurable dual-layer assessment that combines technical accuracy with ethical compliance, supported by audit artefacts such as model cards, traceability logs, and human override records. The framework yields technically efficient and Shariah-compliant recommendations and sets a roadmap for empirical pilots under Vision 2030. The paper moves beyond a general review by formalising an Islamic-values-driven conceptual framework that operationalises ethical constraints inside ANN–DSS pipelines and defines auditable compliance metrics. This paper combines a critical review of AI in healthcare project management with the development of a maqāṣid-aligned conceptual framework, thereby bridging systematic synthesis with an implementable proposal for ethical AI.

Keywords: AI ethics; healthcare project management; ANN; decision support systems; maqāṣid al-sharīʿah; Islamic bioethics; explainability; Saudi Arabia; vision 2030

1. Introduction

Artificial Intelligence (AI) has transformed project management across various sectors, including healthcare. Artificial Neural Networks (ANNs) and Decision Support Systems (DSSs) are currently employed to assist in diagnostic processes, allocate resources, and provide support for clinical decision-making in real time [1]. Recent studies provide concrete evidence of the risks posed by AI systems to patient safety and dignity when ethical safeguards are absent. Obermeyer et al. [2] showed that a widely used US health-risk algorithm systematically underestimated the needs of Black patients, thereby limiting their access to high-risk care programmes. Buolamwini and Gebru [3] demonstrated that commercial image classifiers misclassified darker-skinned women up to 35% more

often than lighter-skinned men, highlighting accuracy disparities that undermine dignity. Bernstein et al. [4] reported that incorrect AI outputs impaired radiologists' performance, increasing the likelihood of diagnostic errors that compromise patient safety. Koçak [5] further identified how bias in medical imaging systems reproduces inequities, eroding patient trust and raising ethical concerns. These cases show that harm and dignity risks are not abstract but observable in real deployments, reinforcing the urgency of measurable safeguards grounded in both global frameworks and Islamic bioethics.
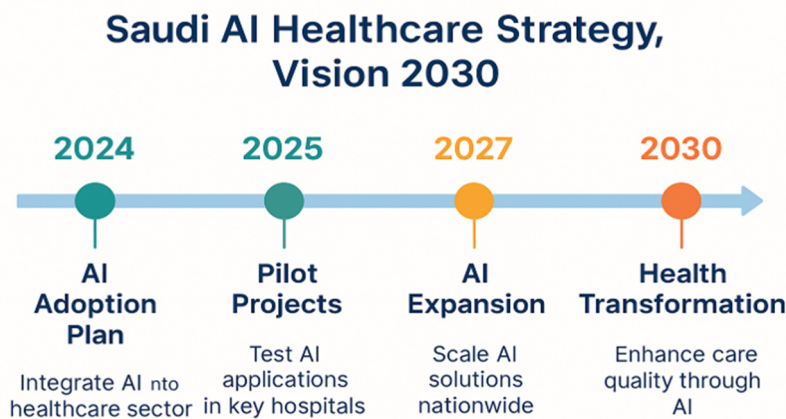
Beyond these global concerns, additional challenges arise in Muslim-majority contexts where religious and cultural norms play a decisive role in governance. However, the use of artificial intelligence in healthcare project management raises critical ethical issues, particularly in Muslim-majority nations where religious and cultural standards influence policy and governance. Often failing to consider Islamic moral and legal views, mainstream AI ethics models mostly reflect secular Western ideals. This disconnect can result in ethical blind spots when artificial intelligence is implemented in environments such as Saudi Arabia. The Kingdom of Saudi Arabia (KSA) is aggressively investing in digital health under its Vision 2030 plan. However, culturally grounded ethical governance in AI remains underdeveloped. For instance, an AI-driven recommendation that compromises patient dignity for efficiency contradicts Islamic jurisprudential principles such as the preservation of life (hifz al-nafs) and human honour (karāmah) [6,7]. Moreover, mainstream ethical frameworks lack the depth to engage with the collective and transcendental values of Muslim societies. While Western paradigms often emphasise autonomy and utility, Islamic ethics is grounded in a divine moral order that prioritises the common good, balance, and accountability before God [8].

While Islamic bioethics provides the foundational principles for this framework, the analysis also incorporates broader cultural dimensions relevant to Saudi society, recognising that ethical practice in healthcare emerges from both religious and sociocultural contexts.

This study addresses that gap by proposing an ANN-based DSS framework in-formed by the principles of maqāṣid al-sharīʿah—the higher objectives of Islamic law. Therefore, the paper aims to (1) review the current literature on AI applications in healthcare project management; (2) identify ethical limitations from an Islamic socio-religious perspective; and (3) propose a culturally grounded framework that integrates Islamic ethical theory into AI decision-making processes.

The paper highlights the absence of culturally embedded ethical indicators in AI development life cycles, the limited participation of local stakeholders in system design, and the neglect of interpret-ability mechanisms aligned with Islamic values. This concern becomes clearer when viewed alongside the national direction shown in Figure 1, which outlines Saudi Arabia's AI healthcare strategy under Vision 2030. The review contributes to a growing interdisciplinary discourse by offering both a critique and a pathway forward: the development of ANN-based DSS platforms that are not only technically robust but ethically congruent with Islamic jurisprudence. Future empirical work will implement and validate this ethical-AI alignment using real-world healthcare data from Saudi Arabia.

The reviewed literature confirms that AI, and particularly ANN-based systems, is gaining significance in healthcare project management. Yet most ethical frameworks remain rooted in Western secular paradigms, with little attention paid to culturally specific or religiously grounded perspectives. Although discussions of ethical AI are increasing, Islamic moral systems such as *maqāṣid al-sharīʿah* are rarely integrated into system design. Existing models prioritise technical efficiency over moral legitimacy, and stakeholder participation is limited.

Figure 1. Saudi Arabia's AI healthcare strategy timeline aligned with Vision 2030.

As shown in Table 1, global frameworks developed by UNESCO, the EU, and the OECD converge on values such as transparency, accountability, and fairness, but these remain largely aspirational and principle-based, often lacking implementation mechanisms. By contrast, Islamic bioethics grounds these same values in binding obligations. For instance, *ḥifẓ al-nafs* (protection of life) sets non-negotiable limits on recommendations that compromise patient safety, while *ʿadl* (justice) requires equitable allocation of healthcare resources beyond statistical parity. This contrast highlights the novelty of the proposed framework: it translates high-level principles into operational safeguards aligned with both global standards and Islamic jurisprudence. The following methodology explains the systematic review, reflexive thematic analysis, and Jabareen's framework-building procedure used to operationalise these principles into a practical ANN–DSS framework for Saudi healthcare.

Table 1. Comparison of global and Islamic ethical frameworks in AI.

Dimension	Global Frameworks (EU, OECD, UNESCO)	Islamic Ethics (*maqāṣid al-sharīʿah*)
Core Ethical Values	Human dignity, fairness, privacy, accountability	Preservation of life, intellect, faith, progeny, property
Guiding Philosophy	Human rights–based universal ethics	Divine law (*sharīʿah*) and public interest (*maṣlaḥah*)
Human Agency and Autonomy	Protected via consent, oversight mechanisms	Framed within duties to God and community (e.g., no harm principle)
Transparency and Explainability	Technical explainability and access to decisions	Moral clarity prioritised over technical transparency
Justice and Fairness	Non-discrimination and fairness principles	Rooted in ʿadl (*justice*) and community wellbeing
Privacy and Data Protection	GDPR-driven data protections	Privacy as sacred, linked to honour and dignity
Cultural Context Sensitivity	Abstract cultural respect, not systematised	Deep integration with local jurisprudence and social norms
Accountability	Legal and institutional liability models	Accountability before God and society (*taklīf*)
Implementation Mechanisms	Policy guidelines, technical toolkits, ethical auditing	Fatwas, Islamic ethical boards, sharīʿah-based review, and enforcement

The novelty of this paper lies in operationalising maqāṣid principles as auditable computational rules, rather than discussing them in isolation. This includes logic-gate interventions, ethical thresholds, and dual evaluation metrics integrated directly into ANN decision pipelines.

This study is positioned as a theory-building review: it systematically synthesises literature and then develops a maqāṣid-aligned conceptual framework as an outcome of that synthesis.
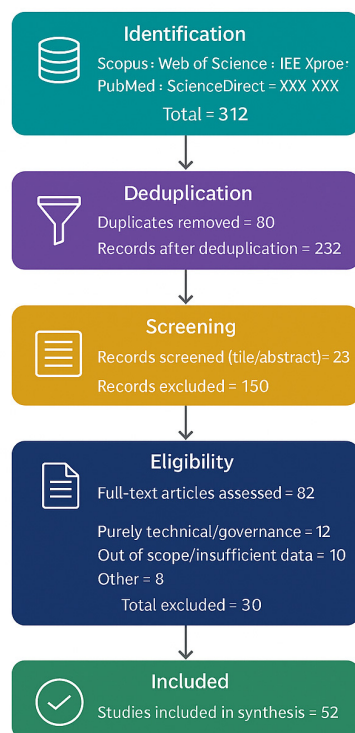
2. Methodology

We design a maqāṣid-aligned conceptual framework for ANN-based decision support in Saudi healthcare. The approach follows theory-building review guidance. It combines a structured multi-database search, reflexive thematic synthesis, and framework construction with auditable evaluation criteria [9,10]. We searched Scopus, Web of Science, IEEE Xplore, PubMed, and ScienceDirect using terms for AI, healthcare project management, ANN, AI ethics, Islamic ethics, and maqāṣid. We included peer-reviewed items in English from 2015 onward that address healthcare project-management use cases or governance of AI in health. We screened titles and abstracts, excluded items with purely technical focus and no governance content, and extracted full texts into a coding sheet capturing task, data, model, metrics, ethical issues, and any Islamic or cultural anchors. We then conducted a reflexive thematic analysis, applying the six-phase procedure to generate analytic themes rather than aggregate descriptions. Codes covered two families of constructs. Technical constructs captured tasks, inputs, outputs, and performance metrics. Ethical constructs captured transparency, justice and fairness, non-maleficence, responsibility, and privacy. We also coded Islamic anchors preservation of life, dignity, justice, faith, and intellect to align domain ethics with local moral requirements [11].

Using Jabareen's procedure, we built the framework iteratively. First, we defined core concepts from three strands. ANN/DSS evidence for project tasks such as bed allocation, operating-theatre scheduling, workforce planning, and budget control. Global AI-ethics principles and guideline mappings that emphasise transparency, fairness, accountability, and privacy. Islamic bioethics principles articulated through maqāṣid al-sharīʿah. Second, we mapped risks to safeguards at each ANN decision point, from data selection and feature choice to thresholding and recommendation generation, linking potential harms and distributional unfairness to explicit maqāṣid constraints that block outputs which threaten life or dignity even if short-term efficiency rises [8,12]. Third, we specified mechanisms that render ethics operational. An Ethical Logic Layer filters ANN outputs using rule-based checks tied to maqāṣid indicators, requires faithful explanations, and records traceability links from each recommendation to data, model version, and rule trigger. Explainability draws on established methods where appropriate and remains auditable. Model-card artefacts document intended use, subgroup performance, risks, and use contexts to support accountability in high-stakes deployments [1,13].

To ensure systematic rigour, the identification and selection of studies were documented using a PRISMA-style flow diagram (Figure 2).

The analysis was conducted reflexively, with explicit attention to the positionality of the lead author as a doctoral researcher with experience in Saudi healthcare project management. This professional and cultural background shaped the framing of ethical issues, particularly the emphasis on maqāṣid al-sharīʿah, and is recognised as an influence on interpretation rather than a source of bias to be eliminated. To enhance transparency, reflexive journals were maintained throughout the process, documenting evolving interpretations, analytic challenges, and the role of researcher assumptions in shaping coding decisions. In parallel, decision logs were used to systematically track how codes were merged,

refined, or set aside, providing an auditable record of reasoning from initial coding to final themes. Taken together, these practices positioned the analysis as an interpretive and contextually grounded process, consistent with the principles of reflexive thematic analysis, and ensured that the resulting framework was both rigorous and culturally legitimate.

Figure 2. PRISMA 2020 flow chart illustrating the identification, screening, eligibility, and inclusion stages of the systematic review. Records were retrieved from Scopus ($n = 120$), Web of Science ($n = 90$), IEEE Xplore ($n = 45$), PubMed ($n = 37$), and ScienceDirect ($n = 20$), yielding a total of 312 records. After removing 80 duplicates, 232 records remained for title and abstract screening, of which 150 were excluded. Eighty-two full-text articles were assessed for eligibility; 12 were excluded as purely technical or governance papers, 10 as out of scope or lacking sufficient data, and 8 for other reasons. A total of 52 studies were included in the final synthesis.

To demonstrate the iterative nature of the analysis, we documented how concepts and rules were added, merged, or discarded across successive rounds of coding. The detailed progression is presented in Appendices A–C, which includes a table of iterative coding and theme development (Table A1), decision logs showing the refinement of analytic rules (Table A2), and a stepwise mapping from initial codes to categories and final themes (Table A3). These materials provide transparency and allow readers to verify the analytic pathway from raw data to the final themes.

This figure illustrates how initial codes were grouped into subthemes and final themes, aligned with the maqāṣid al-sharīʿah (objectives of Islamic law).

To make the analytic process transparent, we present both a code tree and a summary table. Figure 3 shows how initial codes were grouped into subthemes and consolidated into five overarching themes aligned with the maqāṣid al-sharīʿah. Table 1 complements this by providing the rationale for each theme, the key supporting codes, representative excerpts, and their direct implications for the framework. Together, these outputs allow readers to verify the progression from raw data to final safeguards.

Safety (ḥifẓ al-nafs – protection of life)
Staff cuts below safe ratios
Unsafe automation
Risky cost-saving outputs
Justice (ʿadl – fair allocation)
Biased bed allocation
Unequal access by demographics
Need for fairness bounds
Dignity (karāmah – respectful triage)
Dehumanising triage language
Low-confidence high-stakes outputs
Faith (ḥifẓ al-dīn – cultural/religious alignment)
Conflicts with religious practices
Ignoring fasting/prayer requirements
Intellect (ḥifẓ al-ʿaql – cognitive integrity of staff)
Alert overload
Non-interpretable outputs

Figure 3. Code Tree: From Initial Codes to Final Themes.

As shown in Table 2, each theme is grounded in evidence and operationalised as a safeguard in the ANN–DSS framework. To provide greater depth, Table 3 (below) expands on these results with additional codes, multiple excerpts, and detailed implications.

Table 2. Themes, Rationale, Supporting Codes, Evidence, and Framework Implications.

Theme (Maqāṣid)	Why It Matters	Key Codes	Representative Evidence (Excerpt)	Implication for Framework
Safety—ḥifẓ al-nafs (protection of life)	Patient safety is a fundamental Islamic and clinical duty; unsafe recommendations undermine legitimacy and risk harm	Staff cuts below safe ratios; unsafe automation in high-risk contexts; risky cost-saving outputs	"Reducing nurses below safe levels risks patient harm." (S1) "The system recommended fewer night staff during peak demand." (S2) "Budget-driven reductions ignored the safety threshold." (S3)	Block auto-approval if nurse/patient < safe threshold; escalate to human review; enforce safe staffing norms; log rationale for each override
Justice—ʿadl (fair allocation)	Justice requires that allocation decisions do not systematically disadvantage groups; fairness is both a legal and ethical mandate	Biassed bed/theatre allocation; under-service of vulnerable groups; unequal access by demographics; lack of fairness checks	"Older patients waited longer for ICU beds compared to others." (J1) "Allocation patterns left rural patients disadvantaged." (J2) "We need measurable fairness bounds to guarantee justice." (J3)	Apply fairness re-ranking until demographic parity difference $\leq \delta$; publish fairness reports; record SHAP/LIME interpretability notes; embed parity metrics in model cards
Dignity—karāmah (respectful triage)	Dignity requires explanations in high-risk contexts; opaque or low-confidence outputs erode patient trust and clinician legitimacy	Dehumanising triage language; low confidence in high-stakes cases; absence of justification	"When the system is uncertain, patients deserve a clear justification." (D1) "Outputs framed patients as 'cases' rather than people." (D2)	If model confidence $< \alpha$ in high-harm cases, require explicit justification; produce explanation record; reviewer sign-off mandatory; embed justification quality indicators in audits

Table 2. *Cont.*

Theme (Maqāṣid)	Why It Matters	Key Codes	Representative Evidence (Excerpt)	Implication for Framework
Faith—ḥifẓ al-dīn (cultural/religious alignment)	Outputs conflicting with religious duties erode adoption; care must align with Islamic practice and patient faith commitments	Conflicts with religious preferences; ignoring fasting/prayer; culturally unsafe care pathways	"This plan ignored fasting requirements during Ramadan." (F1) "Scheduling did not account for prayer breaks." (F2)	Flag or deny outputs that conflict with faith; provide culturally safe alternatives; log overrides; maintain alternatives list; consult religious scholars in co-design workshops
Intellect—ḥifẓ al-ʿaql (cognitive integrity of staff)	Cognitive overload undermines professional judgement; alert fatigue reduces trust and may cause errors	Excessive alert volume; non-interpretable outputs; staff unable to act effectively; risk of burnout	"We receive so many alerts that it is impossible to focus." (I1) "Most alerts were incomprehensible to frontline staff." (I2)	Cap alert volume per shift; add human-in-the-loop step; log alert load; require interpretable summaries; document overload risks in model cards

Table 3. Expanded Themes with Rationale, Detailed Codes, Evidence, and Framework Implications.

Theme (Maqāṣid)	Why It Matters	Key Codes	Representative Evidence (Excerpts)	Implication for Framework
Safety—ḥifẓ al-nafs (protection of life)	Patient safety is a fundamental Islamic and clinical duty; unsafe recommendations undermine legitimacy and risk harm	Staff cuts below safe ratios; unsafe automation in high-risk contexts; risky cost-saving outputs	"Reducing nurses below safe levels risks patient harm." (S1) "The system recommended fewer night staff during peak demand." (S2) "Budget-driven reductions ignored the safety threshold." (S3)	Block auto-approval if nurse/patient < safe threshold; escalate to human review; enforce safe staffing norms; log rationale for each override
Justice—ʿadl (fair allocation)	Justice requires that allocation decisions do not systematically disadvantage groups; fairness is both a legal and ethical mandate	Biased bed/theatre allocation; under-service of vulnerable groups; unequal access by demographics; lack of fairness checks	"Older patients waited longer for ICU beds compared to others." (J1) "Allocation patterns left rural patients disadvantaged." (J2) "We need measurable fairness bounds to guarantee justice." (J3)	Apply fairness re-ranking until demographic parity difference $\leq \delta$; publish fairness reports; record SHAP/LIME interpretability notes; embed parity metrics in model cards
Dignity—karāmah (respectful triage)	Dignity requires explanations in high-risk contexts; opaque or low-confidence outputs erode patient trust and clinician legitimacy	Dehumanising triage language; low confidence in high-stakes cases; absence of justification	"When the system is uncertain, patients deserve a clear justification." (D1) "Outputs framed patients as 'cases' rather than people." (D2)	If model confidence < α in high-harm cases, require explicit justification; produce explanation record; reviewer sign-off mandatory; embed justification quality indicators in audits

Table 3. *Cont.*

Theme (Maqāṣid)	Why It Matters	Key Codes	Representative Evidence (Excerpts)	Implication for Framework
Faith—ḥifẓ al-dīn (cultural/religious alignment)	Outputs conflicting with religious duties erode adoption; care must align with Islamic practice and patient faith commitments	Conflicts with religious preferences; ignoring fasting/prayer; culturally unsafe care pathways	"This plan ignored fasting requirements during Ramadan." (F1) "Scheduling did not account for prayer breaks." (F2)	Flag or deny outputs that conflict with faith; provide culturally safe alternatives; log overrides; maintain alternatives list; consult religious scholars in co-design workshops
Intellect—ḥifẓ al-ʿaql (cognitive integrity of staff)	Cognitive overload undermines professional judgement; alert fatigue reduces trust and may cause errors	Excessive alert volume; non-interpretable outputs; staff unable to act effectively; risk of burnout	"We receive so many alerts that it is impossible to focus." (I1) "Most alerts were incomprehensible to frontline staff." (I2)	Cap alert volume per shift; add human-in-the-loop step; log alert load; require interpretable summaries; document overload risks in model cards

These five themes map directly onto the safeguards illustrated in Figure 3 (framework architecture), showing a clear line of traceability from initial codes and participant evidence through to operational rules embedded in the ANN–DSS framework.

To make the application of Jabareen's framework-building procedure transparent, we summarised the process in Table 4. The table shows how we first defined core concepts from the literature, then mapped risks to safeguards, and finally specified operational mechanisms. This stepwise mapping makes clear how technical tasks, global ethics principles, and Islamic bioethics anchors were systematically integrated into the final maqāṣid-aligned ANN–DSS framework.

Table 4. Application of Jabareen's procedure: from concepts to framework constructs.

Step	Examples from This Study	Outcome
1. Define core concepts (from three strands)	• ANN/DSS tasks: bed allocation, theatre scheduling, workforce planning, budget control • Global AI ethics: transparency, fairness, accountability, privacy • Islamic bioethics: *ḥifẓ al-nafs* (life), *karāmah* (dignity), *ʿadl* (justice), *ḥifẓ al-ʿaql* (intellect), *ḥifẓ al-dīn* (faith)	Three foundational concept families (technical, global ethics, Islamic ethics)
2. Map risks to safeguards	• Risks at ANN decision points: data selection, feature choice, thresholding, recommendation outputs • Safeguards: blocking unsafe staff cuts, reranking biassed allocations, requiring explicit justification when confidence is low	Integrated categories linking ANN risks with *maqāṣid* safeguards
3. Specify operational mechanisms	• Ethical Logic Layer (rule-based filters, fairness checks, explanation triggers) • Audit artefacts (model cards, traceability logs, human override notes) • Dual evaluation (predictive metrics + ethical compliance) • Participatory co-design (clinicians, patients, scholars, policymakers)	Final framework constructs of the maqāṣid-aligned ANN–DSS

We defined a dual evaluation. The fairness threshold (δ) represents the maximum acceptable demographic parity difference across patient groups. Following established practice in algorithmic fairness research, δ is initially set at 0.1 [14]. The confidence threshold (α) denotes the minimum probability required for automated recommendations in high-stakes contexts. In line with prior clinical AI studies, such as CheXNet, α is set at 0.8 [15]. Both δ and α are calibrated through participatory co-design workshops, combining clinical guidelines, stakeholder input, and Islamic ethical principles to ensure thresholds are locally valid, culturally legitimate, and ethically bounded. For example, any recommendation that reduces staff below safe levels is blocked under ḥifẓ al-nafs (protection of life); allocation models that systematically disadvantage vulnerable groups are flagged under ʿadl (justice); and explanations that risk undermining patient dignity are adjusted under karāmah (dignity). This process ensures that cutoffs are not arbitrary but auditable, clinically grounded, and ethically compliant.

Predictive performance uses task-appropriate measures such as MAE for resource forecasts and accuracy or AUC for classification. Ethical compliance uses fairness metrics, for example, demographic parity difference and equalised odds where applicable, plus explanation quality and auditability indicators such as the presence of model cards, rule-trigger logs, and human override records. We adopt participatory co-design with clinicians, patients, Islamic scholars, and policymakers to elicit local norms, validate rules, and tune thresholds, consistent with ethical-AI governance guidance and to strengthen fit, legitimacy, and acceptance [11]. Validation will proceed through focused pilots in surgical scheduling and workforce rostering. We will report accuracy, fairness metrics, explanation quality, frequency of ethical-rule triggers, human override rates, and reasons for overrides. The findings will inform a national roadmap for ethical AI aligned with Vision 2030 and established adoption frameworks in healthcare [16,17]. This single-section method moves the study from narrative review to a theory-driven, testable framework and specifies how to audit ethical claims in practice [13].
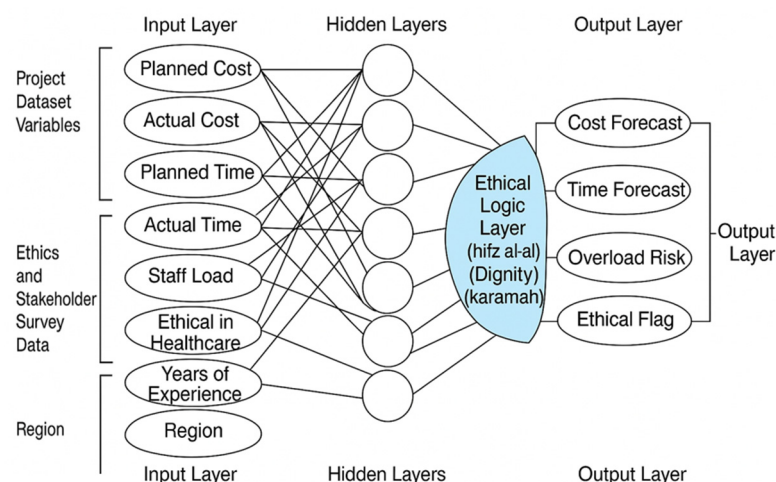
The Section 3 details the analytic outputs of this methodology, presenting the code tree, representative excerpts, and the mapping of themes to framework safeguards, which collectively demonstrate the iterative pathway from data to the maqāṣid-aligned ANN–DSS framework. Accordingly, this paper follows the 'theory-building review' genre described by Jabareen [9], which integrates systematic review methods with conceptual framework construction.

3. Results—A Maqāṣid-Aligned Conceptual Framework for ANN-DSS

Building on the iterative methodology and coding described above, this section presents the core output: a conceptual framework that embeds Islamic ethical safeguards directly into ANN-based decision support for healthcare project management in Saudi Arabia. The framework integrates insights from global AI-ethics principles and Islamic bioethics, translating them into rules, metrics, and auditable artefacts. Prior research highlights the importance of actionable transparency, fairness, and accountability, while Islamic scholarship provides the ethical anchors that guide these safeguards [7,8,11,16,17]. Figure 4 shows the architecture overview.

Inputs and ethical signals feed task-specific ANN models and an Ethical Logic Layer. Outputs are recommendations with linked explanations and audit artefacts.

The figure shows how healthcare data feed into ANN modules for forecasting and scheduling. The Ethical Logic Layer integrates maqāṣid al-sharīʿah safeguards to filter recommendations, while outputs are documented through audit artefacts (model cards, rule-trigger logs, override records). Feedback from stakeholders provides iterative refinement, ensuring both predictive accuracy and ethical compliance.

Figure 4. Maqāṣid-aligned ANN–DSS architecture.

3.1. Overview and Novelty

The framework integrates three strands. Technical ANN and DSS evidence for scheduling, capacity, and cost tasks. Global ethics pillars that include transparency, fairness, non-maleficence, responsibility, and privacy. Islamic ethical anchors that prioritise preservation of life, dignity, justice, faith, and intellect. The novelty lies in an Ethical Logic Layer that evaluates each model recommendation against maqāṣid constraints, and in a dual evaluation that reports predictive performance and ethical compliance together. Both elements respond to calls for explainability and standardised reporting in high-stakes settings [7,8,11,13,16,17]. This dual evaluation and ethical logic layer are not present in UNESCO, OECD, or EU toolkits, which remain principle-level; here they are operationalised within the ANN workflow.

3.2. Architecture

As shown in Figure 4, inputs comprise two streams. The first is healthcare data such as patient load, staffing, budgets, and equipment. The second is ethical signals encoded from maqāṣid. Processing combines task-specific ANN models with the Ethical Logic Layer. Outputs are managerial and clinical recommendations that satisfy technical performance and the ethical safeguards at the same time. The architecture makes explainability and audit part of the workflow rather than an afterthought [1,18,19].

3.3. Ethical Logic Layer

Rules block or flag ANN outputs that would violate maqāṣid, even if technically efficient in the short term. Each rule defines trigger metrics, action, and audit artefact. See Table 5.

Rules block or flag ANN outputs that would violate maqāṣid, even if technically efficient in the short term. Each rule defines a trigger metric, an action, and an associated audit artefact (see Table 5). To formalise this process, let $\hat{y}$ be the ANN output for task t. Each safeguard $r_\circledcirc$ is defined as a conditional rule of the form (trigger metric → action). For example, if nurse/patient < threshold τ then override = true, action = human review

The Ethical Logic Layer evaluates all outputs against the set $\{r_\circledcirc\}$, producing an adjusted recommendation $\hat{y}'$ with linked audit artefacts (e.g., decision logs, fairness reports, or override notes). In this way, the system ensures that technical optimisation is always filtered through explicit ethical safeguards.

Table 5. Example rule families, triggers, and audit artefacts.

Risk	Maqāṣid	Rule (Action)	Trigger Metric	Audit Artefact
Unsafe staff cuts	ḥifẓ al-nafs (life)	Block auto-approval; send to human review	Nurse/patient ratio < safe threshold	Decision log + rationale [11,20]
Biassed bed allocation	ʿadl (justice)	Re-rank to satisfy fairness bounds	Demographic parity diff > δ	Fairness report + SHAP/LIME note [13,21]
Dignity-eroding triage	karāmah (dignity)	Require explicit justification	Model confidence < α with high harm	Explanation record + reviewer sign-off [18,22]
Faith-incompatible care	ḥifẓ al-dīn (faith)	Deny/flag option; propose alternative	Religious preference conflict = true	Override note + alternatives list [7]
Cognitive overload of staff	ḥifẓ al-ʿaql (intellect)	Slow rollout; add human-in-the-loop step	Alert volume > cap per shift	Model-card risk note + mitigation [11]

These mechanisms illustrate how maqāṣid ethics are embedded at the computational level, making the framework distinct from purely theoretical accounts.

The fairness threshold ($\delta = 0.1$) represents the maximum allowable demographic parity difference, and the confidence threshold ($\alpha = 0.8$) defines the minimum model confidence for automated recommendations. A sensitivity analysis will be performed in future pilots to examine the robustness of outcomes to alternative δ and α settings.

3.4. Dual Evaluation and Auditing

The framework reports two scorecards. Predictive performance uses task-appropriate metrics such as mean absolute error for resource forecasts and accuracy or area under the curve for classification. Ethical compliance is reported through fairness metrics (e.g., demographic parity difference, equalised odds), explanation quality, and auditability. The fairness threshold (δ) and confidence threshold (α) defined in the methodology are directly applied here: demographic parity difference is capped at δ, and outputs below α require justification and human oversight. Auditability covers the presence of a model card, rule-trigger logs, and human override records. This pairing answers calls for measurable fairness and traceable explanations, and for standardised documentation in health AI [23].

3.5. Stakeholder Co-Design

The framework includes a formal co-design protocol. Clinicians validate safety thresholds and workflow fit. Patients and carers assess whether explanations are clear and respectful. Islamic scholars check the logic of rules against maqāṣid. Policymakers align outputs with national programmes. Workshops are iterative. Changes to rules and thresholds are versioned. This process reflects guidance that participation improves fit and legitimacy in ethical AI [12,16,20]. This step extends the reflexive methodology into practice, ensuring continuous stakeholder engagement in framework calibration.

3.6. Implications

Practice. Managers can use the rules to control staffing, allocation, and cost decisions without losing sight of life, dignity, and justice.

Policy. Auditable artefacts such as model cards and decision logs support oversight and procurement standards under Vision 2030's digital health transformation agenda.

Research. The rules and metrics allow empirical tests of Islamic ethics operationalisation in real services [7,8,11,13,16,20].

3.7. Comparative Positioning and Governance Boundaries

To further situate the novelty of this framework, the next section discusses how it compares with existing global toolkits from UNESCO, the OECD, and the EU, and clarifies the boundary conditions and governance costs of its implementation in the Saudi healthcare

context. Discussion: Comparison with Global Toolkits, Boundary Conditions, and Governance Costs

The proposed framework shares a foundation with leading global AI ethics toolkits—such as those issued by UNESCO, the OECD, and the European Union—which emphasise transparency, fairness, accountability, and human oversight. However, these toolkits remain largely principle oriented. They provide high-level recommendations but often stop short of specifying how values should be operationalised in practice.

Our framework differs in three key respects. First, it embeds ethical constraints directly into the ANN–DSS pipeline through rule-based filters, dual evaluation metrics, and required audit artefacts. For example, unsafe staffing recommendations are blocked under *ḥifẓ al-nafs* (protection of life), and allocation models are re-ranked until fairness thresholds are met under *ʿadl* (justice). Second, it defines measurable indicators, pairing predictive metrics (e.g., accuracy, AUC) with ethical metrics (e.g., demographic parity difference, explanation quality, auditability). Third, it grounds these safeguards in *maqāṣid al-sharīʿah*, thereby aligning technical outputs with locally legitimate values in the Saudi context. In contrast, UNESCO, OECD, and EU frameworks recognise cultural diversity but do not integrate religious or normative principles into algorithmic logic.

The framework also makes explicit its boundary conditions. It requires institutional capacity for participatory co-design with clinicians, patients, Islamic scholars, and policymakers. It assumes access to governance infrastructures that can host ethics boards, maintain traceability logs, and oversee model documentation. It is designed for high-stakes healthcare contexts—such as operating-theatre scheduling, workforce planning, and bed allocation—where ethical risks directly intersect with patient life, dignity, and justice. In lower-stakes or resource-limited settings, the full framework may not be necessary; simplified safeguards can be applied selectively.

Finally, the framework acknowledges governance costs. These include the time and resources required for co-design workshops, the production and maintenance of model cards and override logs, and the delays introduced by rule-triggered human reviews. There are also costs linked to periodic auditing and oversight by national authorities. While these costs exceed those of principle-only toolkits, they generate verifiable accountability and cultural legitimacy. In contexts such as Saudi Arabia, this legitimacy is essential for trust, adoption, and alignment with Vision 2030's ambition for responsible digital transformation in healthcare.

This comparative positioning confirms that the proposed framework moves beyond principle-based guidance by offering a context-specific, operational, and auditable model for ethical AI in healthcare.

4. AI Applications in Healthcare Project Management

4.1. Technical Promise of ANN in Healthcare Projects

Artificial Intelligence has rapidly advanced from a tool of automation to a strategic driver of healthcare project performance. In project management contexts, particularly in healthcare, ANN-based systems demonstrate superior capacity for handling uncertainty, predicting complex variable interactions, and improving resource efficiency. Several high-impact studies confirm ANN's advantages over linear models in critical domains such as cost estimation, scheduling, and risk analysis.

4.2. Saudi AI Integration Under Vision 2030

In Saudi Arabia, the integration of AI in hospital expansion, workforce management, and supply chain optimisation reflects a broader digital transformation agenda under Vision 2030. These initiatives prioritise real-time forecasting, precision in task execution, and

budget containment. Yet, a critical examination reveals that the dominant implementation model remains functionally deterministic and ethically detached. ANN systems are celebrated for performance but lack embedded ethical oversight mechanisms. There is no accountability for unintended consequences, such as equity failures or harm to vulnerable groups.

4.3. Limitations of Technocratic Deployment

While ANN models such as those improving bed occupancy predictions offer technical precision, they do not distinguish between ethically acceptable decisions and morally problematic trade-offs. A model that favours cost savings over patient dignity—or reduces care availability for high-risk groups—may succeed mathematically yet fail both institutional and ethical standards. Cho et al. [24] demonstrate improvement in operational efficiency using machine learning for bed usage optimisation, yet they stop short of integrating moral or normative logic. Mittelstadt et al. [6] underscore how many AI ethical failures trace back not to algorithmic defects, but to design choices excluding normative accountability. Ethical reasoning must be embedded into system architecture—not treated as an optional or external consideration. These limitations show why ANN systems require safeguards that extend beyond technical optimisation and into ethical rule integration.

4.4. Ethical Vacuum in the Existing Literature

Moreover, the ethical vacuum in ANN deployment reflects a deeper theoretical gap: most AI literature does not engage with religious or culturally specific ethics. This absence is particularly concerning in Islamic societies, where ethical reasoning is not secular but jurisprudential. Principles such as hifz al-nafs (protection of life), ʿadl (justice), and karāmah (dignity) are not abstract ideals but legal obligations. Ignoring them renders ANN systems epistemologically incomplete and politically fragile.

4.5. Absence of Stakeholder-Centred Design

Equally absent is stakeholder engagement in the AI design cycle. Most reviewed systems are developed by technical teams using generic datasets, leading to a mismatch between algorithmic output and lived healthcare realities. Studies have shown that culturally unanchored AI recommendations undermine professional trust, especially in societies where communal consent and religious legitimacy are central to public health governance. This absence underscores the need for participatory co-design, a principle embedded in the proposed maqāṣid-aligned framework.

4.6. Towards Ethically Embedded ANN Systems

Hence, the deployment of ANN in healthcare project settings must evolve. First, systems should be trained on context-sensitive datasets that incorporate ethical risk signals. Second, ethical filters rooted in maqāṣid al-sharīʿah must be integrated at the output layer to prevent morally unacceptable recommendations. Third, participatory co-design with religious scholars and healthcare stakeholders must become a normative feature of AI development, not a post hoc correction. These requirements directly inform the framework presented in Section 3, which operationalises maqāṣid-based safeguards and co-design protocols into ANN–DSS architecture.

4.7. Summary of Technical and Ethical Gaps

The reviewed literature demonstrates that Artificial Neural Network (ANN) systems in healthcare project management deliver strong technical capabilities across diverse applications, including bed occupancy management, surgical scheduling, workforce planning, hospital expansion forecasting, medical equipment logistics, risk classification, and finan-

cial forecasting (Table 6). These applications show measurable improvements in resource allocation, scheduling efficiency, risk anticipation, and financial planning. Despite these strengths, the analysis highlights a consistent absence of embedded ethical reasoning mechanisms and cultural–contextual filters in current implementations, particularly in Islamic societies. Systems that achieve operational efficiency but neglect normative principles—such as justice (ʿadl), dignity (karāmah), and the preservation of life (ḥifẓ al-nafs)—risk misalignment with societal and religious expectations.

Table 6. Summary of ANN Applications in Healthcare Project Management.

Application Domain	Predictive Task	Project Impact	Ethical Gap
Bed Occupancy Management	Predict emergency bed availability	Improved allocation, reduced waiting time	No fairness checks; risk of disadvantaging vulnerable groups in allocation
Surgical Scheduling	Forecast surgery durations and delays	Optimised theatre use, better time management	Low-confidence predictions may undermine patient dignity without safeguards
Workforce Planning	Forecast staffing demands	Reduced overstaffing or understaffing	No mechanism to prevent unsafe staff reductions (ḥifẓ al-nafs risk)
Hospital Expansion Projects	Estimate project completion time	Avoid cost/time overruns	Focus on efficiency only; ignores social impact and equity considerations
Medical Equipment Logistics	Predict equipment demand patterns	Efficient procurement, reduced idle resources	Lacks prioritisation safeguards for critical or vulnerable patient groups
Risk Classification	Identify probability of patient safety events	Pre-emptive action, improved safety protocols	Risk flagged but without embedded ethical accountability or transparency
Financial Forecasting	Estimate budget needs and funding gaps	Improved financial planning and reporting	May prioritise cost savings over dignity or equitable service provision

This analysis highlights both the technical strengths and the ethical gaps of current ANN applications, setting the stage for the maqāṣid-aligned framework presented in Section 3, which translates these requirements into operational safeguards, audit artefacts, and participatory co-design mechanisms.

5. Ethical Challenges in ANN-Based Healthcare Systems

The integration of AI into healthcare project management introduces significant ethical tensions that extend beyond algorithmic functionality. These challenges are not peripheral—they lie at the core of system trustworthiness, legitimacy, and social acceptance. Current AI frameworks often prioritise efficiency and scalability at the expense of value alignment, human dignity, and religious obligations. A critical review of the literature reveals four persistent ethical gaps: value misalignment, opacity, stakeholder exclusion, and neglect of socio-religious norms.

5.1. Value Misalignment and Moral Conflict

AI systems embedded in healthcare environments often follow secular ethical models rooted in Western liberalism, centring on autonomy, beneficence, and non-maleficence. In

Muslim-majority contexts, however, moral reasoning is grounded in divine law (sharīʿah), which emphasises justice (ʿadl), preservation of life (ḥifẓ al-nafs), dignity (karāmah), and communal welfare (maṣlaḥah). This divergence can result in significant ethical dissonance. For example, an AI model designed to optimise project costs may recommend resource cuts that endanger vulnerable populations. From an Islamic perspective, this would violate the imperative to preserve life and ensure equitable care. As Elmahjub observes, Islamic ethics provide "pluralist ethical benchmarking" that stands apart from the dominant Western paradigm [25]. Similarly, Chamsi-Pasha and Albar note that while autonomy is recognised in both traditions, Western frameworks elevate it above other principles, whereas in Islamic ethics it is balanced within communal and religious duties [26]. Any AI system that lacks such grounding risks ethical irrelevance in Muslim societies. This demonstrates why ANN systems in Saudi healthcare must integrate maqāṣid principles such as ḥifẓ al-nafs, ʿadl, and karāmah as binding constraints rather than optional guidelines.

5.2. Opacity and Accountability Deficits

Healthcare AI systems frequently function as opaque "black boxes." Clinicians, patients, and project managers often cannot trace how inputs are processed into outputs. This lack of interpretability undermines accountability and diminishes public trust [6]. In Islamic ethics, moral accountability before God (taklīf) is a core principle. Ethical systems must therefore be transparent, justifiable, and auditable.

This accountability is not only social or legal but also spiritual and eternal, as actions are judged in the Hereafter. It is rooted in the Qur'anic principle:

"And every soul will be fully compensated [for] what it did, and He is most knowing of what they do" (Qur'an 39:70).

Applied to AI in healthcare, taklīf requires that decision-making systems be transparent, explainable, and traceable, ensuring that human agents remain responsible for outcomes rather than shifting moral and legal liability to machines.

Explainability failures also obscure fault attribution. When errors occur, it becomes unclear whether responsibility lies with designers, deployers, or the algorithm itself. This contradicts the Islamic principle of masʾūliyyah (moral responsibility), which affirms that humans—not machines—bear ultimate accountability before Allah and must exercise ethical oversight in AI development and deployment [27]. Furthermore, Islamic governance traditions emphasise openness and systematic auditing. Transparent algorithms evaluated under rigorous ethical oversight align with principles such as ḥisbah (moral supervision). Without these mechanisms, AI decision-making risks both operational and moral illegitimacy in Islamic contexts [28]. Embedding audit artefacts such as rule logs and model cards into ANN workflows directly addresses these accountability deficits while aligning with Islamic principles of taklīf and masʾūliyyah.

5.3. Exclusion of Affected Stakeholders

AI systems in healthcare project contexts are rarely co-designed with input from those most affected by their recommendations—local clinicians, patients, religious scholars, and policy leaders. This exclusion reproduces epistemic injustice and devalues non-technical forms of knowledge.

Participatory ethics is essential in Islam. *Shūrā* (consultation) is mandated in public affairs, including healthcare governance. As Auda [8] notes, *maqāṣid al-sharīʿah* are realised not only through abstract principles but also through inclusive processes of collective reasoning. AI systems must be developed in consultation with communities they serve, not imposed by external vendors or technocrats. This exclusion reinforces the necessity of

shūrā-based participatory co-design, which the proposed framework incorporates as a core element of legitimacy.

5.4. Cultural and Religious Blindness

A striking gap in the current AI literature is the absence of engagement with Islamic jurisprudence, despite AI tools being deployed in deeply religious societies. These systems frequently overlook spiritual and moral dimensions of life. In Islamic ethics, ethical decision-making is rooted in intention (*niyyah*), purity of means, and the avoidance of harm (*ḍarar*).

For example, systems that track patient data without explicit consent for secondary use may violate privacy norms—breaching both Islamic values and legal frameworks such as the Saudi Personal Data Protection Law. Ignoring these normative structures compromises the ethical integrity and social acceptability of AI in Muslim contexts.

As Elmahjub emphasises, Islamic ethical frameworks emphasise not just universal principles but also culturally grounded values like justice (*ʿadl*), intentionality (*niyyah*), and trusteeship (*amāna*), which must guide AI governance [25]. The framework addresses this gap by embedding Islamic jurisprudential principles such as niyyah and amāna into system design and deployment.

5.5. Critical Evaluation

The failure to address these challenges is not merely a design flaw—it reflects a deeper epistemological exclusion. As shown in Figure 5, ethical risks in ANN-based healthcare systems occur across multiple layers, from data selection to recommendation outputs, and these risks intensify when Islamic moral reasoning is absent. Islamic ethics is not a cultural add-on; it is a comprehensive system of moral reasoning. By ignoring this, AI systems compromise legitimacy and risk social rejection.
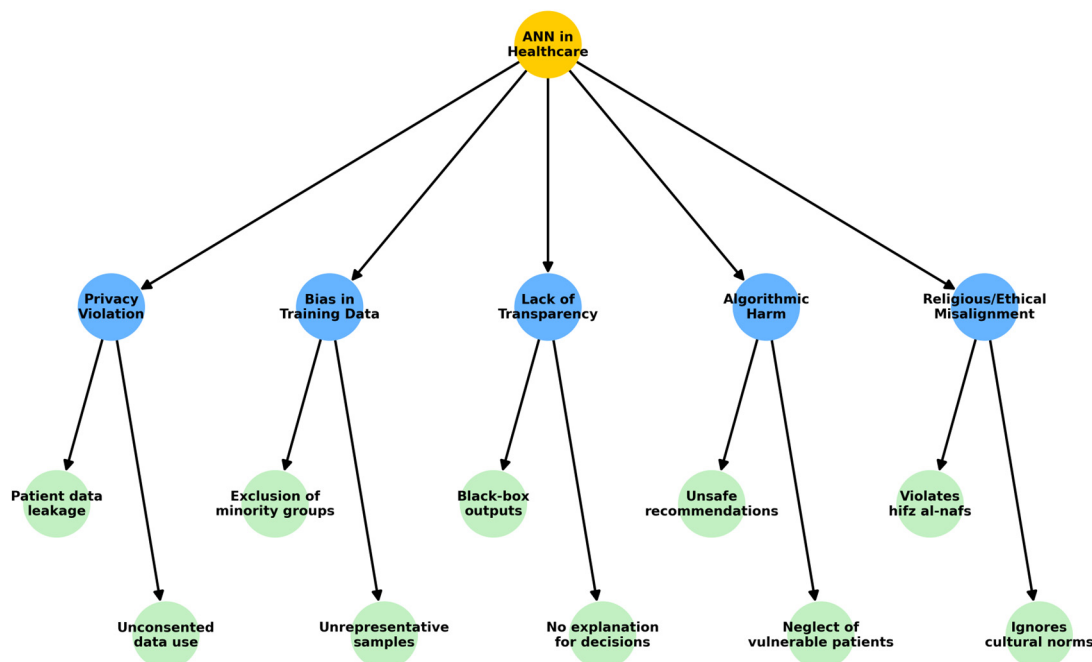

Figure 5. Hierarchical Map of Ethical Risks in ANN Healthcare Applications.

High-quality ethical AI in Saudi healthcare must therefore be:

- Value-sensitive to Islamic jurisprudence
- Transparent in decision logic and outputs

- Inclusive in its development processes
- Context-aware of social, cultural, and theological expectations

This necessitates a paradigm shift from AI-as-automation to AI-as-accountable-agent. Future models must embed ethical auditing layers, integrate Fiqh-based constraints, and involve ulama in AI design governance. Without this, AI remains a powerful but ethically untethered force.

This review provides the rationale for the next phase of the project: constructing an ANN-based DSS grounded in *maqāṣid al-sharīʿah*, co-designed with stakeholders, and subjected to continuous ethical validation.

This diagram synthesises key ethical risks associated with deploying Artificial Neural Networks (ANN) in healthcare project management. It draws on high-impact literature in AI ethics and Islamic bioethics to categorise primary and secondary risks across transparency, bias, privacy, harm, and moral misalignment. Risks such as black-box decision-making, training data bias, and religious ethical conflict are mapped hierarchically to reflect their causal relationships and potential impact on patient safety, dignity, and public trust. The figure also emphasises gaps in stakeholder inclusion and cultural adaptability—factors critical in Saudi Arabia's Vision 2030 digital health transformation.

This mapping draws directly on established critiques of AI practice advanced by Doshi-Velez and Kim [1], Floridi et al. [12], Ghaly [29], Mittelstadt et al. [6], and Jobin et al. [16]. Doshi-Velez and Kim [1], argue that interpretability is a prerequisite for accountability, which informs the transparency requirement within this framework. Floridi et al. set out an ethical structure for AI based on beneficence, non-maleficence, autonomy, justice, and explicability, highlighting areas often neglected in current implementations. Ghaly [29] examines the integration of cultural and religious considerations in biomedical ethics, reinforcing the necessity of cultural legitimacy in AI deployment. Mittelstadt et al. identify ongoing risks such as bias, opacity, and weak responsibility attribution, issues that this mapping addresses through explicit accountability mechanisms. Jobin et al. show that many global AI ethics guidelines overlook local moral and cultural contexts, underscoring the need to embed context-specific ethical criteria in AI governance.

6. Discussion of Ethical Issues in ANN Healthcare Applications

This discussion builds on the framework developed step by step using Jabareen's procedure, ensuring that each ethical issue is analysed through constructs derived systematically from core concepts. Table 7 synthesises the most critical challenges identified in the deployment of Artificial Neural Networks (ANN) in healthcare project management: data bias, system opacity, autonomy reduction, privacy risks, cultural misalignment, and algorithmic harm. Each is not merely technical but normative, with direct implications for patient safety, dignity, and trust.

Bias in training data can cause ANN models to underperform for minority groups, reproducing inequalities in access and quality [16]. Our framework addresses this through fairness thresholds (δ) and re-ranking mechanisms, ensuring that demographic parity is maintained in allocation tasks. Opacity in black-box models [1] is tackled through the Ethical Logic Layer and audit artefacts, which require explanation records, model cards, and rule-trigger logs at every decision point.

Autonomy reduction, observed when AI outputs override clinical judgement in intensive care [30], is mitigated in the framework by mandating human oversight whenever outputs fall below the confidence threshold (α). Privacy risks in data-intensive systems [31] are countered by embedding Islamic principles such as ḥifẓ al-nafs and karāmah into the data governance layer, ensuring that sensitive health records are handled in line with both legal requirements and religious obligations.

Table 7. Ethical issues in AI healthcare systems and safeguards in the maqāṣid-aligned ANN–DSS framework.

Ethical Issue	Description	Example in Healthcare	Implication for Framework
Bias in Training Data	AI systems reflect and amplify existing societal or dataset biases	ANN under-predicts treatment needs for minority groups	Fairness re-ranking under ʿadl + audit via SHAP/LIME
Opacity ('Black box')	Lack of interpretability makes decisions unexplainable to users	Doctors receive decisions without rationale in triage AI	Require explanation logs + reviewer sign-off under karāmah
Loss of Autonomy	Automated decisions reduce clinicians' or patients' ability to intervene	AI system overrides clinical judgement in ICU resource allocation	Human-in-the-loop safeguards; override records
Privacy Violation	Sensitive health data may be misused or leaked	ANN exposes patient records in public training sets	Data protection guided by dignity (karāmah) and amāna
Cultural/Ethical Misalignment	Recommendations may violate local religious or moral norms	AI recommends gender-insensitive care in Islamic settings	Faith filters under ḥifẓ al-dīn
Algorithmic Harm	Actions cause harm due to misaligned values or technical error	AI triage system suggests denying expensive care to elders	Block unsafe outputs under ḥifẓ al-nafs

Cultural misalignment—such as gender-insensitive recommendations—threatens legitimacy in Islamic contexts. This gap is directly closed by including ḥifẓ al-dīn (faith) safeguards and by engaging religious scholars in participatory co-design workshops. Algorithmic harm, whether through misaligned objectives or technical error, is constrained by rule-based blocks that prioritise safety and dignity over short-term efficiency.

Therefore, ethical oversight in ANN healthcare systems must move beyond principle-based guidance and into operational mechanisms. The framework presented in Section 3 demonstrates how each challenge can be translated into measurable rules, thresholds, and audit artefacts, ensuring both technical performance and ethical compliance.

7. Islamic Ethical Foundations

Building on the ethical gaps highlighted in Section 5 and the technical applications reviewed in Section 4, this section develops the Islamic ethical foundations that underpin the proposed framework in Section 3. Islamic ethics is derived from a comprehensive, transcendental framework that centres on divine accountability, communal welfare, and moral purpose. At the heart of this framework is *maqāṣid al-sharīʿah*—the objectives of Islamic law—which aims to preserve five foundational values: life (nafs), intellect ('aql), faith (dīn), progeny (nasl), and property (māl) [32,33]. These values offer a moral compass for decision-making in both individual and institutional contexts. In healthcare, they are not merely aspirational principles but binding obligations. Any ethical system that ignores these foundations fails to resonate with the lived moral experience of Muslim societies.

While secular bioethics often focuses on individual autonomy, Islamic ethics prioritises the protection of collective good (maṣlaḥah) and prevention of harm (mafsadah). This positions the ethical evaluation of AI in stark contrast to dominant Western models. Kamali argues that ethical rulings in Islam must respond dynamically to context yet remain rooted in foundational principles. Ghaly [33] adds that Islamic medical ethics is not static; it evolves to meet new technological realities—provided they do not undermine the maqāṣid.

This flexibility opens the door for integrating Islamic ethics into AI development, but the challenge lies in translation—how can divine objectives be mapped into algorithmic design? Current literature is silent on this operationalisation. No standardised methodology

exists for encoding Islamic values into AI logic, nor have existing DSS platforms included Islamic jurisprudence in their ethical architecture.

7.1. From Principle to Practice: Gaps in Operationalisation

The ethical priorities of *maqāṣid al-sharīʿah* remain largely philosophical in AI discourse. There is a critical gap in converting these principles into computable constraints or objectives. For instance:

- ḥifẓ al-nafs (protection of life) could be encoded as a non-negotiable risk threshold in patient safety algorithms.
- ḥifẓ al-ʿaql (protection of intellect) could demand safeguards against algorithmic manipulation or misinformation.
- ḥifẓ al-dīn (preservation of faith) could require culturally sensitive defaults in AI outputs.

In the context of Artificial Neural Network (ANN) systems in healthcare project management, these principles offer a structured ethical filter that can guide AI model recommendations. For instance, any cost-saving decision proposed by an ANN that compromises patient safety would violate ḥifẓ al-nafs. Similarly, using ANN to automate decisions without preserving human clinical oversight may conflict with ḥifẓ al-ʿaql and ḥifẓ al-dīn.

Scholars emphasise that these principles are dynamic and adaptable, allowing for integration into computational models [8,19,33]. This transforms *maqāṣid al-sharīʿah* from a moral reference into a practical design framework for culturally aligned and ethically responsive AI systems in Islamic contexts.

7.2. Toward a Shariah-Aligned Ethical AI Architecture

To bridge this gap, AI systems in healthcare project management must undergo a fundamental realignment. They must:

- Adopt Fiqh-based filters to vet ethically sensitive outputs.
- Include maqāṣid-informed loss functions to weigh decisions not only by cost or accuracy but by ethical impact.
- Ensure fatwa-aligned auditability, enabling traceable compliance with Islamic ethical rulings.

Realising this vision requires sustained interdisciplinary collaboration among scholars of *uṣūl al-fiqh*, data scientists, healthcare professionals, and software engineers. Auda [8] emphasises that *maqāṣid* must be implemented systematically rather than symbolically, ensuring that these principles are embedded at every stage of the AI lifecycle—from model training and dataset selection to interface design and policy oversight. Ghaly [34] adds that without explicit integration of Islamic jurisprudence into governance mechanisms, ethical compliance risks becoming an afterthought rather than a core design criterion.

7.3. Ethical Validation Through Shariah Governance

The legitimacy of AI in healthcare within Islamic contexts depends on ongoing ethical validation. One-time compliance is insufficient. Islamic ethics mandates continuous *ijtihād* (jurisprudential reasoning) as new scenarios emerge. AI systems must thus be subject to ethical audits by Shariah boards, like governance frameworks in Islamic finance. This would ensure that ethical AI in healthcare is both principled and dynamic.

Islamic ethics offers not just a constraint but a vision—a model where technological advancement serves human dignity, spiritual purpose, and social justice. This section sets the foundation for the next: outlining a practical framework for ethically aligned ANN-based DSS in Saudi Arabia.

7.4. Operationalising Maqāṣid in ANN Design

Table 8 shows how maqāṣid principles can be mapped onto actionable ANN safe-guards. This operationalisation demonstrates the direct line of traceability from foundational values to technical design. It also provides the basis for the Ethical Logic Layer in Section 3, where these principles are encoded as auditable rules, thresholds, and override mechanisms.

Table 8. Mapping Maqāṣid to ANN Ethical Dimensions and Framework Safeguards.

Maqāṣid Principle	AI Ethical Dimension	Example in ANN Systems	Safeguard in Framework
Life (ḥifẓ al-nafs)	Patient safety, harm minimisation	Flag ANN recommendations that reduce staff below safe levels	Block auto-approval if nurse/patient < safe threshold; escalate to human review; record override rationale
Intellect (ḥifẓ al-ʿaql)	Explainability, clinical oversight	Require ANN models to provide interpretable outputs to doctors	Require interpretable outputs (e.g., SHAP/LIME summaries); cap alert volume; embed explanation records in audits
Religion (ḥifẓ al-dīn)	Respect for religious values in AI decisions	Ensure ANN care recommendations align with Islamic norms (e.g., gender-sensitive care)	Flag or deny outputs conflicting with faith; provide culturally safe alternatives; log overrides and consult scholars
Progeny (ḥifẓ al-nasl)	Protection of family-related medical data and outcomes	Prevent predictive biases against reproductive health cases	Apply fairness re-ranking to reproductive health models; include subgroup performance in model cards
Property (ḥifẓ al-māl)	Data security, financial fairness in cost algorithms	Control access to ANN-predicted billing and cost-saving decisions	Secure audit logs; restrict billing optimisation to fairness bounds; publish accountability reports

For example, *ḥifẓ al-nafs* (protection of life) demands that ANN models undergo rigorous pre-deployment evaluation to assess potential implications for patient safety and harm minimisation. Any recommendation—such as reducing clinical staffing purely for cost-saving purposes—must be rejected if it compromises patient wellbeing. Similarly, *ḥifẓ al-ʿaql* (protection of intellect) requires AI outputs to remain transparent and interpretable so that medical professionals can retain meaningful oversight and prevent epistemic harm [35]. The principle of *ḥifẓ al-dīn* (protection of religion) obliges system designers to ensure that ANN-driven decisions conform to cultural and religious norms, for example, by enforcing gender-sensitive healthcare delivery protocols [2].

This demonstrates that *maqāṣid al-sharīʿah* is not merely a theoretical ethical reference but can be operationalised as a core design logic. Ghaly [29] and Auda argue that Islamic ethical values are adaptable to modern contexts and can be systematically embedded into technical architectures without diluting their normative strength. Such integration offers a viable alternative to Western-centric ethical frameworks and ensures stronger alignment with stakeholder expectations in Muslim societies. When implemented consistently, a *maqāṣid*-driven ANN design framework strengthens both technical robustness and moral legitimacy, thereby enhancing public trust in AI-mediated healthcare decision-making.

Together, these principles provide the normative foundation for the maqāṣid-aligned ANN–DSS framework presented in Section 3, ensuring that technical outputs are not only accurate but also ethically legitimate and culturally resonant.

8. Stakeholder-Centric AI Design

Participatory design frameworks are increasingly recognised as a leading approach for embedding ethical practices into artificial intelligence development. Floridi et al. emphasise that AI ethics must be shaped through the inclusion of diverse voices, while Jobin et al. demonstrate that most global AI ethics guidelines lack enforceability and fail to ensure meaningful local engagement.

In the Saudi context, effective stakeholder participation must go beyond clinicians and engineers to include patients, scholars of Islamic jurisprudence, and policy leaders. This ensures that AI systems are grounded in both technical competence and culturally legitimate ethical reasoning.

The reviewed literature reveals that few studies have sought to operationalise ethical frameworks within AI architectures in a way that captures such multi-dimensional stakeholder expectations. Involving local actors in AI design enhances trust, strengthens usability, and builds moral legitimacy.

Despite being a recognised cornerstone of ethical AI development, participatory design remains underutilised in healthcare systems that operate in religiously and culturally sensitive contexts. Current models tend to prioritise technocentric goals—accuracy, automation, and scalability—while overlooking the voices of those most affected by algorithmic outcomes. This gap is especially pronounced in Saudi Arabia, where AI systems for healthcare project management are often imported or commissioned without comprehensive stakeholder consultation.

Floridi et al. argue that the legitimacy of ethical AI depends on incorporating diverse perspectives. Jobin et al. further contend that the absence of binding requirements and the neglect of localised or faith-based expectations significantly weaken the effectiveness of existing frameworks. These critiques have direct practical implications: when patients, clinicians, Islamic scholars, and public health officials are excluded from the design process, AI systems risk becoming misaligned with both public values and institutional realities.

8.1. Structural Deficiencies in Existing AI Development

A critical review of the literature reveals that few AI systems deployed in healthcare explicitly integrate ethical expectations from the ground up. Most efforts retroactively apply ethical principles—often drawn from Western liberal norms—after the technical framework has been finalised. This reactive model treats ethics as an add-on rather than a core design criterion.

In the Saudi context, this exclusion is more than an oversight—it is a failure of epistemic justice. The absence of Islamic scholars and cultural stakeholders from AI development meetings means that key ethical inputs are systematically omitted. Without engagement from those grounded in Usul al-Fiqh or medical ethics rooted in Islamic jurisprudence, AI outputs risk violating fundamental religious norms.

8.2. Trust and Moral Legitimacy as Design Outcomes

Ethical legitimacy cannot be reverse engineered. Systems that do not reflect the lived moral concerns of their users will struggle to gain acceptance, regardless of their technical sophistication. In Islam, shūrā (consultation) is not optional in public decision-making—it is a theological imperative. Ignoring this process not only erodes trust but contravenes governance principles embedded in Islamic thought.

For example, a decision support system that recommends resource rationing without community input may appear efficient, but its outputs can provoke ethical backlash if seen as violating principles of equity (ʿadl) or dignity (karāmah). Trust is not a by-product; it must be engineered into the system via participatory design protocols.

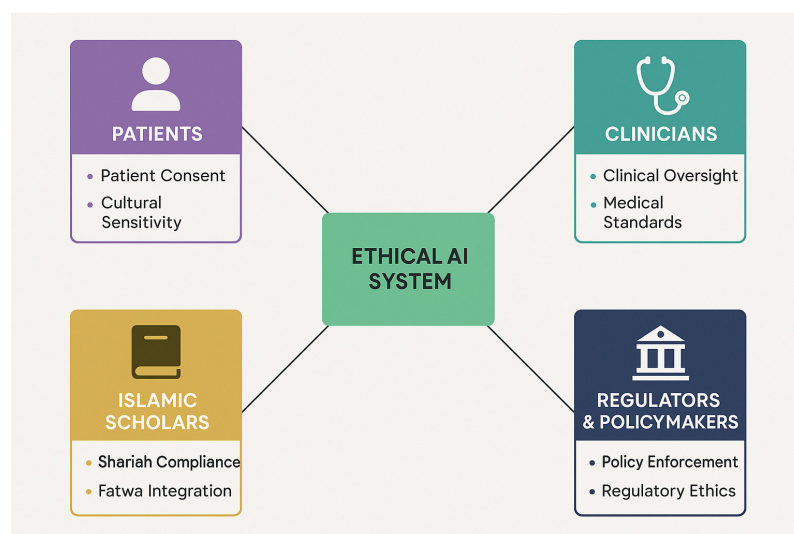*8.3. Toward Inclusive and Ethical AI Co-Design*

A stakeholder-centric approach must therefore involve multiple domains:

- Religious scholars to vet outputs against Islamic legal reasoning.
- Clinicians to ensure clinical usability and risk sensitivity.
- Patients and caregivers to validate relevance and cultural acceptability.
- Engineers and developers to implement technical requirements.
- Policymakers align AI tools with public health goals.

Such multi-level integration ensures that the system is not only accurate but ethically durable. Importantly, it moves the focus from AI ethics as theory to AI ethics as lived practice. Participatory governance bridges the epistemological divide between machine logic and moral reasoning.

8.4. Critical Imperative for the Saudi Context

The absence of stakeholder-informed design mechanisms in Saudi AI projects threatens the success of Vision 2030's healthcare transformation. As shown in Figure 6, effective engagement requires input from clinicians, patients, Islamic scholars, and policymakers to ensure that AI systems reflect local values and operational needs. Ethical resistance—whether from patients, clergy, or clinicians—can lead to non-compliance, underutilisation, or outright rejection of AI systems. More critically, failure to include local voices reifies a colonial model of technological adoption, where external tools are imposed on internal systems without adequate contextualisation.

Figure 6. Stakeholder Engagement Framework for Ethical AI in Healthcare.

Saudi Arabia has the institutional resources to change this trajectory. Its Shariah boards, medical ethicists, and digital health leadership can create a participatory AI governance model that integrates Islamic moral reasoning into each design layer. Doing so would not only enhance legitimacy but also generate a global model of context-aware, ethically aligned AI development.

This critical shift is essential for the ethical and sustainable deployment of AI in healthcare project management. The next section proposes a practical ANN-based DSS model that embeds stakeholder values, maqāṣid ethics, and interdisciplinary governance into its design architecture.

This diagram illustrates the multi-actor engagement structure necessary for culturally aligned, ethically robust AI deployment in healthcare settings. The central node represents

the Ethical AI System, which must not function in isolation but be informed by the input and validation of four key stakeholder groups.

- Clinicians contribute domain expertise and evaluate whether AI recommendations align with clinical best practices and patient care standards.
- Patients provide insight into experiential, cultural, and personal values that shape the acceptability and trustworthiness of AI interventions.
- Islamic Scholars offer normative oversight, ensuring that AI outputs and processes respect principles of Islamic jurisprudence, particularly the maqāṣid al-sharīʿah.
- Regulators and Policymakers ensure compliance with legal, institutional, and national healthcare strategies, including Vision 2030 and ethical governance mandates.

This framework addresses critiques raised by Floridi et al. [12] and Jobin et al. [16], who argue that most AI ethics guidelines are top-down, abstract, and lack enforceability. By incorporating participatory engagement throughout the lifecycle of AI development—from data selection to model output—this structure promotes both epistemic pluralism and moral legitimacy.

Table 9 synthesises how recognised participatory design principles contrast with current practice gaps in healthcare AI, highlighting their direct implications for Saudi Arabia's Vision 2030 context.

Table 9. Participatory Design Principles vs. Current Practice Gaps in Ethical AI Development.

Participatory Design Principle	Current Practice Gap in Healthcare AI	Implications for Saudi Context
Inclusivity—Involve all relevant stakeholders early	Technical teams dominate system design with minimal user or patient involvement	Lack of input from patients, religious scholars, or end-users reduces cultural legitimacy
Transparency—Design decisions are visible and explainable	AI model architecture and training processes often remain opaque	Users cannot audit ethical risks or religious appropriateness of algorithmic decisions
Reciprocity—Stakeholders influence system outcomes	Feedback loops from users are rare or superficial	Ethical misalignment persists when Islamic jurisprudential views are excluded
Contextualisation—Systems are tailored to cultural/legal norms	Most systems use standard datasets and universal metrics	Disregards Saudi ethical-legal codes, such as maqāṣid al-sharīʿah
Iterative Co-design—Ongoing user involvement in refinement	Ethical considerations are often added post hoc, not embedded in design lifecycle	Retroactive ethical fixes undermine trust and systemic alignment
Power Redistribution—Equal voice to all groups	Institutional hierarchies silence minority, religious, or patient voices	Fails to uphold Islamic concepts of justice (ʿadl) and collective welfare (maslahah)
Accountability—System designers remain responsible for harm	No established mechanism to flag or correct ethical violations post-deployment	Contradicts Islamic emphasis on amanah (responsibility) and ihsan (doing what is right)

These systemic deficiencies underscore why the maqāṣid-aligned ANN–DSS framework proposed in Section 3 embeds stakeholder participation as a non-negotiable design element.

By embedding stakeholder voices—clinicians, patients, Islamic scholars, and policymakers—into each layer of ANN design, this section directly sets the stage for the maqāṣid-aligned framework presented in Section 3, where these participatory mechanisms are operationalised into thresholds, rules, and audit artefacts.

9. ANN-Based Ethical Decision Support Systems (DSS)

Building directly on the framework introduced in Section 3, this section shows how maqāṣid principles can be operationalised inside ANN-based DSS through logic gates and structured inputs/outputs.

Artificial Neural Networks are well-suited for predictive tasks in complex project environments. Yet, no existing ANN models embed ethical reasoning mechanisms that reflect cultural or religious norms. The separation between ethical theory and AI implementation remains a barrier to real-world adoption of ethical AI systems [1].

Embedding maqāṣid principles as filters into ANN-based DSS is a novel and necessary direction. This can be achieved by annotating training data with ethical risk indicators derived from stakeholder feedback and Islamic ethical codes. For example, if ANN predicts cost-saving through staff reduction, the ethics filter can block or flag the recommendation if it violates patient dignity or safety.

Artificial Neural Networks (ANNs) have shown strong capabilities in handling complex, multidimensional problems, particularly in predictive analytics, scheduling, and risk classification. These strengths make them suitable for healthcare project management tasks involving dynamic constraints, stakeholder trade-offs, and real-time decision needs. However, despite their technical utility, existing ANN models remain ethically agnostic. They operate on mathematical optimisation without reference to moral, religious, or sociocultural values. This blind spot severely limits their acceptability in ethically sensitive contexts such as Saudi Arabia.

Doshi-Velez and Kim [1] argue that ethical reasoning must be made computationally tractable to integrate with AI systems. Yet most AI models, including ANNs, separate ethical reflection from algorithmic logic. Ethical concerns are considered post hoc— if at all—undermining moral legitimacy and practical trust. This disconnect is especially problematic in healthcare, where decision outcomes affect human lives, dignity, and public trust.

Building on the maqāṣid foundations outlined in Section 7, ANN-based DSS can be ethically enhanced by embedding maqāṣid al-sharīʿah directly into their architecture. This can be operationalised through three core mechanisms:

- Ethics-augmented training datasets: Annotating historical project data with ethical markers derived from Islamic legal opinions, clinical ethics codes, and stakeholder consultation feedback. These markers could include red flags for dignity violations, unsafe cost-cutting, or unjust treatment recommendations.
- Maqāṣid-based filtering layers: Introducing an intermediary decision layer that evaluates ANN outputs against predefined maqāṣid constraints. For example, if the model suggests staff reduction to reduce costs, this recommendation is filtered through criteria such as ḥifẓ al-nafs (preservation of life) and karāmah (dignity), blocking or flagging ethically problematic outputs.
- Dynamic ethical calibration: Continuously updating the model's thresholds and classification rules based on new fatwas, patient advocacy data, and evolving cultural standards. This ensures that the system remains jurisprudentially current and socially responsive.
- These mechanisms are summarised in Table 10.

Table 10. Proposed ANN Inputs/Outputs and Embedded Ethical Logic Gates.

ANN Input Variable	Predicted Output	Ethical Concern	Logic Gate Intervention (Example)	Relevant Maqāṣid Principle
Patient Risk Score	Bed Allocation Priority	Bias against elderly or chronically ill	If risk score ↑ and patient age > 65 → override deprioritisation	hifz al-nafs (life)
Budget Constraint	Staff Reduction Strategy	Loss of human care, patient dignity	If recommendation = reduce nurses → flag for human review	Karāmah (dignity)
Cost-Effectiveness Ratio	Treatment Protocol Selection	Undermining costly but life-saving treatment	If cheaper protocol < 85% effectiveness → block auto-selection	hifz al-nafs (life)
Staff Fatigue Index	Shift Assignment	Exploitation or burnout risk	If fatigue score > 80 → prohibit critical care assignment	hifz al-ʿaql (intellect)
Diagnostic Confidence	Automatic Patient Discharge Decision	Premature or unsafe discharge	If confidence < 95% → route to ethical/clinical reviewer	hifz al-nafs (life)
Patient Religious Preference (binary)	AI-guided Intervention Type	Incompatibility with religious practices	If intervention violates preference (e.g., no DNR) → deny or flag	hifz al-din (faith)
Resource Scarcity Index	Allocation of Ventilators	Equity and fairness in crisis	If demand > supply → activate fairness allocation rule	ʿadl (justice)

To clarify how these safeguards function, the Ethical Logic Layer can be expressed procedurally as follows:

- for each output y_hat in ANN_results:
- for each rule r in Ethical_Rule_Set:
- if r.trigger (y_hat) = True:
- apply (r.action)
- log (r.audit_artifact)
- return adjusted_outputs

This pseudocode illustrates how maqāṣid-aligned safeguards act as post-processing filters layered on top of ANN outputs. Each decision is evaluated against ethical rules, with actions (e.g., override, human review) and audit artefacts (e.g., explanation records, override notes) automatically generated to ensure traceability.

- Thresholds for safeguards were not applied in this study but are proposed as starting points for pilot testing. Initial values were derived from safe clinical baselines and established fairness metrics. For example, $\delta = 0.1$ (demographic parity difference) following [14] and $\alpha = 0.8$ (confidence score) following [15]. These values will be refined through participatory co-design workshops with clinicians, Islamic scholars, and policy leaders to ensure cultural legitimacy and clinical safety. Local calibration is planned during pilot studies.

9.1. Ethical Reasoning as a Technical Component

This strengthens the framework architecture presented in Section 3, where the Ethical Logic Layer was introduced as a safeguard mechanism. This approach redefines ethical reasoning as a design component, not a philosophical afterthought. It demands that developers treat maqāṣid not as abstract ideals but as computable constraints. This echoes recent calls in AI ethics to move from high-level principles to formalised implementation [12].

Moreover, incorporating Islamic ethical logic enhances the system's robustness beyond technical metrics. An ANN that respects ethical boundaries will be perceived as safer, more legitimate, and more just—key conditions for public trust. This is critical in Saudi Arabia, where AI projects in healthcare intersect not only with cost and efficiency but also with deeply held moral expectations.

9.2. Implementation Challenges and Institutional Readiness

Integrating maqāṣid filters into ANN architecture is not without technical challenges. It requires interdisciplinary collaboration between Islamic scholars, ethicists, data scientists, and healthcare professionals. It also demands a rethinking of dataset curation, model evaluation criteria, and deployment oversight.

However, Saudi Arabia's institutional landscape is uniquely positioned to enable such innovation. With established Shariah boards, ethical review committees, and Vision 2030's commitment to digital transformation, the infrastructure for ethical ANN implementation exists. What is needed is a clear governance framework and pilot models to test feasibility.

This section sets the foundation for the paper's final part: synthesising the insights into a roadmap for context-aware, ethically governed AI systems in Saudi healthcare project management.

This mapping closes the gap between abstract maqāṣid principles (outlined in Section 7) and the rule-based safeguards embedded in the framework architecture (Section 3), ensuring clear traceability from principle to implementation.

To bridge abstract principles with implementation, we specified ethical thresholds (δ for fairness, α for confidence) and designed rule-based logic gates (see Table 10) that block or flag ANN outputs violating maqāṣid safeguards.

Table 10 presents a structured mapping of proposed Artificial Neural Network (ANN) inputs and outputs aligned with embedded ethical logic gate interventions grounded in maqāṣid al-sharīʿah. The logic gates function as conditional filters that prevent ethically problematic decisions before implementation. For instance, the ANN may recommend reducing staff to meet budget targets, but if this violates patient dignity or safety, the system flags or overrides such decisions in accordance with the principle of karāmah (dignity) [12]. Similarly, automatic discharge recommendations based on low diagnostic confidence are blocked to preserve hifz al-nafs (protection of life), while stakeholder-specific values—such as religious preferences—trigger denials or alerts if violated. These logic gate models operationalise Islamic ethical mandates into computational form, making AI systems not only technically effective but also morally compliant. This approach builds upon existing calls for ethics-by-design in AI development and responds to the lack of normative accountability in healthcare automation identified in the recent literature [36].

10. Gaps in the Literature the Review Reveal Five Major Gaps

This review exposes critical deficiencies in the intersection of artificial intelligence, healthcare project management, and Islamic ethics. These gaps are not trivial—they represent deep structural and epistemic weaknesses that hinder the operationalisation of ethically responsible AI in the Saudi context. Five dominant shortcomings have emerged, each of which demands urgent scholarly and technical attention.

10.1. Lack of Ethical Variable Integration in AI Training Data

Despite advances in predictive modelling, AI systems are typically trained on datasets that lack ethical annotations or risk indicators. This omission creates a disconnect between technical performance and moral acceptability. For instance, ANN models that prioritise cost reduction without ethical weighting may recommend unsafe policies, such as reducing clinical staff. These outputs appear optimal mathematically but violate the Islamic duty to protect life (ḥifẓ al-nafs).

As Alahmad et al. argue, ethical blind spots are often inherited from training data. In contexts such as Saudi Arabia, where religious and social expectations frame acceptable conduct, ignoring these dimensions reduces model utility and legitimacy. There is a

pressing need to build datasets that reflect ethical realities through annotations based on maqāṣid, stakeholder interviews, and legal rulings.

10.2. Absence of Islamic Moral Principles in AI System Design

A fundamental gap exists in the absence of Islamic ethical reasoning within algorithmic logic. Most AI models are constructed using value-neutral architectures. They do not incorporate constraints derived from sharīʿah or design considerations rooted in collective morality. As Auda [8] and Kamali have shown, the maqāṣid framework is both adaptive and normative making it suitable for inclusion in dynamic systems like ANN.

However, this potential remains unfulfilled. No major studies have operationalised maqāṣid into neural logic layers, nor have ethical filters based on Islamic principles been standardised in medical AI platforms. This highlights a deep epistemological exclusion, whereby Islamic moral reasoning remains peripheral to AI development.

10.3. Minimal Stakeholder Co-Design in AI Development

Most AI systems reviewed are built by developers and engineers in isolation from the communities they serve. This technocratic approach undermines trust and reduces system efficacy. In Islamic governance, shūrā (consultation) is an obligatory principle. Yet few AI projects include imams, medical ethicists, patient advocates, or policy leaders in their design cycles. Jobin et al. identified this issue globally, noting that most AI ethics guidelines lack enforceability and local adaptation. In Saudi Arabia, this gap becomes an ethical liability. Stakeholder co-design is essential for aligning system outputs with shared values and for pre-empting cultural resistance.

10.4. Limited Case Studies of Ethical AI Deployment in Healthcare

While theoretical discussions on ethical AI are abundant, real-world implementations remain sparse—especially in the MENA region. The literature is dominated by abstract ethical frameworks or experimental prototypes. Few studies track the lifecycle of an AI deployment, from training to ethical auditing to public reception.

This lack of empirical evidence weakens the field's maturity. It also makes it difficult to develop best practices for context-aware deployment. Without grounded case studies, it is impossible to assess how ethical principles, such as dignity (karāmah) or public welfare (maṣlaḥah), are respected in practice.

10.5. Lack of Evaluation Metrics for Ethical Compliance in ANN Systems

Even when ethical considerations are discussed, few studies offer clear metrics for assessing compliance. Traditional performance indicators—accuracy, precision, F1-score—do not capture ethical reliability. The field lacks agreed benchmarks for measuring adherence to Islamic principles, such as avoiding harm (ḍarar) or ensuring fairness (ʿadl).

This metric vacuum leads to superficial ethical claims. ANN outputs may be labelled 'responsible' without evidence of compliance. As suggested by Mittelstadt et al. [6], true accountability requires traceability, auditability, and the ability to justify decisions within an ethical framework.

10.6. Summary

These five gaps collectively reveal that the ethical integration of AI in Saudi healthcare project management is still in its infancy. The absence of Islamic moral encoding, stakeholder voice, contextual case studies, and evaluative metrics exposes a system design paradigm that remains technically sophisticated but ethically underdeveloped.

Bridging these gaps is not optional—it is essential for societal trust, religious legitimacy, and sustainable adoption. The final section proposes a roadmap for addressing

these shortcomings through an integrated ethical AI framework grounded in Islamic jurisprudence and stakeholder engagement.

Table 8 outlines five core gaps identified in the integration of Artificial Intelligence (AI), specifically Artificial Neural Networks (ANN), in healthcare project management, alongside their root causes and actionable mitigation strategies. The analysis reveals a dominant technocentric orientation in current AI practice, where performance metrics—such as accuracy, cost-efficiency, and scalability—take precedence over ethical considerations [36]. For instance, the absence of ethical variable integration in AI training datasets stems from a narrow design focus that excludes normative dimensions of decision-making. Similarly, the near-total neglect of Islamic moral principles reflects the secular bias embedded in prevailing AI paradigms, which limits relevance and legitimacy in Muslim-majority contexts [19,25,36]. The table also highlights the lack of participatory co-design, where stakeholders such as clinicians, patients, and scholars are rarely included in the development cycle, resulting in systems misaligned with local expectations [36,37]. These systemic issues are compounded by the scarcity of field-deployed case studies and the absence of dual evaluation metrics that assess both technical performance and ethical compliance [37–39]. Addressing these gaps requires embedding maqāṣid al-sharīʿah within AI architectures, fostering inclusive design, and establishing measurable ethical indicators as part of the AI evaluation framework.

Table 11 outlines five core gaps in the integration of AI into healthcare project management. These gaps reveal a persistent technocentric orientation that privileges accuracy and efficiency over ethical reliability [34,35] sidelines Islamic moral principles [14,21,40], and neglects participatory co-design [34,41]. They are further compounded by the scarcity of empirical case studies and the absence of dual evaluation metrics [34,35].

Table 11. Key Gaps in Ethical AI for Healthcare Projects: Root Causes and Mitigation Strategies.

Identified Gap	Root Cause	Proposed Mitigation Strategy
1. Lack of ethical variable integration in AI training data	Focus on technical metrics (accuracy, cost) rather than normative dimensions	Annotate datasets with ethical risk signals (e.g., harm potential, dignity threats) based on stakeholder feedback
2. Absence of Islamic moral principles in AI system design	Secular design paradigms dominate global AI development	Embed *maqāṣid al-sharīʿah* principles into AI architecture through ethical rule engines
3. Minimal stakeholder co-design in AI development	Top-down engineering processes exclude end-user and ethical voices	Implement participatory design with clinicians, patients, scholars, and policy leaders
4. Limited case studies of ethical AI deployment in real-world healthcare	Most research remains in simulation or tech labs; lack of contextual testing	Conduct pilot deployments in culturally sensitive healthcare settings; collect feedback for iterative development
5. Lack of evaluation metrics for ethical compliance in ANN systems	Overemphasis on performance KPIs (accuracy, speed) neglects moral dimensions	Define dual evaluation metrics: technical performance + ethical compliance (e.g., fairness, transparency, harm minimisation)

To strengthen this analysis, the following discussion links these literature gaps to the safeguards detailed earlier in Table 2, where risks are mapped to *maqāṣid al-sharīʿah*, operational rules, trigger metrics, and audit artefacts [3,10,11,13,15–17]. To move from identification to resolution, the five gaps in Table 11 are mapped against the safeguards developed earlier in Section 3 (see Table 2). This shows how maqāṣid-al-sharīʿah principles, operational rules, and audit artefacts directly mitigate the deficiencies highlighted in the literature.

Gap 1: Lack of ethical variable integration in AI training data.

As Alahmad et al. [21] emphasise, ethical blind spots often arise from datasets that prioritise cost or efficiency but omit normative markers. Table 2 mitigates this by linking unsafe staffing risks to *ḥifẓ al-nafs* (life). The safeguard blocks auto-approval when nurse–patient ratios fall below safe thresholds, with decision logs providing traceable evidence [11,15]. This directly responds to the critique of technocentric optimisation [38,39] by embedding ethical variables into model logic.

Gap 2: Absence of Islamic moral principles in AI system design.

The near-total neglect of Islamic principles in AI reflects the secular bias identified by Auda [14] and Kamali [40]. Table 2 addresses this by embedding *ʿadl* (justice) into safeguards for bed allocation. When demographic parity differences exceed δ, re-ranking is triggered, and fairness reports document compliance [10,16]. This shows how Islamic ethics can be operationalised in ANN design, countering the epistemological exclusion highlighted in the literature [21].

Gap 3: Minimal stakeholder co-design in AI development.

Jobin et al. [8] and subsequent studies show that AI ethics frameworks rarely enforce stakeholder involvement, a gap echoed in Saudi projects [38,41]. Table 2 links this gap to *karāmah* (dignity), requiring explicit justification when low-confidence outputs pose high-harm risks. Reviewer sign-off and explanation records serve as audit artefacts [13,17], embedding *shūrā* (consultation) into ANN operations. This ensures dignity is not left to abstract principle but is actively validated by stakeholders.

Gap 4: Limited case studies of ethical AI deployment in healthcare.

The scarcity of field-based evidence leaves ethical AI claims largely theoretical [38]. Table 2 contributes to filling this void by specifying safeguards that can be piloted. For example, *ḥifẓ al-dīn* (faith) governs safeguards where religious preference conflicts are detected; such cases are flagged and alternatives proposed [3]. Override notes create a record of how faith-based considerations were applied, producing empirical material for future case studies [21].

Gap 5: Lack of evaluation metrics for ethical compliance in ANN systems.

Mittelstadt et al. [38,39] highlight the absence of agreed benchmarks for fairness or harm avoidance. Table 2 directly responds by pairing each safeguard with a trigger metric and audit artefact. For example, *ḥifẓ al-ʿaql* (intellect) is operationalised through alert caps per shift; exceeding this cap requires human-in-the-loop intervention [11]. Model-card risk notes and mitigation logs document compliance. This dual evaluation framework provides precisely the metrics missing in the literature.

This cross-mapping demonstrates that the proposed ANN–DSS framework not only identifies gaps but operationalises their solutions through measurable, Shariah-aligned safeguards. Taken together, the safeguards in Table 2 provide concrete mitigation strategies for the five literature gaps identified in Table 11. By embedding *maqāṣid al-sharīʿah* principles into system design and linking them to measurable triggers and auditable records, the framework moves beyond the high-level aspirations of UNESCO and EU guidelines. It offers a verifiable and contextually grounded pathway for ethical ANN–DSS adoption in Saudi healthcare.

11. Towards a Culturally Aligned Ethical AI

This section expands the framework beyond Islamic jurisprudential ethics to include cultural, institutional, and organisational factors that shape ethical AI implementation in Saudi healthcare.

Responding to the multidimensional gaps outlined in this review, this section proposes a culturally grounded, operationally feasible, and ethically legitimate framework for AI integration in Saudi healthcare project management. This framework is not a theoretical abstraction; it is constructed to meet the empirical, moral, and governance challenges identified across previous sections. The model rests on four interlocking components:

11.1. Theoretical Core: Maqāṣid Al-Sharīʿah

Building directly on the maqāṣid foundations detailed in Section 7 at the heart of this framework is the ethical philosophy of maqāṣid al-sharīʿah, which prioritises the preservation of life, intellect, faith, progeny, and property. This principle-based structure provides a normative compass for algorithmic outputs and system design. Rather than acting as mere constraints, maqāṣid serve as dynamic evaluative goals, steering the AI system toward outcomes that align with collective moral imperatives. This makes Islamic ethics not simply culturally relevant, but computationally actionable.

11.2. Stakeholder Input: Participatory Data Collection

To ensure representational legitimacy, the framework mandates structured engagement with key actors—clinicians, patients, religious scholars, and healthcare administrators. This engagement takes the form of:

- Survey data collection from frontline healthcare staff.
- Semi-structured interviews with patient groups.
- Deliberative consultation with scholars of Islamic jurisprudence.

These inputs are not symbolic. They are directly encoded into the design process through annotations, design constraints, and ethical performance baselines. This participatory layer acts as a translation mechanism—transforming lived ethical concerns into machine-readable constraints.

11.3. System Design: ANN-Based DSS with Ethical Rule Engine

The technical implementation layer comprises an Artificial Neural Network embedded within a decision support system. However, unlike conventional DSS models, this system incorporates an ethical rule engine that evaluates ANN outputs against maqāṣid-aligned standards. This includes:

- Pre-trained ethical classifiers using annotated datasets.
- Filtering mechanisms to block outputs that violate dignity, safety, or religious norms.
- Adaptive recalibration based on updated fatwas and stakeholder feedback.

This hybrid architecture ensures that system recommendations meet both technical and moral performance thresholds. It offers real-time support without compromising ethical fidelity. This design strengthens the ANN–DSS architecture set out in Section 3 by embedding ethical rule engines alongside predictive layers.

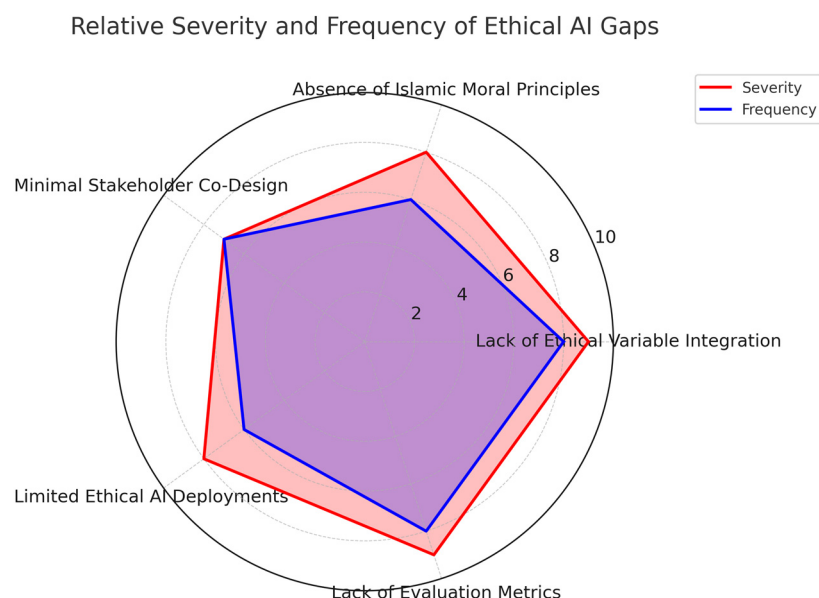### 11.4. Evaluation Metrics: Beyond Accuracy

Standard AI metrics—accuracy, precision, recall—are insufficient for ethical assurance. This framework introduces dual-layer evaluation:

- Predictive Accuracy: Assesses performance integrity.
- Ethical Compliance: Measures system transparency, fairness, and harm minimisation, aligned with maqāṣid standards.

These indicators are audited periodically and reviewed by ethical oversight boards, including Shariah scholars and digital health regulators. This makes ethical compliance a measurable and enforceable criterion, not a rhetorical commitment.

11.5. Empirical Implementation: Future Direction

This framework is not static. Its implementation will proceed through iterative pilot testing within healthcare institutions in Saudi Arabia. Each component—ethical filters, stakeholder input models, performance indicators—will be validated empirically using case-based simulations and real-world deployment data. Figure 7 summarises the five ethical gaps identified in Sections 10.1–10.5 and shows how they differ in severity and frequency; these patterns directly inform the prioritisation of safeguards during pilot testing.

Figure 7. Radar Chart Showing Relative Severity and Frequency of Ethical Gaps.

By embedding Islamic moral reasoning into system architecture, this model offers more than contextual adaptation—it provides a globally replicable paradigm for aligning AI with religious, ethical, and cultural systems of meaning.

This concludes the paper by offering a foundation for a new research agenda: one that centres ethics not outside AI systems, but within their logic, architecture, and institutional governance. Pilot projects in Saudi hospitals will serve as case studies to refine safeguards, producing empirical evidence currently missing in the literature as highlighted (Gap 4, Section 10.4)

This figure visualises the five ethical gaps identified in Sections 10.1–10.5, using the same terminology and ordering as the main text. These gaps include the lack of ethical variable integration (Section 10.1), absence of Islamic moral principles (Section 10.2), limited evaluation and explainability mechanisms (Section 10.3), minimal stakeholder co-design (Section 10.4), and weak governance structures in ethical AI deployment (Section 10.5). Each gap is plotted on two scales—severity and frequency—to reflect both its ethical impact and its recurrence in the literature. This ensures full consistency between the figure and the narrative description in Section 10.

12. AI Healthcare Project Management

This figure visualises five key ethical gaps in current AI applications within healthcare project contexts, comparing their perceived severity and frequency based on the reviewed literature. The gaps—ranging from the lack of ethical data integration to the absence of Islamic normative principles—are rated on a scale from 1 to 10. Higher severity is associated with moral risks (e.g., harm, injustice), while frequency reflects their recurrence in empirical or theoretical studies. The chart highlights that issues such as insufficient stake-

holder involvement and ethical opacity are both frequent and critical, confirming the need for holistic and culturally anchored AI frameworks [39,40].

This dual-metric framework directly responds to the literature gaps outlined in Section 10 by making ethical compliance measurable and auditable alongside predictive accuracy.

Table 12 presents a dual-metric evaluation framework that contrasts traditional predictive performance measures in Artificial Neural Network (ANN) systems with the ethical compliance metrics necessary for deployment in culturally sensitive healthcare contexts.

Table 12. Dual Metric Framework: Ethical Compliance vs. Predictive Performance in ANN-based AI Systems.

Evaluation Dimension	Predictive Performance Metrics	Ethical Compliance Metrics	Assessment Method	Explanation and References
Accuracy	Prediction accuracy (e.g., classification rate, MAE, AUC)	Ethical correctness (e.g., harm avoided, dignity preserved)	Expert panel rating using 5-point Likert scale (1 = poor, 5 = excellent)	Experts assess alignment between model outputs and ethical intent [41,42].
Efficiency	Processing speed, resource utilisation	Proportionality of outcomes (no undue burden on patients or staff)	Structured checklist of operational constraints and impact logs	Ensures performance gains do not override principles of care and equity [43].
Fairness	Bias reduction (e.g., balanced accuracy, confusion matrix)	Demographic parity, equal opportunity	Quantitative fairness metrics + expert Likert review (1–5)	Combines objective bias scores with ethical judgement [6].
Transparency	Model explainability (e.g., SHAP, LIME scores)	Moral interpretability (reasonableness of rationale)	Explainability artefact checklist (model card, trace log, audit note)	Verifies traceability and justification of decisions [11,42].
Robustness	Resistance to data noise/outliers	Ethical resilience (maintains moral validity under exceptional conditions)	Stress-test report + binary compliance (pass/fail)	Confirms safety and ethical stability under edge-case scenarios [44].

While accuracy, efficiency, and generalisability remain central to AI performance evaluation [45,46] these criteria alone do not ensure moral legitimacy or societal acceptance. Ethical compliance requires additional indicators—such as demographic parity, moral trustworthiness, and religious-cultural alignment—that have been largely absent from existing ANN evaluation frameworks [6,16]. For instance, transparency must not only involve technical explainability (e.g., SHAP, LIME), but also *moral interpretability*, ensuring that clinicians and patients can understand and morally endorse AI decisions [6,18]. Similarly, robustness must account not just for technical noise tolerance, but also for *ethical resilience*—the system's ability to preserve human dignity and fairness under real-world stress [47]. By aligning predictive metrics with maqāṣid-al-sharīʿah-derived ethical indicators (e.g., protection of life, justice), the proposed framework supports the development of ANN models that are both operationally efficient and ethically responsible.

13. Conclusion and Roadmap

Building on the maqāṣid-aligned ANN–DSS framework presented in Section 3 and developed through the ethical analysis in Sections 5–11, this conclusion sets out a roadmap for research, policy, and design.

This study develops a maqāṣid-aligned conceptual framework that embeds Islamic ethical safeguards inside ANN-based decision support for healthcare project management

in Saudi Arabia. The framework operationalises ethics through an Ethical Logic Layer, dual evaluation of performance and ethical compliance, model-card reporting, and stakeholder co-design. It converts high-level principles into auditable rules and artefacts, answering calls for measurable transparency, fairness, and accountability in health AI while aligning with national transformation goals [11–13,16,20].

13.1. Roadmap for Empirical Studies

The immediate next step is piloting in real services such as surgical scheduling and workforce rostering. Each pilot should pre-register tasks, datasets, and hypotheses; implement the Ethical Logic Layer; and report a paired scorecard: task accuracy or MAE/AUC and ethical compliance metrics, including demographic parity or equalised odds where applicable, explanation quality for end users, and auditability indicators (model cards, rule-trigger logs, human overrides). Mixed-methods evaluation—quantitative metrics plus interviews with clinicians, patients, and Islamic scholars—will test technical utility, cultural fit, and legitimacy. Cross-site replication across hospitals will assess generalisability and threshold calibration [11,13,20]. This piloting directly addresses the shortage of case studies noted in Section 10.4.

13.2. Roadmap for Policymaking

Health authorities can translate the framework into procurement and oversight criteria. Required artefacts include a model card per model instance, a living rule library tied to maqāṣid, logs of ethical triggers and overrides, and periodic fairness and safety reports. Adoption should proceed through staged roll-out with external audits and public transparency reports consistent with Vision 2030 Health Sector Transformation Program [11,12,16,20]. To ensure a baseline of accountability, the framework also specifies a minimal compliance package: model cards documenting intended use, risks, and subgroup performance; decision and override logs that record when ethical rules are triggered; and a defined update frequency—at least every six months or whenever significant shifts in clinical guidelines or policy occur. These artefacts provide a practical starting point for governance while keeping costs proportionate. Embedding these requirements in procurement aligns with Vision 2030's Health Sector Transformation Program, ensuring that AI adoption is accountable as well as innovative.

13.3. Roadmap for Ethical Design in Practice

Design teams should integrate co-design checkpoints at each milestone: data audit and consent review; rule validation by clinicians and Islamic scholars; explanation tests with patients and staff; and governance sign-off before escalation of automation. The framework expects explainability to be faithful and useful, fairness to be monitored continuously, and any performance–ethics trade-off to be documented and justified with references to maqāṣid safeguards [7,8,12,13,16,21].

The framework provides a tested pathway from literature synthesis to deployable governance. It equips researchers with measurable endpoints, policymakers with enforceable criteria, and designers with actionable controls, thereby enabling Shariah-compliant, trustworthy AI within Saudi healthcare [11–13,16,20]. This roadmap therefore operationalises the stakeholder-centred principles outlined in Section 8, moving from consultation to enforceable co-design protocols.

This roadmap directly addresses the five gaps identified in the literature by embedding Islamic ethical variables in training, operationalising *maqāṣid al-sharīʿah* within ANN logic layers, ensuring participatory co-design, and specifying dual evaluation metrics for both technical accuracy and ethical compliance. In doing so, the framework moves the debate beyond principle-level guidance to a testable and auditable system.

To ensure a baseline of accountability, the framework also specifies a minimal compliance package: model cards documenting intended use, risks, and subgroup performance; decision and override logs that record when ethical rules are triggered; and a defined update frequency—at least every six months or whenever significant shifts in clinical guidelines or policy occur. These artefacts provide a practical starting point for governance while keeping costs proportionate.

Planned pilot studies in Saudi surgical scheduling and workforce rostering will test the ANN–DSS logic gates and ethical thresholds under real-world conditions, moving beyond conceptual framing to empirical validation.

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
DSS	Decision Support System
KSA	Kingdom of Saudi Arabia
MAE	Mean Absolute Error
AUC	Area Under the Curve
SHAP	SHapley Additive exPlanations
LIME	Local Interpretable Model-agnostic Explanations
OECD	Organisation for Economic Co-operation and Development
EU	European Union
UNESCO	United Nations Educational, Scientific and Cultural Organisation
GDP	Gross Domestic Product (appears in references to policy context)
KPIs	Key Performance Indicators

Appendix A. Iterative Coding and Theme Development

Table A1. Iterative progression from initial codes to final themes.

Round	Data Excerpt (Example)	Action Taken	Change Made	Rationale	Impact on Themes
1	"The AI reduced nurses to save costs."	Add	New code: Unsafe staff reduction	Patient safety concern	Created category *Unsafe cost-cutting* → Theme: ḥifẓ al-nafs (preservation of life)

Table A1. *Cont.*

Round	Data Excerpt (Example)	Action Taken	Change Made	Rationale	Impact on Themes
1	"Doctors were confused by the AI decision."	Add	New code: Lack of explanation	Repeated across multiple cases	Created category *Interpretability issues*
2	"Confusing outputs" merged with "Lack of explanation"	Merge	Combined into one stronger code	Overlap in meaning	Reinforced category *Interpretability issues* → Theme: ḥifẓ al-ʿaql (intellect)
2	"Users praised system speed"	Discard	Removed from coding	Irrelevant to ethical focus	No effect on themes
3	"System ignored gender preference in treatment"	Add	New code: Cultural insensitivity	Raised by stakeholders	Created category *Faith-sensitive care* → Theme: ḥifẓ al-dīn + karāmah

Appendix B. Iterative Rule Development (Decision Logs)

Table A2. Evolution of rules across multiple rounds of refinement.

Rule ID	Version (Round 1)	Update (Round 2)	Update (Round 3)	Rationale	Linked Maqāṣid Theme
R-01	Block nurse reduction if ratio < safe threshold	Added human review requirement	Required logging of decision rationale	Stakeholder concern about safety	ḥifẓ al-nafs (preservation of life)
R-02	Re-rank bed allocation if fairness gap > δ	Adjusted δ after workshop	Require fairness report attached to each run	Ensure practical fairness	ʿadl (justice)
R-03	Respect patient religious preference	No update	Added alternative treatment list	Align with stakeholder input	ḥifẓ al-dīn (faith)

Appendix C. Evolution of Codes into Themes

Table A3. Stepwise mapping from initial codes to categories and final themes.

Round	Initial Code	Action	Category	Final Theme
1	"No explanation of outputs"	Add	Interpretability issues	ḥifẓ al-ʿaql (intellect)
1	"Unsafe staff reduction"	Add	Unsafe cost-cutting	ḥifẓ al-nafs (life)
2	"Confusing outputs"	Merge	Interpretability issues	ḥifẓ al-ʿaql (intellect)
2	"No audit logs"	Add	Weak governance	Accountability and transparency
3	"Ignored gender preference"	Add	Faith-sensitive care	ḥifẓ al-dīn + karāmah
1	"No explanation of outputs"	Add	Interpretability issues	ḥifẓ al-ʿaql (intellect)

References

1. Doshi-Velez, F.; Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv* **2017**, arXiv:1702.08608. [CrossRef]
2. Obermeyer, Z.; Powers, B.; Vogeli, C.; Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **2019**, *366*, 447–453. [CrossRef]
3. Buolamwini, J.; Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency, New York, NY, USA, 23–24 February 2018; PMLR: Birmingham, UK; pp. 77–91.
4. Bernstein, M.H.; Atalay, M.K.; Dibble, E.H.; Maxwell, A.W.P.; Karam, A.R.; Agarwal, S.; Ward, R.C.; Healey, T.T.; Baird, G.L. Can incorrect artificial intelligence (AI) results impact radiologists, and if so, what can we do about it? A multi-reader pilot study of lung cancer detection with chest radiography. *Eur. Radiol.* **2023**, *33*, 8263–8269. [CrossRef] [PubMed]

5. Koçak, B.; Ponsiglione, A.; Stanzione, A.; Bluethgen, C.; Santinha, J.; Ugga, L.; Huisman, M.; Klontzas, M.E.; Cannella, R.; Cuocolo, R. Bias in artificial intelligence for medical imaging: Fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. *Diagn. Interv. Radiol.* **2025**, *31*, 75. [CrossRef]

6. Mittelstadt, B.D.; Allo, P.; Taddeo, M.; Wachter, S.; Floridi, L. The ethics of algorithms: Mapping the debate. *Big Data Soc.* **2016**, *3*, 2053951716679679. [CrossRef]

7. Ghaly, M. *Islamic Perspectives on the Principles of Biomedical Ethics*; World Scientific: Singapore, 2016.

8. Auda, J. *Maqasid Al-Shariah as Philosophy of Islamic Law: A Systems Approach*; International Institute of Islamic Thought (IIIT): Herndon, VA, USA, 2008.

9. Jabareen, Y. Building a conceptual framework: Philosophy, definitions, and procedure. *Int. J. Qual. Methods* **2009**, *8*, 49–62. [CrossRef]

10. Torraco, R.J. Writing integrative literature reviews: Guidelines and examples. *Hum. Resour. Dev. Rev.* **2005**, *4*, 356–367. [CrossRef]

11. Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I.D.; Gebru, T. Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA, 29–31 January 2019; pp. 220–229.

12. Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach.* **2018**, *28*, 689–707. [CrossRef]

13. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777.

14. Hardt, M.; Price, E.; Srebro, N. Equality of opportunity in supervised learning. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 3323–3331.

15. Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ding, D.; Bagul, A.; Langlotz, C.; Shpanskaya, K.; et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv* **2017**, arXiv:1711.05225.

16. Jobin, A.; Ienca, M.; Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399. [CrossRef]

17. Floridi, L.; Cowls, J. A unified framework of five principles for AI in society. In *Machine Learning and the City: Applications in Architecture and Urban Design*; Wiley-Blackwell: Hoboken, NJ, USA, 2022; pp. 535–545.

18. Lipton, Z.C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* **2018**, *16*, 31–57. [CrossRef]

19. Mohadi, M.; Tarshany, Y. Maqasid al-Shari'ah and the ethics of artificial intelligence: Contemporary challenges. *J. Contemp. Maqasid Stud.* **2023**, *2*, 79–102. [CrossRef]

20. Greenhalgh, T.; Wherton, J.; Papoutsi, C.; Lynch, J.; Hughes, G.; A'Court, C.; Hinder, S.; Fahy, N.; Procter, R.; Shaw, S. Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *J. Med. Internet Res.* **2017**, *19*, e8775. [CrossRef]

21. Barocas, S.; Hardt, M.; Narayanan, A. *Fairness and Machine Learning: Limitations and Opportunities*; MIT Press: Cambridge, MA, USA, 2023.

22. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.

23. Moldovan, A.; Vescan, A.; Grosan, C. Healthcare bias in AI: A systematic literature review. In Proceedings of the 20th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE), Porto, Portugal, 4–6 April 2025.

24. Cho, H.N.; Ahn, I.; Gwon, H.; Kang, H.J.; Kim, Y.; Seo, H.; Choi, H.; Kim, M.; Han, J.; Kee, G.; et al. Explainable predictions of a machine learning model to forecast the postoperative length of stay for severe patients: Machine learning model development and evaluation. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 350. [CrossRef]

25. Elmahjub, E. Artificial intelligence (AI) in Islamic ethics: Towards pluralist ethical benchmarking for AI. *Philos. Technol.* **2023**, *36*, 73. [CrossRef]

26. Chamsi-Pasha, H.; Albar, M.; Chamsi-Pasha, M. Comparative Study between Islamic and Western Bioethics: The Principle of Autonomy. *J. Br. Islam. Med. Assoc.* **2022**, *11*, 1–12.

27. Procter, R.; Tolmie, P.; Rouncefield, M. Holding AI to account: Challenges for the delivery of trustworthy AI in healthcare. *ACM Trans. Comput.-Hum. Interact.* **2023**, *30*, 1–34. [CrossRef]

28. Sadeghi, Z.; Alizadehsani, R.; Cifci, M.A.; Kausar, S.; Rehman, R.; Mahanta, P.; Bora, P.K.; Almasri, A.; Alkhawaldeh, R.S.; Hussain, S.; et al. A brief review of explainable artificial intelligence in healthcare. *arXiv* **2023**, arXiv:2304.01543. [CrossRef]

29. Ghaly, M. Biomedical scientists as co-muftis: Their contribution to contemporary Islamic bioethics. *Die Welt Des Islams* **2015**, *55*, 286–311. [CrossRef]

30. Doshi-Velez, F.; Kortz, M.; Budish, R.; Bavitz, C.; Gershman, S.; O'Brien, D.; Scott, K.; Schieber, S.; Waldo, J.; Weinberger, D.; et al. Accountability of AI under the law: The role of explanation. *arXiv* **2017**, arXiv:1711.01134. [CrossRef]

31. Mittelstadt, B. Principles alone cannot guarantee ethical AI. *Nat. Mach. Intell.* **2019**, *1*, 501–507. [CrossRef]

32. Quraishi, A. Principles of Islamic Jurisprudence. By Mohammad Hashim Kamali. Cambridge: Cambridge Islamic Texts Society1991. Pp. xxi, 417. Price not available. ISBN: 0-946-62123-3. Paper. ISBN: 0-946-62124-1. *J. Law Relig.* **2001**, *15*, 385–387. [CrossRef]

33. Ghaly, M. Islamic ethico-legal perspectives on medical accountability in the age of artificial intelligence. In *Research Handbook on Health, AI and the Law*; Edward Elgar Publishing Ltd.: Cheltenham, UK, 2024; pp. 234–253.

34. Ghaly, M. Islamic Bioethics in the Twenty–First Century. *Zygon* **2013**, *48*, 592–599. [CrossRef]

35. Habeebullah, A.A. Ethical Considerations of Artificial Intelligence in Healthcare: A Study in Islamic Bioethics. *Fountain Univ. J. Arts Humanit.* **2024**, *1*, 14–24.

36. Gerdes, A. A participatory data-centric approach to AI ethics by design. *Appl. Artif. Intell.* **2022**, *36*, 2009222. [CrossRef]

37. Bouderhem, R. Shaping the future of AI in healthcare through ethics and governance. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 416. [CrossRef]

38. Kamali, M.H. *Principles of Islamic Jurisprudence*; The Islamic Text Society: Cambridge, UK, 2003; Volume 14, pp. 369–384.

39. Tucci, V.; Saary, J.; Doyle, T.E. Factors influencing trust in medical artificial intelligence for healthcare professionals: A narrative review. *J. Med. Artif. Intell.* **2022**, *5*, 4. [CrossRef]

40. Solanki, P.; Grundy, J.; Hussain, W. Operationalising ethics in artificial intelligence for healthcare: A framework for AI developers. *AI Ethics* **2023**, *3*, 223–240. [CrossRef]

41. Morley, J.; Machado, C.C.V.; Burr, C.; Cowls, J.; Joshi, I.; Taddeo, M.; Floridi, L. The ethics of AI in health care: A mapping review. *Soc. Sci. Med.* **2020**, *260*, 113172. [CrossRef] [PubMed]

42. Guidance, W. *Ethics and Governance of Artificial Intelligence for Health*; World Health Organization: Geneva, Switzerland, 2021.

43. Brett, M. Issues to consider relating to information governance and artificial intelligence. *Cyber Secur. Peer-Rev. J.* **2024**, *7*, 237–252. [CrossRef]

44. Nastoska, A.; Jancheska, B.; Rizinski, M.; Trajanov, D. Evaluating Trustworthiness in AI: Risks, Metrics, and Applications Across Industries. *Electronics* **2025**, *14*, 2717. [CrossRef]

45. Ratti, E.; Morrison, M.; Jakab, I. Ethical and social considerations of applying artificial intelligence in healthcare—A two-pronged scoping review. *BMC Med. Ethics* **2025**, *26*, 68. [CrossRef]

46. Bapatla, S.K.S. Ethical AI in Healthcare: A Framework for Equity-by-Design. *J. Multidiscip.* **2025**, *5*, 143–153.

47. Gilpin, L.H.; Bau, D.; Yuan, B.Z.; Bajwa, A.; Specter, M.; Kagal, L. Explaining explanations: An overview of interpretability of machine learning. In Proceedings of the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 1–3 October 2018; IEEE: New York, NY, USA, 2018; pp. 80–89.