INTERNATIONAL JOURNAL OF ADVANCED RESEARCH
ISSN: 3106-7980: 3106-7972
Volume 3, Issue 2, 2026

DYNAMIC APPROACHES TO MACHINE LEARNING ETHICS IN SOFTWARE ENGINEERING: A COMPARATIVE EXPLORATION OF PAKISTAN AND SAUDI ARABIA

Abdul Mannan[*1], Dr. Nadeem Ahmad Malik[2], Muhammad Huzaifa Noor[3], Muhammad Ali[4], Mehmood Siddique[5]

*[*1]Management Engineer, Department of Engineering for Innovation, University of Salento, Lecce, Italy*
[2]Director IT Services, PMAS Arid Agriculture University, Rawalpindi, Pakistan
[3]Research Scholar, Department of Software Engineering, National University of Modern Languages (NUML), Islamabad, Pakistan
[4]Assistant Professor, Department of Computing and Innovation, Riphah International University, Sahiwal, Pakistan
[5]Research Scholar, Faculty of Computing, Riphah International University, Islamabad, Pakistan

[*1]abdul.mannan17900@gmail.com, [2]nadeem.malik@uaar.edu.pk, [3]huzaifanoor.web@gmail.com, [4]m.ali@riphahsahiwal.edu.pk, [5]mehmoodmeo@hotmail.com

DOI: https://doi.org/10.5281/zenodo.21507803

Keywords
Machine learning ethics, Software engineering, Responsible AI, Ethical governance, Comparative study

Article History
Received: 28 April 2026
Accepted: 15 June 2026
Published: 30 June 2026

Corresponding Author: *
Abdul Mannan

Abstract

The increasing integration of machine learning (ML) into software engineering has transformed software development while intensifying ethical concerns related to fairness, transparency, accountability, and governance. Grounded in Socio-Technical Systems (STS) Theory and the Responsible AI Framework, this study explores how software professionals operationalize ethical principles throughout the Software Development Life Cycle (SDLC) in Pakistan and Saudi Arabia. A qualitative comparative design was employed using semi-structured interviews with 16 software engineers, AI engineers, machine learning specialists, software architects, and technology managers from organizations actively developing or deploying ML-enabled software. The interview data were analyzed using Braun and Clarke's reflexive thematic analysis. Five interrelated themes emerged: ethical awareness in machine learning development, fairness and bias mitigation, transparency and explainability, organizational governance and regulatory compliance, and cross-cultural perspectives on responsible AI implementation. The findings reveal that ethical machine learning is fundamentally a socio-technical process shaped by interactions among technical practices, organizational governance, professional expertise, and institutional environments. The study extends the application of STS Theory to software engineering, operationalizes the Responsible AI Framework within the SDLC, and proposes a Socio-Technical Ethical Governance Framework to support trustworthy machine learning development across emerging digital economies.

Introduction

Background of the Research

Machine learning (ML) has become one of the most transformative technologies influencing contemporary software engineering, fundamentally changing the way software systems are designed, developed, tested, deployed, and maintained. Recent advances in artificial intelligence (AI) have enabled software applications to perform increasingly sophisticated tasks such as predictive analytics, intelligent automation, natural language processing, computer vision, and autonomous decision-making. Consequently, software engineering has evolved from traditional rule-based programming toward data-driven development approaches in which machine learning algorithms continuously learn from large volumes of data to improve system performance and user experiences (Abbas et al., 2025; Ain et al., 2025). Furthermore, AI-powered tools have accelerated software development through automated code generation, intelligent testing, defect prediction, project management, and software maintenance, making machine learning an indispensable component of modern software engineering practices (Alenezi & Akour, 2025).

The widespread adoption of machine learning has significantly increased organizational dependence on AI-driven software across healthcare, finance, education, cybersecurity, tourism, manufacturing, and public administration. While these intelligent systems improve efficiency, innovation, and decision quality, they simultaneously introduce complex ethical challenges related to fairness, transparency, accountability, explainability, privacy, and governance (Alvear et al., 2025). As a result, software engineers are increasingly expected to integrate ethical considerations throughout the Software Development Life Cycle (SDLC), ensuring that AI systems are technically robust while remaining socially responsible and aligned with human values (Sohag et al., 2023).

The importance of responsible AI is particularly evident in emerging digital economies such as Pakistan and Saudi Arabia. Pakistan's software industry is expanding rapidly through increasing software exports, digital entrepreneurship, and AI-focused startups that are integrating machine learning into commercial software solutions (Ashraf et al., 2025; Khan et al., 2026). Meanwhile, Saudi Arabia has positioned artificial intelligence as a strategic pillar of Saudi Vision 2030, investing heavily in smart government initiatives, healthcare modernization, digital transformation, intelligent tourism, and AI-enabled public services (Alahmari et al., 2022; Alsahafi et al., 2023; Khan, 2025; AlShwaier & Abdullah, 2025). These developments highlight the growing need for ethical governance frameworks capable of supporting responsible machine learning implementation within diverse organizational and cultural contexts.

Problem Statement

Despite the rapid advancement of machine learning technologies, ethical governance has not progressed at the same pace as technological innovation. Many organizations continue to prioritize technical performance, automation, and business outcomes while lacking standardized ethical software engineering practices that adequately address fairness, bias mitigation, transparency, explainability, accountability, and responsible AI governance throughout the software development lifecycle. Consequently, software professionals frequently encounter ethical dilemmas when designing, training, deploying, and maintaining machine learning applications.

Existing literature has largely focused on improving algorithmic performance, model accuracy, and technical optimization, whereas comparatively limited attention has been devoted to understanding ethical decision-making from the perspective of software engineers and AI practitioners (Oyeniran et al., 2022; Aljawawdeh et al., 2024). Recent studies have emphasized the importance of embedding AI ethics into agile software engineering processes; however, empirical qualitative evidence exploring practitioners' lived experiences remains relatively scarce (Amugongo et al., 2025). Moreover, very few comparative studies have investigated how

software professionals navigate ethical challenges across different national contexts, particularly between Pakistan and Saudi Arabia, creating a significant gap in understanding responsible AI implementation within emerging digital economies (Khan et al., 2026).

Research Questions

To address these gaps, this study seeks to answer the following research questions:

RQ1: How do software professionals perceive ethical challenges during machine learning software development?

RQ2: How are fairness and bias managed throughout software engineering processes?

RQ3: How is explainability incorporated into machine learning applications?

RQ4: How do organizations govern ethical AI implementation?

RQ5: How do ethical practices differ between Pakistan and Saudi Arabia?

Research Objectives

The primary objective of this study is to explore how software professionals perceive and implement ethical machine learning practices throughout the Software Development Life Cycle (SDLC) within the contexts of Pakistan and Saudi Arabia. Specifically, the study aims to:

1. To explore ethical awareness and ethical decision-making practices among software professionals involved in machine learning development.
2. To examine organizational strategies for promoting fairness and mitigating algorithmic bias throughout the machine learning development process.
3. To investigate how transparency and explainability are incorporated into AI-enabled software systems to support responsible software engineering.
4. To analyze organizational governance mechanisms, policies, and regulatory practices that facilitate the ethical implementation of machine learning applications.
5. To compare responsible machine learning practices and ethical software engineering approaches between Pakistan and Saudi Arabia from a socio-technical perspective.

Significance of the Research

The findings of this research contribute to academic, practical, and policy domains. Academically, the study extends the growing body of Responsible AI literature by integrating ethical governance into software engineering research and providing comparative qualitative evidence from two emerging digital economies. It also enriches understanding of how ethical principles are operationalized throughout the software development lifecycle rather than being treated solely as technical or algorithmic concerns.

From a practical perspective, the study offers valuable insights for software organizations, AI developers, software architects, and technology managers seeking to strengthen ethical governance in machine learning projects. The findings may support organizations in developing internal AI ethics policies, improving bias mitigation strategies, enhancing model transparency, and integrating responsible AI practices into software engineering workflows.

From a policy perspective, the research provides evidence that can assist governments and regulatory bodies in designing context-specific AI governance frameworks. The findings are particularly relevant to Saudi Arabia's Vision 2030 digital transformation agenda and Pakistan's expanding software industry, where ethical AI standards are becoming increasingly important for sustainable technological development (Khan, 2025; AlShwaier & Abdullah, 2025; Tariq et al., 2025). By encouraging collaboration among policymakers, academia, and industry, this study contributes to the development of trustworthy, transparent, and socially responsible machine learning systems capable of supporting long-term digital innovation.

Literature Review

Introduction

The rapid advancement of machine learning (ML) has fundamentally reshaped software engineering by enabling intelligent software systems capable of

learning from data, adapting to dynamic environments, and supporting complex decision-making processes. Across sectors such as healthcare, finance, education, cybersecurity, tourism, and public administration, machine learning has become a core component of software development, allowing organizations to improve operational efficiency, automate decision-making, and deliver personalized digital services (Abbas et al., 2025; Ain et al., 2025; Alenezi & Akour, 2025; Alsehani, 2026). While these technological developments have accelerated innovation, they have simultaneously introduced significant ethical concerns regarding algorithmic fairness, transparency, accountability, privacy, explainability, and governance. Consequently, software engineering is no longer viewed solely as a technical discipline but increasingly as a socio-technical process in which software quality depends on the interaction between technological capabilities, human expertise, organizational practices, and societal expectations (Alvear et al., 2025; Sohag et al., 2023; Hanif et al., 2026).

Recent literature consistently argues that responsible AI has become essential for ensuring that machine learning systems generate socially beneficial outcomes while minimizing unintended ethical consequences (Javed et al., 2023; Oyeniran et al., 2022; Tariq et al., 2025). Rather than treating ethics as a post-development compliance activity, contemporary software engineering emphasizes integrating ethical principles throughout the Software Development Life Cycle (SDLC), including requirements engineering, data collection, model development, validation, deployment, and continuous monitoring (Amugongo et al., 2025; Abbas et al., 2025). However, despite growing scholarly attention, much of the existing literature remains dominated by technical discussions of AI algorithms and software performance, with comparatively limited attention devoted to understanding how software professionals operationalize ethical principles within organizational settings, particularly in emerging digital economies. To address this gap, the present study synthesizes existing research through the complementary perspectives of Socio-Technical Systems (STS) Theory and the Responsible AI Framework, providing a comprehensive foundation for examining ethical machine learning practices in software engineering.

Theoretical Framework: Socio-Technical Systems Theory and Responsible AI Framework

This study is primarily grounded in Socio-Technical Systems (STS) Theory, which argues that organizational effectiveness depends on the interaction between technical systems and social systems rather than technological capability alone. Applied to software engineering, STS suggests that machine learning systems are shaped by algorithms, software professionals, organizational governance, regulatory frameworks, and stakeholder interactions. Consequently, ethical software engineering emerges through the alignment of technological innovation with organizational practices, human expertise, and institutional environments rather than technical optimization alone (Abbas et al., 2025; Alenezi & Akour, 2025; Alvear et al., 2025). The theory is particularly relevant because ethical issues such as dataset bias, explainability, and governance arise from socio-technical interactions involving developers, managers, users, and policymakers (Sohag et al., 2023; Amugongo et al., 2025; Aljawawdeh et al., 2024).

Complementing STS, the Responsible AI Framework provides the operational principles of fairness, transparency, accountability, and human oversight that guide ethical machine learning implementation. Fairness minimizes discrimination, transparency promotes explainable decision-making, accountability ensures organizational responsibility, and human oversight preserves meaningful human control over AI systems (Javed et al., 2023; Oyeniran et al., 2022). These principles are strengthened through explainable AI, governance mechanisms, ethical auditing, and continuous monitoring (Hanif et al., 2026; Tariq et al., 2025; Khan, 2025; Amugongo et al., 2025). Together, STS Theory explains why ethical challenges emerge, while the Responsible AI Framework explains how they can be

addressed, providing the conceptual foundation for the five themes investigated in this study.

Ethical Awareness throughout the Software Development Lifecycle

The growing adoption of machine learning has transformed ethical awareness from a peripheral concern into a fundamental component of contemporary software engineering. Earlier software engineering research primarily emphasized functionality, reliability, and software quality; however, the increasing deployment of AI-enabled applications has expanded developers' responsibilities to include broader societal considerations such as fairness, accountability, transparency, and user trust (Abbas et al., 2025; Ain et al., 2025; Alenezi & Akour, 2025). As organizations increasingly deploy machine learning systems in sensitive domains including healthcare, finance, education, cybersecurity, and public administration, software engineers are expected to anticipate ethical risks alongside technical requirements. Consequently, ethical reasoning is becoming embedded throughout the Software Development Life Cycle rather than being treated as a post-deployment compliance exercise (Amugongo et al., 2025).

From the perspective of STS Theory, ethical awareness develops through interactions among developers, organizational culture, professional norms, regulatory expectations, and stakeholder needs. Software engineers do not make ethical decisions independently; rather, their decisions are influenced by multidisciplinary collaboration, organizational leadership, client expectations, legal requirements, and societal values. Sohag et al. (2023) argue that human-centred software engineering requires integrating ethical thinking into every stage of AI system development to ensure that technological innovation remains aligned with human well-being. Likewise, Alvear et al. (2025) contend that responsible software engineering demands balancing technological advancement with social responsibility, while Hanif et al. (2026) demonstrate that ethical awareness improves software reliability by encouraging explainability, accountability, and continuous human oversight. Collectively, the literature suggests that ethical awareness represents the foundation of responsible machine learning because it shapes how software professionals collect data, design algorithms, evaluate model performance, and govern intelligent software systems within increasingly complex socio-technical environments.

Explainability and Transparency in AI Systems

The growing complexity of machine learning algorithms has intensified scholarly attention toward explainability and transparency as essential dimensions of responsible software engineering. While sophisticated machine learning models often achieve superior predictive performance, their opaque decision-making processes may reduce user confidence, complicate software maintenance, and hinder organizational accountability. Consequently, software engineering research increasingly advocates the adoption of Explainable Artificial Intelligence (XAI), which seeks to make algorithmic decisions understandable, interpretable, and justifiable to both technical and non-technical stakeholders (Hanif et al., 2026; Aljawawdeh et al., 2024). Transparency has therefore evolved from a desirable software attribute into an ethical necessity that enables organizations to identify model errors, validate predictions, support regulatory compliance, and strengthen public trust in AI-enabled software systems.

Within Socio-Technical Systems (STS) Theory, explainability is understood as a socio-technical capability rather than merely a technical feature. Machine learning systems operate within organizational ecosystems involving software engineers, project managers, regulators, business executives, clients, and end users, all of whom require different levels of understanding of AI-generated decisions. Consequently, effective explainability facilitates communication among these stakeholders, improves collaborative decision-making, and enhances organizational accountability throughout the Software Development Life Cycle (SDLC). Hanif et al. (2026) demonstrate that explainable AI

significantly improves ethical decision-making in cybersecurity software by enabling professionals to justify automated threat detection and reduce uncertainty during incident response. Similarly, Aljawawdeh et al. (2024) argue that explainability strengthens human-computer interaction by improving users' understanding of AI-supported recommendations, thereby increasing system acceptance and reducing ethical concerns associated with opaque decision-making.

Transparency also contributes to organizational governance by enabling software teams to evaluate model performance continuously, identify unintended consequences, and maintain accountability throughout system deployment and maintenance. Tariq et al. (2025) contend that transparent AI systems facilitate ethical auditing and regulatory compliance by documenting how machine learning models generate predictions and by providing evidence for organizational decision-making. Likewise, Javed et al. (2023) emphasize that explainability is particularly critical in high-risk applications such as healthcare because interpretable machine learning models enable practitioners to understand clinical recommendations before making critical decisions. Collectively, these studies suggest that transparency should be embedded throughout software engineering processes rather than treated as a post-development requirement, reinforcing the socio-technical principle that trustworthy AI emerges through collaboration between technology and its human stakeholders.

Organizational Governance and Ethical Compliance

Although software engineers play a central role in developing responsible machine learning systems, ethical implementation ultimately depends upon organizational governance structures that institutionalize ethical decision-making throughout software development. As AI technologies become increasingly embedded within organizational operations, governance mechanisms such as AI ethics committees, organizational policies, ethical risk assessments, algorithmic auditing, compliance monitoring, documentation standards, and continuous ethics training have become essential for ensuring responsible innovation (Khan, 2025; Alsumari & Alrwais, 2025). These governance practices help organizations balance technological advancement with legal obligations, stakeholder expectations, and societal values while reducing ethical risks associated with automated decision-making.

Socio-Technical Systems Theory explains organizational governance as the interaction between technological systems and organizational structures. Ethical software engineering therefore depends not only on technically competent developers but also on leadership commitment, institutional policies, collaborative decision-making, and supportive organizational cultures. Khan (2025) argues that effective AI governance requires integrating regulatory compliance with organizational strategy to ensure that machine learning systems align with national digital transformation agendas while protecting public interests. Similarly, Amugongo et al. (2025) demonstrate that embedding ethical governance within agile software development enables organizations to identify ethical risks continuously throughout iterative development rather than relying solely on post-deployment evaluations. Alsumari and Alrwais (2025) further report that structured AI project management methodologies improve governance effectiveness by clearly defining stakeholder responsibilities, accountability mechanisms, and ethical review procedures. Together, these studies indicate that organizational governance functions as the bridge connecting technical innovation with responsible software engineering, ensuring that ethical principles remain embedded throughout the entire software development lifecycle.

Cross-Cultural Perspectives of Responsible AI in Emerging Digital Economies

The implementation of responsible AI is shaped by national priorities, institutional environments, regulatory maturity, and organizational culture. Pakistan and Saudi Arabia represent two rapidly evolving digital economies that have embraced machine learning to accelerate technological

innovation while following distinct developmental pathways. Pakistan's expanding software industry has experienced significant growth through increasing software exports, entrepreneurial innovation, and AI adoption across private-sector organizations, whereas Saudi Arabia has integrated artificial intelligence into broader national transformation initiatives through Saudi Vision 2030, smart government programs, intelligent healthcare, digital tourism, and innovation-driven economic diversification (AlShwaier & Abdullah, 2025; Khan, 2025; Alahmari et al., 2022; Alsahafi et al., 2023). Despite these contextual differences, both countries increasingly recognize responsible AI as a prerequisite for sustainable digital transformation.

From the perspective of STS Theory, ethical software engineering reflects interactions between technology and the social, institutional, and cultural environments within which software systems are developed. Consequently, organizations operating in different countries may implement similar ethical principles through different governance structures and organizational practices. AlShwaier and Abdullah (2025) demonstrate that Saudi Arabia's innovation ecosystem promotes responsible AI by aligning entrepreneurial activities with national digital policies, while Mohamad et al. (2025) show that higher education institutions across the Middle East increasingly emphasize ethical AI competencies to prepare future software engineers. Conversely, Khan et al. (2026) report that Pakistani software organizations are progressively integrating ethical considerations into software development despite comparatively less formalized governance frameworks, relying more heavily on professional responsibility and organizational commitment. Barefah and Altalhi (2024) further highlight the importance of stakeholder collaboration and organizational participation in improving software engineering practices within Saudi organizations. These studies collectively indicate that although fairness, transparency, accountability, and human oversight represent universal ethical principles, their practical implementation remains strongly influenced by national socio-technical environments, reinforcing the need for context-specific yet globally aligned AI governance frameworks.

Gaps in Existing Research

The existing literature provides valuable insights into artificial intelligence, machine learning, and software engineering; however, several significant theoretical and empirical gaps remain. First, the majority of previous studies primarily examine technical dimensions of AI, including algorithm development, predictive performance, software automation, machine learning optimization, and AI-assisted software engineering, while comparatively limited attention has been devoted to understanding how software professionals experience and operationalize ethical principles during software development (Abbas et al., 2025; Ain et al., 2025; Alenezi & Akour, 2025). Second, existing research has largely concentrated on conceptual discussions or quantitative investigations, providing relatively little qualitative evidence regarding ethical decision-making, organizational governance, and the lived experiences of software engineers responsible for designing and deploying machine learning systems (Alvear et al., 2025; Amugongo et al., 2025; Sohag et al., 2023).

Furthermore, previous studies frequently investigate fairness, transparency, explainability, governance, and accountability as independent constructs rather than examining their interrelationships throughout the Software Development Life Cycle from a socio-technical perspective (Javed et al., 2023; Oyeniran et al., 2022; Tariq et al., 2025; Hanif et al., 2026). Limited attention has also been given to understanding how organizational culture, leadership, regulatory environments, and stakeholder collaboration collectively influence ethical machine learning implementation, despite the central assumptions of Socio-Technical Systems Theory (Aljawawdeh et al., 2024; Khan, 2025; Alsumari & Alrwais, 2025). Additionally, comparative cross-country investigations remain

scarce, particularly between emerging digital economies such as Pakistan and Saudi Arabia, where AI adoption is expanding rapidly under different institutional and policy environments (AlShwaier & Abdullah, 2025; Mohamad et al., 2025; Barefah & Altalhi, 2024; Alahmari et al., 2022; Alsahafi et al., 2023). Existing empirical evidence also pays limited attention to how software engineers, AI specialists, software architects, and technology managers collectively perceive ethical responsibilities during machine learning development (Khan et al., 2026; Rehmat et al., 2025).

Addressing these limitations, the present study adopts Socio-Technical Systems Theory as its primary theoretical lens and the Responsible AI Framework as its operational framework to examine ethical machine learning within software engineering. Through a comparative qualitative investigation involving software professionals from Pakistan and Saudi Arabia, the study extends existing literature by providing practitioner-centred evidence on ethical awareness, fairness, transparency, governance, and cross-cultural implementation. In doing so, it contributes a more holistic understanding of responsible AI by demonstrating how ethical machine learning emerges from the interaction of technological capabilities, organizational governance, professional practice, and national institutional contexts, thereby enriching both software engineering scholarship and the broader Responsible AI literature (Alvear et al., 2025; Amugongo et al., 2025; Khan et al., 2026).

Methodology

Research Design

This study adopted a qualitative research methodology grounded in the interpretivist paradigm to explore software professionals' perceptions and experiences regarding machine learning ethics in software engineering. A qualitative approach was considered appropriate because the study sought to understand participants' subjective interpretations, organizational experiences, and ethical decision-making practices rather than quantify predefined variables. The interpretivist perspective acknowledges that ethical practices are socially constructed and shaped by organizational culture, professional experience, and national contexts. Furthermore, the research employed a comparative qualitative design to examine similarities and differences in responsible AI practices between Pakistan and Saudi Arabia. Semi-structured interviews were selected as the primary data collection method because they provide flexibility for participants to discuss complex ethical issues while allowing the researcher to probe emerging themes and contextual experiences. This methodological approach aligns with recommendations for qualitative inquiry and reflexive thematic analysis presented by Braun et al. (2022) and Byrne (2022).

Sampling Strategy

Purposive sampling was employed to recruit participants possessing substantial knowledge and practical experience in machine learning and software engineering. The study included sixteen professionals working in organizations actively developing, deploying, or managing machine learning applications. To facilitate meaningful cross-country comparison, the sample comprised eight participants from Pakistan and eight from Saudi Arabia. Participants represented diverse professional roles, including software engineers, AI engineers, machine learning specialists, software architects, and technology managers from industries such as information technology, healthcare, financial technology, cybersecurity, e-commerce, telecommunications, and digital consulting. All participants possessed a minimum of five years of professional experience and had direct involvement in AI-enabled software development or governance activities.

Data Collection

Data were collected through semi-structured online interviews conducted using secure video conferencing platforms. Each interview lasted approximately 45–60 minutes and followed a structured interview guide developed from the Responsible AI literature, focusing on ethical

awareness, fairness and bias mitigation, explainability, governance mechanisms, and cross-cultural perspectives on responsible AI implementation. Participants were encouraged to share practical experiences, organizational practices, and challenges encountered during machine learning software development. With participants' permission, all interviews were audio-recorded and subsequently transcribed verbatim to ensure data accuracy and facilitate systematic analysis (Braun et al., 2022).

Data Analysis

The interview data were analyzed using Braun and Clarke's Reflexive Thematic Analysis, a rigorous qualitative analytical approach suitable for identifying patterns across participants' experiences. Following Braun et al. (2022), Byrne (2022), and Hole (2024), the analysis progressed through six iterative phases: (1) familiarization with the interview transcripts through repeated reading; (2) generation of initial codes representing meaningful units of data; (3) development of preliminary themes by grouping related codes; (4) review and refinement of themes to ensure coherence and distinctiveness; (5) definition and naming of final themes that accurately reflected participants' experiences; and (6) production of the final analytical narrative supported by representative participant quotations. This systematic process enhanced the credibility, consistency, and interpretive depth of the findings.

Ethical Considerations

Ethical principles were rigorously maintained throughout the research process. Participation was entirely voluntary, and informed consent was obtained from all participants before data collection commenced. Participants were informed about the study's objectives, confidentiality procedures, and their right to withdraw at any stage without consequence. To preserve anonymity, each participant was assigned a unique identification code (P1–P16), and all personally identifiable information was removed from the interview transcripts. Audio recordings and transcribed data were securely stored using password-protected digital storage accessible only to the research team. The findings are reported anonymously to protect participants' professional identities while ensuring the integrity and trustworthiness of the research.

Participants' Profile

The table below presents the demographic and professional characteristics of all interview participants, including their country, gender, professional designation, years of software engineering experience, industry sector, organizational type, highest qualification, and primary area of expertise.

Table 1: Participant Profiles

Participant	Country	Gender	Professional Role	Organization Type	Industry Sector	Experience (Years)
P1	Pakistan	Male	Senior Software Engineer	Software House	Information Technology	7
P2	Pakistan	Male	AI Engineer	Technology Company	Artificial Intelligence	8
P3	Pakistan	Female	Machine Learning Specialist	FinTech Company	Financial Technology	6
P4	Pakistan	Male	Software Architect	Enterprise Solutions	Information Technology	12
P5	Pakistan	Female	Technology Manager	Healthcare Technology Firm	Healthcare	10

Participant	Country	Gender	Professional Role	Organization Type	Industry Sector	Experience (Years)
P6	Pakistan	Male	Senior ML Engineer	Cybersecurity Company	Cybersecurity	9
P7	Pakistan	Female	Software Engineer	E-Commerce Company	E-Commerce	5
P8	Pakistan	Male	AI Solutions Consultant	Digital Consultancy	Information Technology	11
P9	Saudi Arabia	Male	Software Architect	Government Technology Agency	Smart Government	13
P10	Saudi Arabia	Female	AI Engineer	Healthcare Organization	Healthcare	7
P11	Saudi Arabia	Male	Machine Learning Specialist	Telecommunications Company	Telecommunications	8
P12	Saudi Arabia	Female	Technology Manager	Financial Services Organization	Banking and Finance	12
P13	Saudi Arabia	Male	Senior Software Engineer	Smart City Project	Public Sector	9
P14	Saudi Arabia	Female	AI Solutions Architect	Digital Innovation Center	Digital Innovation	10
P15	Saudi Arabia	Male	ML Engineer	Cloud Technology Company	Cloud Computing	6
P16	Saudi Arabia	Female	Director of AI Programs	Government Digital Authority	Public Administration	15

Findings and Discussion

The thematic analysis of sixteen semi-structured interviews conducted with software engineers, AI engineers, machine learning specialists, software architects, and technology managers from Pakistan and Saudi Arabia generated five interrelated themes that explain how software professionals perceive and implement ethical machine learning throughout the Software Development Life Cycle (SDLC). Viewed through the lens of Socio-Technical Systems (STS) Theory, the findings indicate that ethical machine learning is not determined solely by algorithmic design or technical expertise but emerges from continuous interactions among software professionals, organizational governance, technological infrastructures, regulatory environments, and stakeholder expectations. Across both countries, participants consistently emphasized that responsible AI practices are increasingly integrated into software engineering processes; however, their implementation depends upon organizational maturity, institutional support, and national digital transformation priorities. The first three themes highlight ethical awareness, fairness and bias mitigation, and transparency and explainability as the foundational socio-technical mechanisms supporting responsible machine learning development.

Theme 1: Ethical Awareness in Machine Learning Development

Ethical awareness emerged as the foundation of responsible machine learning development, with participants consistently acknowledging that software engineering responsibilities now extend beyond producing technically accurate algorithms to developing AI systems that align with societal

values, organizational objectives, and user expectations. Participants explained that ethical considerations are increasingly incorporated into project planning, data preparation, model development, testing, deployment, and system maintenance, reflecting a gradual shift from technology-driven software engineering towards a more human-centred and socio-technical approach.

A senior software architect from Saudi Arabia (P9) explained, *"Today we cannot evaluate AI systems only by accuracy. We also ask whether the model is fair, whether users understand its decisions, and whether it could unintentionally discriminate against certain groups."* Likewise, a Pakistani AI engineer (P2) stated, *"Our development teams now discuss ethical risks during project meetings because clients increasingly expect responsible AI solutions instead of only technically advanced systems."*

Participants further explained that ethical awareness develops through continuous interaction between developers, organizational leadership, clients, regulatory expectations, and multidisciplinary project teams. Rather than being viewed as an individual responsibility, ethical decision-making was described as a collaborative organizational process supported by professional learning, management commitment, and stakeholder engagement. Technology managers indicated that organizations investing in ethics training, AI governance policies, and interdisciplinary collaboration exhibited higher levels of ethical awareness among software professionals. Conversely, several Pakistani participants reported that ethical discussions often relied on personal initiative because formal organizational mechanisms remained relatively underdeveloped.

These findings strongly reflect the assumptions of Socio-Technical Systems Theory, which proposes that organizational outcomes emerge through interactions between technical and social subsystems rather than technological capabilities alone. Ethical awareness therefore represents a socio-technical capability developed through collaboration among software professionals, organizational structures, governance mechanisms, and external stakeholders. The findings support Abbas et al. (2025), who argue that software engineering has evolved beyond technical performance towards responsible AI development. They also align with Sohag et al. (2023), who advocate integrating human-centred values throughout software engineering, and with Amugongo et al. (2025), who demonstrate that ethical reflection should accompany every stage of agile software development. Similarly, Alvear et al. (2025) emphasize that responsible software engineering requires balancing technological innovation with societal responsibility. Extending these studies, the present findings demonstrate that ethical awareness is increasingly institutionalized within software engineering practices; however, its maturity remains dependent upon organizational culture, governance capability, leadership commitment, and national digital ecosystems. Compared with Saudi Arabia, ethical awareness in Pakistan was more frequently driven by professional responsibility and client expectations than by standardized organizational frameworks.

Theme 2: Fairness and Bias Mitigation

Fairness and bias mitigation emerged as one of the most significant ethical priorities discussed by participants. Software professionals unanimously recognized that algorithmic bias represents both a technical and organizational challenge capable of undermining software quality, user trust, and organizational credibility. Participants consistently argued that fairness begins with responsible data management because biased, incomplete, or historically unrepresentative datasets inevitably influence model behaviour regardless of algorithmic sophistication.

A machine learning specialist from Pakistan (P3) explained, *"A technically accurate model can still be ethically unacceptable if the training data represent only one segment of society."* Similarly, a Saudi AI engineer (P10) stated, *"Before deployment we perform repeated fairness testing because bias affects user trust more than prediction accuracy."*

Participants described fairness as a continuous process rather than a one-time technical

assessment. Common practices included dataset balancing, data cleaning, fairness metrics, independent model validation, bias testing, multidisciplinary dataset reviews, and post-deployment monitoring. Several respondents explained that fairness assessments require collaboration among software engineers, domain experts, project managers, and business stakeholders to identify demographic, cultural, and contextual biases that purely technical evaluations may overlook. This collaborative approach reflects the socio-technical nature of responsible machine learning, where technical solutions are strengthened through organizational knowledge and stakeholder participation.

Differences between the two countries were evident in governance maturity. Participants from Saudi Arabia described relatively formalized fairness assessment procedures supported by organizational policies and national AI initiatives, whereas Pakistani participants indicated that fairness evaluations were often determined by project scope, client expectations, and available organizational resources. Nevertheless, respondents from both countries agreed that fairness should remain an ongoing activity throughout the software lifecycle because machine learning models continuously evolve as new data become available.

These findings support Oyeniran et al. (2022), who identify dataset quality as a critical determinant of algorithmic fairness, and Javed et al. (2023), who argue that bias mitigation should be integrated throughout model development rather than addressed after deployment. Likewise, Tariq et al. (2025) recommend continuous fairness auditing as an essential component of responsible AI governance. Interpreted through STS Theory, fairness extends beyond algorithm optimization to encompass organizational governance, interdisciplinary collaboration, data governance practices, and institutional accountability. The present study therefore extends previous literature by demonstrating that fairness is simultaneously a technical requirement, an organizational capability, and a socio-technical process shaped by interactions between technology, people, and governance structures.

Theme 3: Transparency and Explainability

Transparency and explainability constituted another dominant theme across participant interviews. Respondents consistently emphasized that machine learning systems should provide understandable explanations for automated decisions, particularly in high-impact domains such as healthcare, finance, cybersecurity, and public administration. Participants explained that explainability enhances stakeholder confidence by allowing users, managers, regulators, and developers to understand how machine learning models generate predictions and recommendations.

A software architect from Pakistan (P4) remarked, *"Clients no longer accept black-box models without asking how decisions are made. Explainability has become a business requirement as much as a technical requirement."* Similarly, a Saudi technology manager (P12) noted, *"Transparency reduces resistance toward AI because decision-makers are more comfortable when they understand why the system reached a particular conclusion."*

Participants further highlighted that explainability supports software maintenance, debugging, quality assurance, regulatory compliance, and communication between technical and non-technical stakeholders. AI engineers reported using feature importance analysis, interpretable models, visualization techniques, and documentation practices to improve understanding of machine learning outputs throughout software development. Several respondents acknowledged that balancing predictive performance with explainability remains challenging, particularly when deploying complex deep learning architectures. Nevertheless, they considered transparency essential for maintaining long-term user trust and organizational accountability.

From an STS perspective, explainability functions as a mechanism that strengthens interactions between technical systems and their surrounding social environment. Rather than merely

improving algorithm interpretation, transparency facilitates communication among developers, organizational leaders, clients, regulators, and end users, thereby supporting collaborative decision-making and responsible software governance. Although participants from both countries regarded explainability as increasingly important, Saudi respondents described stronger organizational emphasis due to evolving regulatory frameworks and government-supported AI initiatives, whereas Pakistani organizations often implemented explainability according to client requirements and project complexity.

These findings are consistent with Hanif et al. (2026), who argue that explainable AI improves organizational accountability and software reliability by making algorithmic decisions interpretable. They also support Aljawawdeh et al. (2024), who found that transparency enhances human-computer interaction and user trust in AI-enabled systems. Furthermore, Tariq et al. (2025) contend that explainability is fundamental for ethical AI governance because it enables accountability and informed oversight. Extending previous research, the present study demonstrates that transparency is not merely a technical characteristic but a socio-technical capability that enables collaboration, governance, stakeholder engagement, and organizational trust throughout the software development lifecycle.

Theme 4: Organizational Governance and Regulatory Compliance

Organizational governance emerged as one of the strongest determinants of responsible machine learning implementation across both countries. Participants consistently emphasized that ethical machine learning cannot depend solely on the ethical judgment or technical competence of individual software engineers but instead requires structured organizational governance, clearly defined policies, institutional accountability, and regulatory alignment. Respondents explained that organizations with formal AI governance frameworks were more successful in embedding ethical principles consistently throughout the Software Development Life Cycle (SDLC), enabling software teams to identify, monitor, and mitigate ethical risks before and after system deployment.

A technology manager from Pakistan (P5) explained, *"Without clear organizational policies, ethical decisions often depend on the experience of individual developers. Governance provides consistency across projects."* Similarly, the Director of AI Programs from Saudi Arabia (P16) stated, *"Every AI project should include ethical risk assessment, documentation, and periodic compliance reviews because governance builds long-term public trust."*

Participants identified numerous governance mechanisms currently adopted within their organizations, including AI ethics committees, ethical review boards, algorithmic risk assessment frameworks, documentation standards, data governance protocols, compliance monitoring, internal auditing procedures, explainability reviews, and continuous ethics training for software professionals. Several respondents explained that governance should accompany every stage of software engineering, beginning with project planning and requirements analysis and continuing through data acquisition, model development, validation, deployment, maintenance, and post-deployment monitoring. Rather than functioning as isolated compliance activities, these governance practices were viewed as continuous organizational processes that improve software quality, reduce ethical uncertainty, and strengthen stakeholder confidence.

The findings strongly reflect the assumptions of Socio-Technical Systems Theory, which argues that technological outcomes are shaped by interactions between technical systems and organizational structures. Participants consistently described governance as the mechanism that connects software engineers, AI specialists, project managers, legal experts, business executives, regulators, and end users within a coordinated decision-making process. Several software architects emphasized that multidisciplinary collaboration allows organizations to identify ethical risks from technical, legal, commercial, and societal perspectives before machine learning

systems are deployed. Consequently, governance was perceived not merely as regulatory compliance but as an organizational capability that facilitates communication, shared responsibility, and continuous ethical improvement.

Important cross-country differences also emerged. Participants from Saudi Arabia reported comparatively mature governance structures supported by national digital transformation strategies, organizational investment in AI governance, and increasing regulatory expectations associated with Saudi Vision 2030. Many organizations had established formal compliance procedures, standardized documentation, periodic ethical audits, and dedicated governance teams responsible for monitoring AI projects. Conversely, participants from Pakistan acknowledged growing organizational awareness regarding responsible AI but indicated that governance practices often remained informal, project-specific, and highly dependent upon organizational leadership, client expectations, and available resources. Nevertheless, respondents from both countries agreed that governance maturity is improving as organizations recognize that ethical machine learning contributes to software quality, organizational reputation, and public trust.

These findings support Khan (2025), who argues that sustainable AI implementation requires governance systems capable of balancing technological innovation with ethical responsibility and regulatory compliance. Similarly, Amugongo et al. (2025) demonstrate that integrating ethical governance into agile software development enables continuous ethical evaluation throughout iterative software engineering processes. Alsumari and Alrwais (2025) further emphasize that structured AI project management methodologies improve organizational accountability by clearly defining governance responsibilities and ethical oversight mechanisms. Likewise, Tariq et al. (2025) contend that organizational governance transforms ethical principles into operational software engineering practices through continuous monitoring and risk management. Extending these studies, the present findings demonstrate that governance functions as the socio-technical mechanism through which fairness, transparency, accountability, and human oversight are systematically institutionalized within software engineering rather than remaining individual ethical aspirations.

Theme 5: Cross-Cultural Perspectives on Responsible AI

The comparative analysis revealed both convergence and divergence in responsible machine learning practices between Pakistan and Saudi Arabia. Participants from both countries unanimously agreed that ethical machine learning has become an indispensable component of contemporary software engineering and that fairness, transparency, accountability, and governance are increasingly viewed as essential software quality attributes rather than optional ethical considerations. However, important differences emerged regarding organizational maturity, institutional support, regulatory environments, and national digital transformation strategies that shape how ethical principles are translated into software engineering practice.

Participants from Pakistan described a rapidly expanding software industry characterized by entrepreneurial innovation, increasing software exports, international collaboration, and growing adoption of artificial intelligence across multiple sectors. Although ethical awareness has increased considerably, respondents explained that responsible AI implementation is primarily influenced by organizational leadership, developers' professional commitment, client expectations, and project-specific requirements rather than comprehensive national governance frameworks. A Pakistani AI consultant (P8) observed, *"Most organizations want to build responsible AI, but ethical implementation depends on available resources and management commitment."* Several respondents further noted that limited organizational resources and inconsistent ethics training occasionally restrict the formal integration of responsible AI practices within software engineering projects.

Conversely, participants from Saudi Arabia consistently highlighted the significant influence of Saudi Vision 2030 and national digital transformation initiatives in shaping organizational AI governance. Respondents explained that government policies have encouraged organizations to formalize AI governance through standardized documentation, ethical risk assessments, regulatory compliance mechanisms, and continuous monitoring of machine learning systems. A software architect (P9) stated, *"Government initiatives have encouraged organizations to treat AI ethics as a strategic priority rather than simply a technical issue."* Likewise, another participant (P14) explained that increasing regulatory expectations have strengthened organizational commitment toward transparency, accountability, documentation standards, and explainable AI practices.

Despite these contextual differences, respondents from both countries identified several common challenges affecting responsible machine learning implementation. These included the rapid evolution of AI technologies, the absence of universally accepted ethical standards, limited availability of specialized ethics training, balancing innovation with regulatory compliance, and the need for greater collaboration among software engineers, policymakers, universities, industry leaders, and professional associations. Participants further emphasized that future AI governance should encourage international knowledge sharing while allowing organizations sufficient flexibility to respond to local cultural, regulatory, and organizational contexts.

From the perspective of Socio-Technical Systems Theory, these findings demonstrate that responsible machine learning is shaped not only by technological capabilities but also by broader institutional, cultural, and organizational environments. Although the ethical principles underpinning Responsible AI remain universally applicable, their implementation differs according to governance maturity, national priorities, regulatory structures, and organizational capabilities. These observations support AlShwaier and Abdullah (2025), who report that Saudi Arabia's innovation ecosystem increasingly promotes responsible AI through national policy initiatives and institutional investment. Similarly, Mohamad et al. (2025) highlight the growing emphasis on AI ethics education and professional competency across the Middle East, while Khan et al. (2026) demonstrate that software engineering organizations within emerging economies increasingly integrate ethical considerations into development practices. Barefah and Altalhi (2024) further emphasize the importance of stakeholder collaboration in strengthening software engineering processes. Extending these studies, the present research demonstrates that responsible AI implementation reflects the interaction between organizational governance, technological capability, professional expertise, and national institutional environments, thereby reinforcing the explanatory power of STS Theory.

Synthesis of Findings

The five themes collectively demonstrate that responsible machine learning is a multidimensional socio-technical phenomenon emerging through continuous interaction among software professionals, technological systems, organizational governance, regulatory environments, and stakeholder expectations. Rather than functioning as independent ethical principles, ethical awareness, fairness, transparency, governance, and cross-cultural influences collectively reinforce one another throughout the Software Development Life Cycle. This interconnectedness strongly supports the assumptions of Socio-Technical Systems Theory, which argues that technological outcomes cannot be separated from the social and organizational systems within which they are developed and implemented.

Ethical awareness establishes the foundation by encouraging software professionals to recognize the broader societal consequences of AI-enabled software systems. Fairness and bias mitigation ensure equitable algorithmic outcomes through responsible data governance, multidisciplinary collaboration, and continuous model evaluation. Transparency and explainability facilitate

communication between technical and non-technical stakeholders while improving software quality, accountability, and user trust. Organizational governance institutionalizes these ethical principles through formal policies, AI ethics committees, risk assessment frameworks, documentation standards, auditing procedures, and regulatory compliance mechanisms. Finally, the comparative findings demonstrate that although Pakistan and Saudi Arabia differ in governance maturity, institutional support, and regulatory development, software professionals across both countries share a common commitment to developing trustworthy, transparent, and socially responsible machine learning systems.

Collectively, these findings extend the Responsible AI Framework by demonstrating that fairness, transparency, accountability, and human oversight become effective only when embedded within supportive socio-technical systems comprising organizational leadership, governance structures, professional expertise, stakeholder collaboration, and enabling institutional environments. Consequently, responsible machine learning should be viewed not merely as a technical software engineering objective but as a socio-technical capability requiring continuous collaboration among software engineers, AI specialists, organizational leaders, policymakers, academic institutions, and regulatory authorities. This integrated perspective forms the basis of the proposed Socio-Technical Ethical Governance Framework presented in Figure 1, which synthesizes the study's principal theoretical and practical contributions.

Table 2: Thematic Analysis of Findings

Theme	Sub-themes	Key Findings (Codes)	STS & Responsible AI Interpretation	Supporting Literature
Ethical Awareness in ML Development	Ethical consciousness; Responsible software engineering; Ethical decision-making	AI literacy, ethical responsibility, professional accountability, human-centered development	Ethical awareness emerged through interactions between software professionals, organizational culture, and development practices, operationalizing accountability and human oversight.	Abbas et al. (2025); Sohag et al. (2023); Amugongo et al. (2025)
Fairness and Bias Mitigation	Dataset quality; Inclusive design; Bias testing	Dataset bias, fairness metrics, balanced sampling, model validation, continuous monitoring	Fairness was achieved through collaboration between technical processes and organizational governance, supporting equitable AI development and continuous accountability.	Oyeniran et al. (2022); Javed et al. (2023); Tariq et al. (2025)
Transparency and Explainability	Explainable AI; User trust; Interpretability	Explainable models, feature importance, decision transparency, stakeholder communication	Transparency strengthened interactions between AI systems and stakeholders, improving trust, human oversight, and software quality within the socio-technical environment.	Hanif et al. (2026); Aljawawdeh et al. (2024); Tariq et al. (2025)

Theme	Sub-themes	Key Findings (Codes)	STS & Responsible AI Interpretation	Supporting Literature
Organizational Governance and Regulatory Compliance	AI governance; Ethical policies; Compliance mechanisms; AI auditing	Governance frameworks, ethics committees, documentation, risk assessment, internal audits	Organizational structures coordinated people, technology, and regulatory processes to institutionalize Responsible AI throughout the SDLC.	Khan (2025); Amugongo et al. (2025); Alsumari & Alrwais (2025)
Cross-Cultural Perspectives on Responsible AI	National priorities; Organizational culture; Regulatory maturity	Saudi Vision 2030, organizational commitment, client expectations, institutional support, collaboration	National and organizational contexts shaped the implementation of Responsible AI, demonstrating that ethical software engineering is influenced by broader socio-technical ecosystems.	AlShwaier & Abdullah (2025); Mohamad et al. (2025); Khan et al. (2026); Barefah & Altalhi (2024)

Summary of Findings

The findings demonstrate that responsible machine learning has become an increasingly important component of contemporary software engineering in both Pakistan and Saudi Arabia. Guided by Socio-Technical Systems Theory and the Responsible AI Framework, the study shows that ethical machine learning is not achieved through technical excellence alone but through the interaction of software professionals, organizational governance, technological processes, regulatory environments, and stakeholder expectations. Across all participant groups, ethical considerations were viewed as integral to the Software Development Life Cycle, with software engineers increasingly recognizing their responsibility to develop AI systems that are not only accurate and efficient but also fair, transparent, accountable, and socially responsible. The five themes collectively indicate that ethical awareness forms the foundation for responsible software engineering, while fairness and bias mitigation, transparency and explainability, organizational governance, and supportive institutional environments collectively enable the practical implementation of ethical AI. Participants reported growing adoption of fairness assessments, explainability techniques, ethical risk evaluations, governance mechanisms, and multidisciplinary collaboration during machine learning development. However, the maturity and consistency of these practices varied considerably across organizations. Saudi Arabian organizations generally demonstrated more structured governance systems supported by national digital transformation initiatives, formal organizational policies, and evolving regulatory frameworks. In contrast, Pakistani organizations relied more heavily on organizational leadership, developers' professional commitment, client expectations, and project-specific ethical practices, reflecting a comparatively less formalized governance environment.

Despite these contextual differences, participants from both countries identified similar challenges, including inconsistent governance structures, evolving regulatory requirements, limited organizational ethics policies, insufficient specialized ethics training, and the need to balance technological innovation with responsible AI implementation. Respondents consistently emphasized that strengthening collaboration among software organizations, policymakers, academic institutions, and professional bodies is essential for establishing sustainable ethical governance frameworks capable of supporting trustworthy machine learning systems.

Overall, the findings extend existing Responsible AI literature by demonstrating that ethical software engineering is fundamentally a socio-

technical process in which technical capabilities, organizational governance, professional expertise, and institutional contexts operate together to shape responsible machine learning implementation. The study therefore provides empirical support for Socio-Technical Systems Theory while illustrating how the principles of the Responsible AI Framework are operationalized within software engineering practice across two emerging digital economies.

.

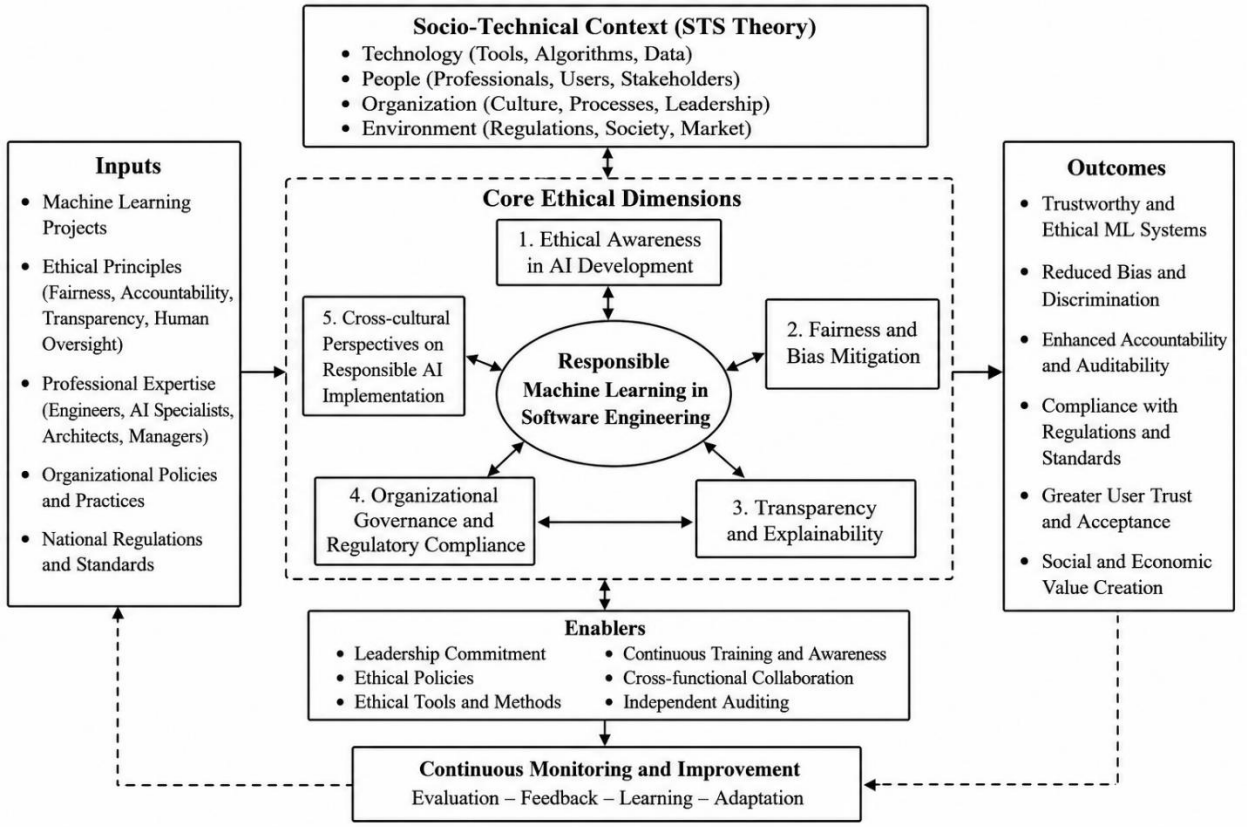

Figure 1 **Proposed Socio-Technical Ethical Governance Framework for Responsible Machine Learning in Software Engineering**

Conclusion

The rapid integration of machine learning into software engineering has transformed software development by creating intelligent systems capable of supporting complex decision-making across diverse application domains. However, alongside these technological advancements, software engineers increasingly face ethical challenges associated with fairness, transparency, accountability, governance, and responsible AI implementation. Guided by Socio-Technical Systems (STS) Theory and the Responsible AI Framework, this study explored how software professionals in Pakistan and Saudi Arabia perceive and operationalize ethical machine learning throughout the Software Development Life Cycle (SDLC). The findings demonstrate that responsible machine learning cannot be achieved through technical excellence alone but requires the effective integration of technological capabilities, organizational governance, professional expertise, regulatory support, and stakeholder engagement.

The comparative findings reveal that ethical awareness, fairness and bias mitigation, transparency and explainability, organizational governance, and national institutional contexts collectively shape responsible software engineering practices. These themes illustrate that ethical machine learning is fundamentally a socio-

technical process in which software engineers, organizational leaders, AI specialists, policymakers, and end users jointly influence the development and deployment of trustworthy AI systems. While participants from both Pakistan and Saudi Arabia demonstrated a strong commitment to responsible AI, organizations in Saudi Arabia generally exhibited more mature governance structures supported by national digital transformation initiatives, whereas organizations in Pakistan relied more heavily on organizational leadership, professional responsibility, and client expectations. Despite these contextual differences, participants consistently emphasized the need for stronger governance frameworks, standardized organizational policies, continuous ethics education, and greater collaboration among industry, academia, and government institutions. Overall, this study contributes comparative empirical evidence demonstrating that sustainable machine learning development depends upon integrating ethical principles into every stage of software engineering while simultaneously strengthening the socio-technical systems that support responsible AI implementation.

Theoretical Implications

This study makes several important theoretical contributions to the literature on machine learning ethics, Responsible AI, and software engineering. First, it extends the application of Socio-Technical Systems Theory to ethical software engineering by demonstrating that responsible machine learning emerges through the interaction of technological systems, software professionals, organizational governance, institutional environments, and stakeholder expectations rather than through algorithmic performance alone. The findings therefore reinforce the central assumptions of STS Theory by illustrating that ethical software development depends upon the alignment of technical and social subsystems throughout the SDLC.

Second, the study complements the Responsible AI Framework by providing empirical evidence demonstrating how the principles of fairness, transparency, accountability, and human oversight are operationalized through socio-technical interactions during software development rather than functioning as isolated ethical principles (Javed et al., 2023). Third, by employing a qualitative comparative design, the study contributes rich empirical evidence to a research domain that has been dominated by conceptual and quantitative investigations, thereby enhancing understanding of software professionals' experiences, organizational practices, and ethical decision-making processes (Amugongo et al., 2025; Alvear et al., 2025). Finally, the comparative analysis advances theoretical understanding of responsible AI implementation within emerging digital economies by illustrating how governance maturity, organizational culture, and national digital transformation initiatives influence ethical software engineering across different institutional contexts (Khan et al., 2026).

Practical Recommendations

The findings suggest several practical recommendations for strengthening responsible machine learning practices within software engineering organizations. First, organizations should embed structured ethical checkpoints throughout every stage of the Software Development Life Cycle, ensuring that ethical risks are systematically evaluated during requirements engineering, data acquisition, model development, testing, deployment, maintenance, and continuous monitoring. Second, organizations should establish dedicated AI governance structures, including ethics committees, governance boards, and multidisciplinary review teams responsible for overseeing ethical decision-making, monitoring compliance, and conducting periodic assessments of machine learning projects.

Third, fairness assessments, bias testing, explainability evaluations, algorithmic audits, and documentation reviews should become routine organizational practices to support continuous ethical improvement rather than one-time compliance activities (Tariq et al., 2025). Fourth, software engineers, AI specialists, project

managers, and technology leaders should receive continuous professional development in AI ethics, responsible software engineering, and regulatory compliance to strengthen organizational ethical capability (Hanif et al., 2026). Finally, governments, universities, professional associations, and software organizations should strengthen collaborative partnerships to develop context-specific AI governance frameworks, standardized ethical guidelines, and industry best practices that support responsible innovation while remaining aligned with national digital transformation priorities in Pakistan and Saudi Arabia (Khan, 2025; Alsumari & Alrwais, 2025).

Limitations and Future Research Directions

Despite its contributions, this study has several limitations. First, the research employed a qualitative design involving sixteen participants; therefore, the findings provide in-depth contextual insights but should not be interpreted as statistically generalizable across all software organizations. Second, the study focused exclusively on software professionals from Pakistan and Saudi Arabia, limiting the transferability of findings to other geographical, cultural, and institutional settings. Third, the research examined the perspectives of software engineers, AI specialists, software architects, and technology managers without incorporating the views of policymakers, regulators, business executives, or end users who also influence ethical AI implementation. Finally, the cross-sectional nature of the study captured participants' experiences at a single point in time and therefore could not examine how organizational governance and ethical practices evolve as machine learning technologies mature.

Future research should employ mixed-methods and longitudinal designs to examine the evolution of responsible machine learning practices across different stages of organizational AI maturity. Comparative investigations involving additional countries from the Middle East, South Asia, Africa, and other emerging digital economies would further enrich understanding of cross-cultural socio-technical governance. Future studies should also investigate sector-specific ethical challenges within healthcare, finance, education, cybersecurity, manufacturing, and public administration, where AI systems have significant societal implications. Given the rapid emergence of generative AI and AI-assisted software engineering, researchers should explore their influence on developer responsibility, organizational governance, software quality, and ethical decision-making. Finally, future research should develop and empirically validate socio-technical Ethical AI Maturity Models capable of assessing organizational readiness, governance capability, and responsible AI implementation across the Software Development Life Cycle.

REFERENCES

Abbas, T., Rathore, S. A., Turki, A., Khan, S., Alghushairy, O., & Daud, A. (2025). Enhancing Software Engineering With AI: Innovations, Challenges, and Future Directions. *IET Software*, *2025*(1), 5691460.

Ain, Q. U., Haq, M., Jilani, A. A. A., & Sohail, K. (2025). A Comparison Between Traditional Software Engineering Practices and AI-Driven Methodologies. In *Generative AI in Software Engineering* (pp. 57-92). IGI Global Scientific Publishing.

Alahmari, N., Alswedani, S., Alzahrani, A., Katib, I., Albeshri, A., & Mehmood, R. (2022). Musawah: a data-driven ai approach and tool to co-create healthcare services with a case study on cancer disease in Saudi Arabia. *Sustainability*, *14*(6), 3313.

Alenezi, M., & Akour, M. (2025). Ai-driven innovations in software engineering: a review of current practices and future directions. *Applied Sciences*, *15*(3), 1344.

Aljawawdeh, H., Fawareh, H., Al Refai, M., Alshraideh, F., Salman, R., Dabaa't, S. A., ... & Khouj, M. (2024, December). Ethical considerations in AI-driven HCI systems in software engineering. In *2024 25th International Arab Conference on Information Technology (ACIT)* (pp. 1-6). IEEE.

Alsahafi, R., Alzahrani, A., & Mehmood, R. (2023). Smarter sustainable tourism: data-driven multi-perspective parameter discovery for autonomous design and operations. *Sustainability*, *15*(5), 4166.

Alsehani, F. N. (2026). Harnessing Artificial Intelligence and Digital Technologies for Sustainable Healthcare Delivery in Saudi Arabia: A Comprehensive Review, Issues, and Future Perspectives. *Sustainability*, *18*(3), 1461.

AlShwaier, A., & Abdullah, M. I. (2025). Artificial Intelligence and Innovation Environments: Enhancing Startup Performance in the Kingdom of Saudi Arabia. *International Journal on Management Education and Emerging Technology (IJMEET)*, *3*(2), 1-15.

Al-Sulbi, K., & Attaallah, A. (2025). Symmetric bug prediction in software requirement by machine learning algorithms. *Scientific Reports*, *15*(1), 38276.

Alsumari, W., & Alrwais, O. (2025). Artificial Intelligence Project Management Methodologies: Insights from Field Experts in Saudi Arabia. *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, *6*(2), 246-258.

Alvear, A., Carvajal, C. L., Vergara, I., & Msalmah, Y. (2025, March). Software Engineering and Society: Navigating the Ethical Challenges of Artificial Intelligence. In *World Conference on Information Systems and Technologies* (pp. 483-493). Cham: Springer Nature Switzerland.

Al-Zahrani, A. M., & Alasmari, T. M. (2024). Exploring the impact of artificial intelligence on higher education: The dynamics of ethical, social, and educational implications. *Humanities and social sciences communications*, *11*(1), 1-12.

Amugongo, L. M., Kriebitz, A., Boch, A., & Lütge, C. (2025). Operationalising AI ethics through the agile software development lifecycle: a case study of AI-enabled mobile health applications. *AI and Ethics*, *5*(1), 227-244.

Ashraf, N., Arzu, F., Abrar, K., & Anwar, M. (2025). Narratives of SMEs on Access to Finance: Barriers and Opportunities in Pakistan's Banking Sector. *The Critical Review of Social Sciences Studies*, *3*(4), 262-277.

Barefah, A., & Altalhi, M. (2024). Exploring the critical benefits and challenges of social network site-based requirement elicitation in Saudi Arabia. *Sustainability*, *16*(22), 9794.

Braun, V., Clarke, V., & Hayfield, N. (2022). 'A starting point for your journey, not a map': Nikki Hayfield in conversation with Virginia Braun and Victoria Clarke about thematic analysis. *Qualitative research in psychology*, *19*(2), 424-445.

Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & quantity*, *56*(3), 1391-1412.

Hanif, I., Hussain, M., Ibrar, M., Hafeez, A., & Yasin, A. (2026). Secure AI Applications for Cybersecurity: Ethical and Explainable Threat Detection in Modern Software Systems. In *Ethical AI Applications for Software Engineering and Development* (pp. 275-304). IGI Global Scientific Publishing.

Hole, L. (2024). Handle with care; considerations of Braun and Clarke's approach to thematic analysis. *Qualitative research journal*, *24*(4), 371-383.

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

Khan, J. A. (2025). Governing Machine Learning, Data Engineering and Data Science in Healthcare: Ethical, Practical, and Regulatory Implications for Saudi Vision 2030. *Review of Education, Administration & Law*, *8*(4), 525-541.

Khan, R. A. S., Abrar, K., Parveen, S., Bhurgri, S. S., & Panhwar, A. J. (2026). *Integrating generative AI into software development lifecycles: A comparative qualitative study of engineering practices in Pakistan and the USA. The Asian Bulletin of Big Data Management,* 6(1), 329–345.

Mohamad, M., Chowdhury, D., Russell, C., Balikel, A., & Kawalek, P. (2025). How GenAI transforms computer engineering education: The case of the Middle East and North Africa. *Journal of Engineering Education Transformations*, 3-19.

Oyeniran, C. O., Adewusi, A. O., Adeleke, A. G., Akwawa, L. A., & Azubuko, C. F. (2022). Ethical AI: Addressing bias in machine learning models and software applications. *Computer Science & IT Research Journal*, *3*(3), 115-126.

Razzaq, K., & Shah, M. (2025). Advancing cybersecurity through machine learning: A scientometric analysis of global research trends and influential contributions. *Journal of Cybersecurity and Privacy*, *5*(2), 12.

Rehmat, M. A., Hassan, H., Khan, H. A., & Abrar, K. (2025). Artificial Intelligence in the Classroom: Teachers' Lived Experiences and Ethical Concerns in Educational Integration. *ASSAJ*, *4*(02), 629-645.

Sohag, S. R., Pasha, S. M. M., & Ali, M. M. (2023). Advanced approaches to achieve adaptive ethical and AI-driven human-centric software engineering. *Internafional Journal of Scienfific Research in Modern Science and Technology*, *2*, 19-31.

Tariq, B., Ashraf, M. R., & Rashid, U. (2025). Ethical Imperatives in AI design: a comprehensive framework for risk mitigation and responsible innovation. *Ubiquitous Technology Journal*, *1*(2), 61-73.