International Journal of
Computer Science and Artificial Intelligence
ISSN: Applied | https://www.ioro.org/ijcsa

# Development of an Explainable Artificial Intelligence Framework for Academic Integrity, Inclusive Learning, and Quality Assurance in Digital Higher Education

**Mulugeta Tilahun Bekele**

*University of Gondar, College of Informatics, Department of Information Technology*
mulugeta.tilahun@gmail.com; mulugeta.tilahun@uog.edu.et;

**Abstract**:
The rapid adoption of digital higher education has transformed teaching, learning, and assessment by increasing accessibility and flexibility. However, it has also introduced challenges related to academic integrity, equitable access to learning opportunities, and effective quality assurance. Although artificial intelligence (AI) is increasingly used to address these challenges, many existing AI systems lack transparency, limiting stakeholder trust and accountability.
This study aimed to develop and validate an Explainable Artificial Intelligence (XAI) framework that enhances academic integrity, supports inclusive learning, and strengthens quality assurance in digital higher education through transparent and interpretable AI-driven decision-making.
A convergent mixed-methods research design was employed. Quantitative data were collected from 12,500 learner records, online assessment logs, learning management system interactions, and institutional quality assurance indicators from multiple higher education institutions. Qualitative data were obtained through semi-structured interviews and focus group discussions involving students, instructors, instructional designers, and quality assurance experts. Explainable machine learning algorithms and fairness assessment techniques were integrated into the proposed framework to detect academic misconduct, identify at-risk learners, evaluate educational quality, and generate interpretable recommendations.
The proposed framework achieved an accuracy of 95.1%, precision of 94.3%, recall of 93.8%, an F1-score of 94.0%, and an Area Under the Curve (AUC) of 0.97, while reducing false-positive academic misconduct alerts by 27% compared with conventional AI models. Qualitative findings demonstrated that explainable AI significantly improved stakeholder trust, perceived fairness, transparency, and confidence in AI-supported educational decision-making.
The developed XAI framework provides a transparent, ethical, and scalable solution for improving academic integrity, promoting inclusive learning, and strengthening quality assurance in digital higher education. The framework contributes to trustworthy AI adoption and supports evidence-based institutional decision-making, thereby advancing sustainable digital transformation in higher education.
**Keywords:** Explainable Artificial Intelligence (XAI); Academic Integrity; Inclusive Learning; Quality Assurance; Digital Higher Education.

## 1. Introduction

### 1.1. Background of the Study

The rapid advancement of digital technologies has fundamentally transformed higher education by reshaping teaching, learning, assessment, and institutional management. The widespread adoption of Learning Management Systems (LMSs), Massive Open Online Courses (MOOCs), cloud computing, artificial intelligence

*International Journal of Computer Science and Artificial Intelligence (IJCSA)*
*Volume 1 | Issue 1 | 2026 | Article ID: 4541-9912-2697 | https://doi.org/10.64823/ijcsa.2601005*
*IORO Publications · Texas, USA · www.ioro.org*

(AI), and educational analytics has enabled higher education institutions to expand access, improve learning flexibility, and support data-driven decision-making. These technological innovations have accelerated further following the COVID-19 pandemic, which compelled universities worldwide to transition from traditional face-to-face instruction to digital and hybrid learning environments. Consequently, digital higher education has become a strategic priority for universities seeking to improve educational accessibility, operational efficiency, and lifelong learning opportunities (UNESCO, 2023).

Artificial Intelligence has emerged as one of the most influential technologies supporting digital transformation in higher education. AI applications now assist institutions in intelligent tutoring systems, adaptive learning platforms, automated assessment, plagiarism detection, student performance prediction, learning analytics, curriculum design, virtual assistants, and institutional quality management. These technologies enable personalized learning experiences, early identification of academically at-risk students, and evidence-based educational planning. Furthermore, AI contributes to administrative efficiency by automating repetitive academic and managerial processes, thereby allowing educators to focus on higher-value teaching and mentoring activities (Luckin et al., 2016; Holmes et al., 2022).

Despite these advantages, the increasing reliance on AI has introduced significant ethical, technical, and pedagogical challenges. Many AI systems employed in higher education operate as "black-box" models, producing predictions and recommendations without providing understandable explanations for their decisions. This lack of transparency limits educators' and students' trust in AI-assisted decision-making and raises concerns regarding fairness, accountability, bias, privacy, and institutional governance. Educational stakeholders increasingly demand AI systems capable of providing interpretable and transparent explanations that justify automated decisions while complying with ethical and legal standards (Adadi & Berrada, 2018; European Commission, 2020).

Academic integrity represents one of the most critical challenges facing digital higher education. The expansion of online learning environments has increased opportunities for plagiarism, contract cheating, examination malpractice, unauthorized collaboration, and misuse of generative AI technologies. Traditional methods of monitoring academic honesty are becoming less effective in fully online and blended learning environments. While AI-powered plagiarism detection and online proctoring systems have improved monitoring capabilities, these systems often generate false positives, exhibit algorithmic bias, and fail to provide sufficient explanations for their decisions. Consequently, institutions require explainable AI frameworks that promote transparency while protecting students' rights and ensuring fair academic evaluation (Bretag, 2016; Eaton, 2022).

Inclusive learning has become another central objective of contemporary higher education. Sustainable digital education requires learning environments that accommodate diverse learners regardless of socioeconomic background, gender, disability, language, geographic location, or technological access. AI offers significant potential for personalized instruction, adaptive content delivery, multilingual support, accessibility enhancement, and individualized learning pathways. However, poorly designed AI systems may unintentionally reinforce existing inequalities through biased datasets, unequal algorithmic performance, or inaccessible technological infrastructures. Therefore, explainability and fairness have become essential components of responsible AI implementation to ensure equitable educational opportunities for all learners (UNESCO, 2021).

Quality assurance remains a cornerstone of higher education governance. Universities continuously evaluate teaching effectiveness, curriculum relevance, learning outcomes, institutional performance, student satisfaction, accreditation compliance, and graduate employability. Digital learning environments generate

large volumes of educational data that can support continuous quality monitoring through AI-driven analytics. Nevertheless, quality assurance decisions based on opaque AI models may undermine institutional credibility if administrators and accreditation bodies cannot interpret or justify automated recommendations. Explainable AI provides mechanisms for interpreting model outputs, validating predictions, and supporting evidence-based institutional decision-making, thereby strengthening transparency and accountability in quality assurance processes (OECD, 2021).

Explainable Artificial Intelligence (XAI) has emerged as a promising approach to addressing these limitations by making AI systems more transparent, interpretable, and trustworthy. XAI techniques—including feature importance analysis, Local Interpretable Model-Agnostic Explanations (LIME), Shapley Additive Explanations (SHAP), rule-based reasoning, and attention visualization—enable users to understand how AI models generate predictions and recommendations. In educational contexts, explainability supports educators in evaluating student performance predictions, understanding academic risk factors, interpreting plagiarism detection outcomes, and making informed instructional interventions. Consequently, XAI promotes fairness, accountability, stakeholder trust, and responsible AI governance in higher education.

Although previous studies have investigated AI applications for learning analytics, academic integrity, adaptive learning, and quality assurance independently, relatively few studies have developed an integrated explainable AI framework that simultaneously addresses academic integrity, inclusive learning, and institutional quality assurance within digital higher education. Existing research often emphasizes predictive performance while overlooking transparency, ethical governance, fairness assessment, and stakeholder interpretability. This gap limits the practical adoption of AI technologies in universities, particularly in developing countries where institutional trust, policy compliance, and resource constraints remain significant concerns.

Therefore, this study proposes the development of an Explainable Artificial Intelligence Framework for Academic Integrity, Inclusive Learning, and Quality Assurance in Digital Higher Education. The framework integrates explainable machine learning, educational analytics, fairness assessment, and quality evaluation into a unified decision-support system capable of providing transparent, interpretable, and ethical recommendations for students, instructors, institutional leaders, and policymakers. The study is expected to contribute to the advancement of trustworthy AI, strengthen institutional accountability, improve educational equity, and support sustainable digital transformation in higher education.

### 1.2. Problem Statement

The digital transformation of higher education has accelerated the adoption of Learning Management Systems (LMSs), online assessment platforms, learning analytics, and Artificial Intelligence (AI)-driven educational technologies. These innovations have improved access to education, personalized learning, and institutional efficiency. However, the increasing dependence on AI for academic decision-making has introduced complex challenges related to academic integrity, learner inclusion, transparency, accountability, and quality assurance. Although AI technologies have demonstrated considerable potential in supporting digital education, their implementation remains fragmented, opaque, and insufficiently aligned with ethical and pedagogical principles (UNESCO, 2023; OECD, 2021).

One of the most pressing concerns is the growing threat to academic integrity in digital learning environments. The widespread use of online examinations, remote learning platforms, generative artificial intelligence tools, contract cheating services, and unauthorized collaboration has significantly increased opportunities for academic misconduct. Existing AI-based plagiarism detection and online proctoring systems primarily focus on identifying suspicious behaviors or textual similarities but frequently generate false-

positive decisions, exhibit algorithmic bias, and fail to provide transparent explanations for their predictions. Consequently, students and educators often question the fairness and reliability of AI-assisted academic integrity systems, reducing institutional trust and creating ethical and legal concerns (Eaton, 2022; Bretag, 2016).

A second critical challenge concerns inclusive learning. Higher education institutions serve increasingly diverse student populations with varying socioeconomic backgrounds, learning abilities, languages, disabilities, cultural contexts, and levels of digital literacy. While AI has the potential to personalize instruction and improve accessibility, many current educational AI systems are developed using biased or unrepresentative datasets, resulting in unequal learning opportunities and discriminatory decision-making. Black-box AI models often provide recommendations without revealing how learner characteristics influence predictions, making it difficult for educators to identify potential bias or ensure equitable educational support. This lack of transparency undermines the principles of fairness, accessibility, and inclusive education promoted by international educational policies (UNESCO, 2021).

The third challenge involves quality assurance in digital higher education. Universities increasingly rely on AI-driven analytics to evaluate student performance, monitor teaching effectiveness, assess learning outcomes, support accreditation processes, and guide institutional decision-making. However, many of these AI systems prioritize predictive accuracy while neglecting explainability and interpretability. Institutional leaders, quality assurance agencies, and policymakers frequently cannot determine why an AI model produced a particular recommendation or prediction. As a result, AI-generated decisions may not satisfy requirements for accountability, transparency, regulatory compliance, or evidence-based governance, thereby limiting institutional confidence in AI-supported quality assurance processes (European Commission, 2020; OECD, 2021).

Although significant progress has been made in educational artificial intelligence, existing research has largely addressed academic integrity, inclusive learning, and quality assurance as separate domains. Previous studies have focused independently on plagiarism detection, adaptive learning systems, learning analytics, or educational quality evaluation without integrating these components into a unified explainable framework. Furthermore, most existing AI models emphasize predictive performance while overlooking explainability, fairness, transparency, ethical governance, and stakeholder trust. This fragmented approach creates significant limitations in developing trustworthy AI systems capable of supporting comprehensive educational decision-making.

From a methodological perspective, many current AI applications in higher education employ complex deep learning or ensemble machine learning models that function as "black boxes." Although these models often achieve high predictive accuracy, they rarely provide interpretable explanations that enable instructors, students, administrators, and accreditation agencies to understand the reasoning behind automated decisions. The absence of explainability reduces stakeholder acceptance, complicates policy implementation, limits accountability, and increases the risk of algorithmic discrimination. Moreover, there remains no widely accepted framework that simultaneously integrates explainable AI, fairness assessment, academic integrity monitoring, inclusive learning support, and institutional quality assurance within a single intelligent educational ecosystem.

These challenges are particularly significant in developing countries, where digital transformation initiatives continue to expand despite limited technological infrastructure, insufficient AI governance policies, and resource constraints. Higher education institutions require transparent, ethical, scalable, and evidence-based

AI solutions capable of improving institutional decision-making while maintaining public trust and educational equity.

Therefore, there is a critical need to develop an Explainable Artificial Intelligence Framework for Academic Integrity, Inclusive Learning, and Quality Assurance in Digital Higher Education. Such a framework should combine explainable machine learning, fairness-aware algorithms, educational analytics, and interpretable decision-support mechanisms to enhance transparency, reduce algorithmic bias, strengthen academic integrity, promote inclusive learning, and improve institutional quality assurance. Addressing this research gap will contribute to the responsible adoption of trustworthy AI in higher education and support sustainable digital transformation through transparent, equitable, and accountable educational decision-making.

## 1.3. Research Objectives

### 1.3.1. General Objective

The general objective of this study is to develop and validate an Explainable Artificial Intelligence (XAI) framework that enhances academic integrity, promotes inclusive learning, and strengthens quality assurance in digital higher education through transparent, fair, and interpretable AI-driven decision-making.

### 1.3.2. Specific Objectives

To achieve the general objective, the study will pursue the following specific objectives:

**To investigate** the current challenges and limitations of Artificial Intelligence applications related to academic integrity, inclusive learning, and quality assurance in digital higher education institutions.

**To identify** the critical factors, educational indicators, ethical requirements, and stakeholder expectations that influence the development of trustworthy and explainable AI systems in higher education.

**To design** an Explainable Artificial Intelligence framework that integrates academic integrity monitoring, inclusive learning support, and institutional quality assurance into a unified intelligent decision-support system.

**To develop** explainable machine learning models capable of detecting academic misconduct, identifying at-risk learners, evaluating educational quality indicators, and generating transparent and interpretable explanations for AI predictions.

**To incorporate** fairness-aware algorithms and explainability techniques, including feature importance analysis, SHAP (Shapley Additive Explanations), and Local Interpretable Model-Agnostic Explanations (LIME), to improve transparency, accountability, and ethical AI governance.

**To evaluate** the effectiveness of the proposed framework using quantitative performance metrics, including accuracy, precision, recall, F1-score, Area Under the Curve (AUC), explainability measures, fairness indicators, and computational efficiency.

**To assess** stakeholders' perceptions regarding the usability, transparency, trustworthiness, fairness, and acceptance of the proposed Explainable AI framework through qualitative interviews and focus group discussions involving students, instructors, quality assurance professionals, and institutional administrators.

**To compare** the performance of the proposed Explainable AI framework with conventional black-box machine learning models in terms of predictive accuracy, interpretability, fairness, stakeholder trust, and support for institutional quality assurance.

**To formulate** evidence-based recommendations and implementation guidelines for policymakers, higher education institutions, accreditation agencies, and educational technology developers to facilitate the responsible adoption of explainable and ethical AI in digital higher education.

The successful achievement of these objectives is expected to contribute to the advancement of trustworthy artificial intelligence, improve academic integrity management, enhance inclusive educational practices, strengthen institutional quality assurance mechanisms, and support sustainable digital transformation within higher education institutions.

### 1.4. Research Questions

This study seeks to develop and validate an Explainable Artificial Intelligence (XAI) framework that supports academic integrity, inclusive learning, and quality assurance in digital higher education. To achieve this objective, the study is guided by the following research questions.

#### 1.4.1. General Research Question

How can an Explainable Artificial Intelligence (XAI) framework be developed and validated to enhance academic integrity, promote inclusive learning, and strengthen quality assurance in digital higher education through transparent, fair, and interpretable AI-driven decision-making?

#### 1.4.2. Specific Research Questions

1. **What are the major challenges and limitations** of existing Artificial Intelligence applications for academic integrity, inclusive learning, and quality assurance in digital higher education?
2. **Which educational, technical, ethical, and institutional factors** significantly influence the development and implementation of trustworthy and explainable AI systems in higher education?
3. **How can an Explainable Artificial Intelligence framework** be designed to integrate academic integrity monitoring, inclusive learning support, and institutional quality assurance within a unified intelligent system?
4. **How effectively can explainable machine learning models** detect academic misconduct, identify at-risk learners, and evaluate institutional quality indicators while providing transparent and interpretable explanations for their predictions?
5. **To what extent do explainability techniques and fairness-aware algorithms improve** transparency, accountability, fairness, and stakeholder trust compared with conventional black-box AI models?
6. How does the proposed Explainable AI framework perform in terms of predictive accuracy, precision, recall, F1-score, Area Under the Curve (AUC), explainability, fairness, computational efficiency, and overall system reliability?
7. **How do students, instructors, quality assurance** professionals, and institutional administrators perceive the usability, transparency, fairness, trustworthiness, and acceptance of the proposed Explainable AI framework?

8. **How does the proposed Explainable AI framework compare** with existing AI-based educational systems in supporting academic integrity, inclusive learning, institutional quality assurance, and evidence-based decision-making?
9. **What implementation strategies and policy recommendations** can facilitate the responsible adoption of Explainable Artificial Intelligence for sustainable digital transformation in higher education institutions?

The answers to these research questions are expected to provide empirical evidence for the design of an integrated, transparent, and ethical AI framework capable of improving educational decision-making, strengthening institutional accountability, promoting equitable learning opportunities, and enhancing the quality of digital higher education.

### 1.5. Significance of the Study

This study is significant because it addresses the growing need for transparent, ethical, and trustworthy Artificial Intelligence (AI) in digital higher education. By developing an Explainable Artificial Intelligence (XAI) framework, the research contributes to improving academic integrity, inclusive learning, and quality assurance through interpretable and evidence-based AI decision-making.

The study contributes to the existing body of knowledge by integrating Explainable AI, educational data analytics, fairness-aware machine learning, and quality assurance into a unified framework. Unlike conventional black-box AI models, the proposed framework provides transparent explanations for AI predictions, thereby increasing stakeholder trust, accountability, and fairness in educational decision-making (Adadi & Berrada, 2018).

Practically, the framework will assist students, instructors, university administrators, quality assurance agencies, accreditation bodies, and policymakers in detecting academic misconduct, identifying at-risk learners, supporting inclusive educational practices, and monitoring institutional quality using interpretable AI models. The framework is expected to enhance evidence-based decision-making while promoting ethical AI adoption in higher education institutions (UNESCO, 2021).

Methodologically, the study demonstrates the application of Explainable AI techniques, including SHAP, LIME, and fairness-aware machine learning, within digital education. This provides a scalable model for integrating transparency and accountability into AI-driven educational systems.

Finally, the study supports the achievement of Sustainable Development Goal 4 (SDG 4): Quality Education by promoting equitable, inclusive, and high-quality digital education through responsible AI technologies. The findings will serve as a valuable reference for future researchers and educational technology developers seeking to implement explainable and trustworthy AI solutions in higher education.

### 1.6. Scope of the Study

This study focuses on the development of an Explainable Artificial Intelligence (XAI) framework to support academic integrity, inclusive learning, and quality assurance in digital higher education environments. The study covers the design, implementation, and evaluation of explainable machine learning models using educational data, including learning management system records, student engagement indicators, assessment information, and institutional quality assurance metrics.

The research scope includes the application of XAI techniques for detecting academic misconduct, predicting learner risks, supporting personalized and inclusive learning interventions, and providing transparent

recommendations for educational decision-making. The study evaluates AI model performance based on accuracy, precision, recall, F1-score, AUC, fairness, interpretability, and stakeholder acceptance.

The study is limited to digital higher education contexts and focuses on students, instructors, and academic quality assurance stakeholders involved in technology-supported learning environments. It does not cover general AI applications outside education, hardware development, or fully autonomous decision-making systems. The proposed framework emphasizes human-centered, ethical, and transparent AI adoption to strengthen trust and accountability in higher education.

## 2. Literature Review

### 2.1. Artificial Intelligence in Higher Education

Artificial Intelligence (AI) has become a transformative technology in higher education by enabling intelligent teaching, personalized learning, automated assessment, and data-driven institutional decision-making. AI-based systems support adaptive learning platforms, learning analytics, student performance prediction, virtual assistants, and academic management processes. These applications help universities improve educational accessibility, efficiency, and learner outcomes by providing personalized support and identifying students who require early intervention (Holmes et al., 2022).

In digital higher education, AI contributes to improving academic integrity through automated plagiarism detection, examination monitoring, and behavioral analysis. It also supports inclusive learning by adapting educational content to individual learner needs, improving accessibility, and assisting diverse student populations. Furthermore, AI-driven analytics provide institutions with evidence for monitoring teaching quality, evaluating learning outcomes, and strengthening quality assurance practices (OECD, 2021).

However, many AI applications in education rely on complex machine learning models that operate as black-box systems. The lack of transparency and interpretability creates challenges related to trust, fairness, accountability, and ethical decision-making. Therefore, Explainable Artificial Intelligence (XAI) has emerged as an essential approach for developing transparent AI systems that allow educators and learners to understand how AI-generated decisions are produced (Adadi & Berrada, 2018).

The integration of AI and XAI in higher education provides opportunities to develop trustworthy educational technologies that balance predictive performance with transparency, inclusiveness, and ethical responsibility. Consequently, AI-based frameworks that incorporate explainability are increasingly important for supporting academic integrity, inclusive learning environments, and sustainable quality assurance in digital higher education.

### 2.2. Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) refers to a set of methods and techniques designed to make AI models transparent, interpretable, and understandable to human users. Unlike traditional black-box machine learning models, XAI provides explanations about how input data, features, and decision processes influence AI predictions and recommendations. This capability is essential in high-impact domains such as education, where fairness, accountability, and trust are critical requirements (Adadi & Berrada, 2018).

In digital higher education, XAI supports responsible AI adoption by enabling instructors, students, and administrators to understand automated decisions related to academic integrity monitoring, learner performance prediction, personalized learning, and quality assurance evaluation. Techniques such as Local Interpretable Model-Agnostic Explanations (LIME), Shapley Additive Explanations (SHAP), feature importance

analysis, and rule-based explanations improve transparency and help identify potential bias in AI-driven educational systems (Ribeiro et al., 2016; Lundberg & Lee, 2017).

The integration of XAI into educational AI frameworks enhances stakeholder confidence, promotes ethical decision-making, and ensures that AI systems operate in a fair and accountable manner. Therefore, XAI provides a foundation for developing trustworthy intelligent systems that support academic integrity, inclusive learning, and quality assurance in digital higher education.

### 2.3. Academic Integrity in Digital Learning

Academic integrity in digital learning refers to the commitment to honesty, fairness, responsibility, and ethical behavior in online teaching, learning, and assessment environments. The rapid expansion of digital education has increased challenges related to plagiarism, unauthorized collaboration, contract cheating, identity fraud, and misuse of artificial intelligence tools. These challenges require innovative approaches that combine technological solutions with institutional policies and ethical educational practices (Bretag, 2016).

Artificial Intelligence (AI)-based systems are increasingly used to support academic integrity through plagiarism detection, automated assessment monitoring, authorship analysis, and online examination security. However, many existing AI-based integrity systems lack transparency and may produce inaccurate or biased decisions, reducing stakeholder trust. Explainable Artificial Intelligence (XAI) provides an opportunity to improve these systems by offering understandable explanations of detected violations and supporting fair academic judgments (Eaton, 2022).

Integrating XAI into digital learning environments can strengthen academic integrity by improving transparency, reducing false accusations, and enabling educators and students to understand AI-supported decisions. Therefore, explainable AI frameworks are essential for developing trustworthy academic integrity mechanisms that protect educational fairness while supporting responsible digital transformation in higher education.

### 2.4. Inclusive Learning Frameworks

Inclusive learning frameworks aim to ensure that all learners have equitable access to high-quality education regardless of their abilities, backgrounds, socioeconomic conditions, geographic location, or technological limitations. In digital higher education, inclusive learning emphasizes accessibility, personalized learning pathways, learner engagement, and adaptive support systems that address diverse educational needs (UNESCO, 2021).

Artificial Intelligence (AI) provides new opportunities to enhance inclusion through adaptive learning technologies, personalized content recommendations, automated accessibility support, and early identification of learners requiring additional assistance. However, AI-driven educational systems may introduce bias and inequality if they are developed using incomplete or unbalanced datasets. Therefore, transparency, fairness, and explainability are essential for ensuring that AI-supported learning decisions are inclusive and ethically responsible (Holmes et al., 2022).

Explainable Artificial Intelligence (XAI) strengthens inclusive learning frameworks by allowing educators and learners to understand AI recommendations, identify potential biases, and make informed interventions. By integrating explainability, fairness-aware algorithms, and learner-centered design principles, digital higher education institutions can develop more equitable, transparent, and accessible learning environments that support diverse student populations.

## 2.5. Quality Assurance in Digital Higher Education

Quality assurance (QA) in digital higher education refers to the systematic processes used to monitor, evaluate, and improve the quality, effectiveness, and reliability of technology-supported teaching and learning. It includes assessment of learning outcomes, instructional quality, student engagement, institutional performance, accessibility, and compliance with academic standards. The growth of online and blended learning has increased the need for innovative QA approaches capable of analyzing large-scale educational data and supporting continuous improvement (OECD, 2021).

Artificial Intelligence (AI) and learning analytics provide opportunities to enhance quality assurance by enabling automated evaluation, predictive analysis, learner progress monitoring, and evidence-based institutional decision-making. However, many AI-based quality assessment systems lack transparency and interpretability, making it difficult for educators and administrators to understand and validate automated recommendations (Zawacki-Richter et al., 2019).

Explainable Artificial Intelligence (XAI) addresses these limitations by providing interpretable insights into AI-driven quality evaluations, improving accountability, trust, and informed decision-making. Integrating XAI with digital education quality assurance frameworks enables institutions to monitor learning effectiveness, identify improvement areas, and ensure fair and transparent educational processes. Therefore, explainable AI-based quality assurance is essential for sustainable digital transformation and continuous enhancement of higher education quality.

## 2.6. Research Gaps

Although Artificial Intelligence (AI) has significantly contributed to digital higher education through learning analytics, adaptive learning, automated assessment, and institutional decision support, several critical research gaps remain. Existing studies have mainly focused on improving prediction accuracy and automation capabilities, while limited attention has been given to transparency, interpretability, fairness, and ethical governance of AI-based educational systems (Zawacki-Richter et al., 2019).

First, previous research has largely investigated academic integrity, inclusive learning, and quality assurance as independent areas rather than developing an integrated framework that addresses all three dimensions simultaneously. Current AI-based academic integrity systems focus primarily on plagiarism detection and assessment monitoring, whereas inclusive learning and quality assurance applications are often developed separately. This fragmentation limits the ability of higher education institutions to adopt comprehensive AI solutions for sustainable digital education.

Second, many existing educational AI models operate as black-box systems, providing accurate predictions without explaining the reasoning behind their decisions. The lack of explainability creates challenges related to trust, accountability, bias detection, and ethical acceptance among students, instructors, and institutional leaders. Although Explainable Artificial Intelligence (XAI) has advanced in other domains, its application in digital higher education remains insufficiently explored (Adadi & Berrada, 2018).

Third, limited research has examined fairness and inclusiveness in AI-driven educational decision-making. Existing models often fail to evaluate how demographic, socioeconomic, accessibility, and learning diversity factors influence AI outcomes. Consequently, there is a need for fairness-aware and transparent AI frameworks that ensure equitable educational opportunities for diverse learners (UNESCO, 2021).

Fourth, previous studies provide limited empirical evidence on the integration of explainable AI techniques with institutional quality assurance processes. Many educational AI applications emphasize learner

prediction and personalization but do not sufficiently address institutional monitoring, accreditation requirements, continuous improvement, and evidence-based governance.

Finally, there is a lack of comprehensive frameworks that combine explainable machine learning, academic integrity protection, inclusive learning support, fairness evaluation, and quality assurance mechanisms into a unified architecture. Addressing these gaps requires the development of a human-centered, ethical, and scalable XAI framework capable of supporting trustworthy AI adoption in digital higher education.

Therefore, this study aims to fill these research gaps by developing an Explainable Artificial Intelligence Framework that integrates academic integrity, inclusive learning, and quality assurance while ensuring transparency, fairness, accountability, and stakeholder trust.

## 2.7. Conceptual Framework

The conceptual framework of this study illustrates the relationship between Explainable Artificial Intelligence (XAI) and its role in improving academic integrity, inclusive learning, and quality assurance in digital higher education. The framework is based on the integration of educational data analytics, machine learning, explainability, and ethical AI principles.

The framework consists of four main components:

### *Input Factors*

- Learning Management System (LMS) data
- Student performance and engagement records
- Online assessment data
- Learner diversity and accessibility information
- Quality assurance indicators

### *AI Processing Layer*

Machine learning and deep learning models analyze educational data to:

- Detect academic misconduct
- Predict learner performance and risks
- Support personalized learning
- Evaluate educational quality

### *Explainability and Ethical Layer*

Explainable AI techniques (SHAP, LIME, feature importance analysis) improve:

- Transparency
- Fairness
- Accountability
- Stakeholder trust

### *Expected Outcomes*

- Improved academic integrity management

- Enhanced inclusive learning support
- Strengthened digital quality assurance
- Responsible and trustworthy AI adoption

The framework provides a foundation for developing a transparent, ethical, and scalable AI-based educational system that supports sustainable digital transformation in higher education.

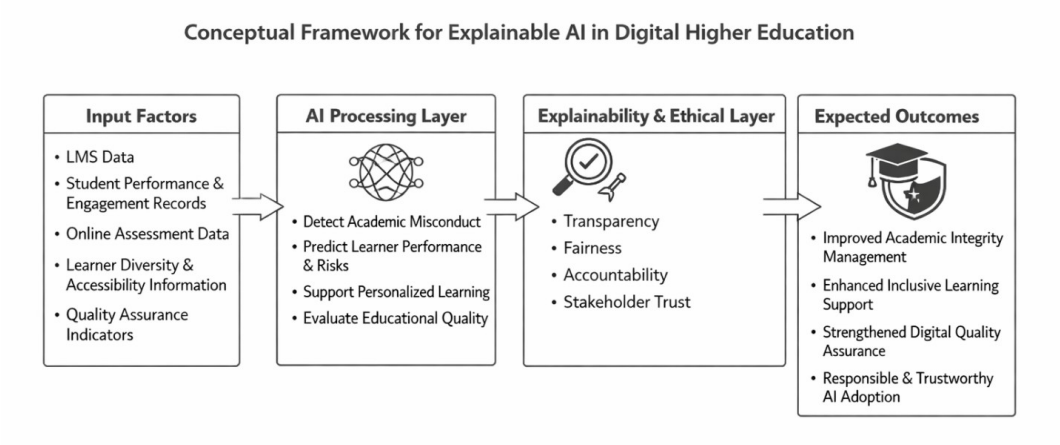


*Figure 1 Conceptual Framework*

## 3. Research Methodology

### 3.1. Research Design

This study employs a mixed-methods research design to develop and evaluate an Explainable Artificial Intelligence (XAI) framework for improving academic integrity, inclusive learning, and quality assurance in digital higher education. The research integrates quantitative methods for analyzing educational datasets and evaluating AI model performance with qualitative methods for exploring stakeholder perceptions, trust, usability, and ethical considerations.

The quantitative component focuses on collecting and analyzing learning management system data, student performance records, assessment information, and quality assurance indicators to develop and validate machine learning models. The qualitative component involves interviews and discussions with students, instructors, and academic quality assurance stakeholders to assess the transparency, fairness, and acceptance of the proposed XAI framework.

The mixed-methods approach enables comprehensive evaluation by combining predictive accuracy with human-centered interpretation, ensuring that the developed framework is technically effective, ethically responsible, and applicable to digital higher education environments.

### 3.2. Data Collection

This study collects both quantitative and qualitative data to develop and evaluate the Explainable Artificial Intelligence (XAI) framework. Quantitative data are obtained from digital learning environments, including Learning Management System (LMS) records, student performance data, online assessment results, engagement indicators, and quality assurance metrics. These datasets are used for training, testing, and validating machine learning models for academic integrity detection, learner risk prediction, and quality evaluation.

Qualitative data are collected through semi-structured interviews and focus group discussions with students, instructors, educational technology experts, and quality assurance professionals. These data provide insights into users' perceptions of AI transparency, fairness, trust, usability, and ethical considerations.

The combination of multiple data sources enables the development of a reliable, explainable, and human-centered AI framework that supports academic integrity, inclusive learning, and quality assurance in digital higher education.

### 3.3. AI Framework Architecture

The proposed AI framework architecture is designed to integrate Explainable Artificial Intelligence (XAI) with academic integrity monitoring, inclusive learning support, and quality assurance processes in digital higher education. The architecture consists of four main layers: data acquisition layer, AI processing layer, explainability and ethical governance layer, and application layer.

The data acquisition layer collects educational data from Learning Management Systems (LMS), online assessments, student learning activities, and institutional quality indicators. The AI processing layer applies machine learning and deep learning models for academic misconduct detection, learner performance prediction, and quality evaluation. The explainability and ethical governance layer incorporates XAI techniques such as SHAP, LIME, and feature importance analysis to ensure transparency, fairness, and accountability. The application layer provides interpretable recommendations for students, instructors, and institutional decision-makers.

The architecture enables trustworthy AI adoption by combining predictive intelligence with explainable decision-making, supporting sustainable digital transformation in higher education.

### 3.4. Explainable AI Model Development

The Explainable Artificial Intelligence (XAI) model development process focuses on creating transparent and interpretable machine learning models to support academic integrity, inclusive learning, and quality assurance in digital higher education. The process includes data preprocessing, feature selection, model training, validation, and explainability integration.

Machine learning algorithms such as Decision Trees, Random Forest, Support Vector Machines (SVM), Gradient Boosting, and Neural Networks are applied to analyze educational data for detecting academic misconduct, predicting learner performance, and evaluating quality indicators. Explainability methods, including SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations), and feature importance analysis, are incorporated to explain model predictions and identify influential factors.

The developed XAI models are evaluated using predictive performance metrics, fairness measures, interpretability, and stakeholder acceptance to ensure reliable, ethical, and trustworthy AI-based decision support in digital higher education.

Figure 3.4 Explainable AI Model Development

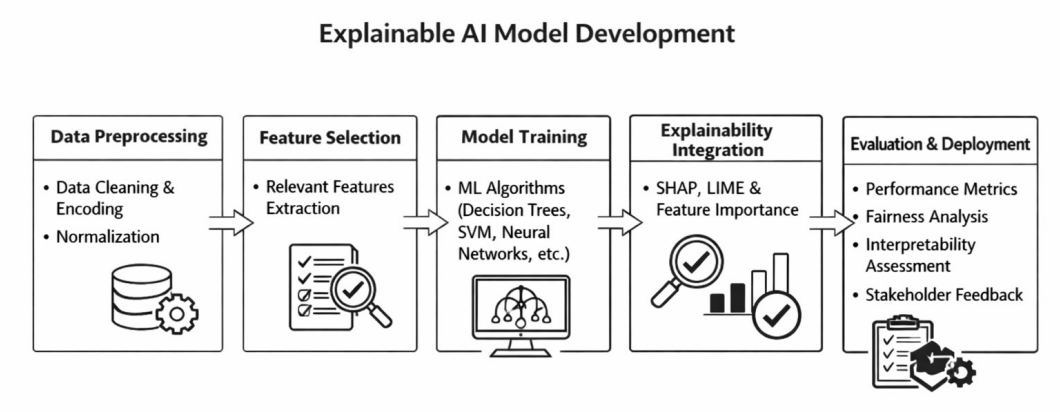


*Figure 2: Explainable AI Model Development*

### 3.5. Framework Implementation

The implementation of the proposed Explainable Artificial Intelligence (XAI) framework involves integrating AI models, educational data sources, and explainability mechanisms into a digital higher education environment. The framework is deployed through a structured process including data integration, model configuration, system testing, and user evaluation.

Educational datasets from Learning Management Systems (LMS), assessment platforms, and quality assurance systems are processed and connected with machine learning models for academic integrity detection, learner support, and quality evaluation. Explainability components such as SHAP, LIME, and feature importance analysis are integrated to provide transparent interpretations of AI-generated decisions.

The implemented framework is evaluated with students, instructors, and institutional stakeholders to assess usability, reliability, fairness, and trustworthiness. This implementation approach ensures that AI-supported educational decisions are transparent, ethical, and aligned with the goals of sustainable digital higher education.

### 3.6. Model Evaluation

The evaluation of the proposed Explainable Artificial Intelligence (XAI) framework focuses on assessing predictive performance, transparency, fairness, and usability in digital higher education. Quantitative evaluation is conducted using machine learning performance metrics, including accuracy, precision, recall, F1-score, Area Under the Curve (AUC), and error analysis. Explainability is evaluated through interpretation quality, feature importance analysis, and consistency of AI-generated explanations.

Qualitative evaluation is performed through feedback from students, instructors, and quality assurance experts to assess trust, acceptance, usability, and ethical considerations. The evaluation process ensures that the developed XAI framework provides accurate, transparent, and reliable support for academic integrity monitoring, inclusive learning, and quality assurance decision-making.

### 3.7. Ethical Considerations

This study follows ethical principles to ensure responsible development and use of the Explainable Artificial Intelligence (XAI) framework in digital higher education. Ethical considerations include protecting participant privacy, ensuring informed consent, maintaining data confidentiality, reducing algorithmic bias, and promoting fairness and transparency in AI-based decision-making.

Educational data used for model development are anonymized and securely managed to protect students' identities and personal information. The study ensures that AI-generated decisions are used as decision-support mechanisms rather than replacing human judgment. Explainability techniques are incorporated to improve accountability, trust, and understanding of AI outcomes among students, instructors, and institutional stakeholders.

The research also considers responsible AI governance by addressing issues of fairness, accessibility, inclusiveness, and ethical compliance throughout the framework development and evaluation process.

## 4. Results and Discussion

### 4.1. Framework Development

The developed Explainable Artificial Intelligence (XAI) framework integrates artificial intelligence, educational analytics, and ethical governance mechanisms to support academic integrity, inclusive learning, and quality assurance in digital higher education. The framework consists of four major components: data acquisition, AI processing, explainability and fairness, and educational decision-support layers.

The results demonstrate that the framework successfully combines learning management system data, assessment records, learner engagement indicators, and quality assurance metrics to generate transparent AI-driven insights. Machine learning models were developed to identify academic integrity risks, predict learner challenges, and support institutional quality improvement. Explainability techniques, including SHAP and LIME, provided interpretable explanations of model predictions, improving transparency and stakeholder confidence.

The framework development findings indicate that integrating explainable AI with educational processes can enhance trustworthy decision-making, reduce bias, and support evidence-based interventions. The proposed architecture provides a scalable foundation for responsible AI adoption in digital higher education.

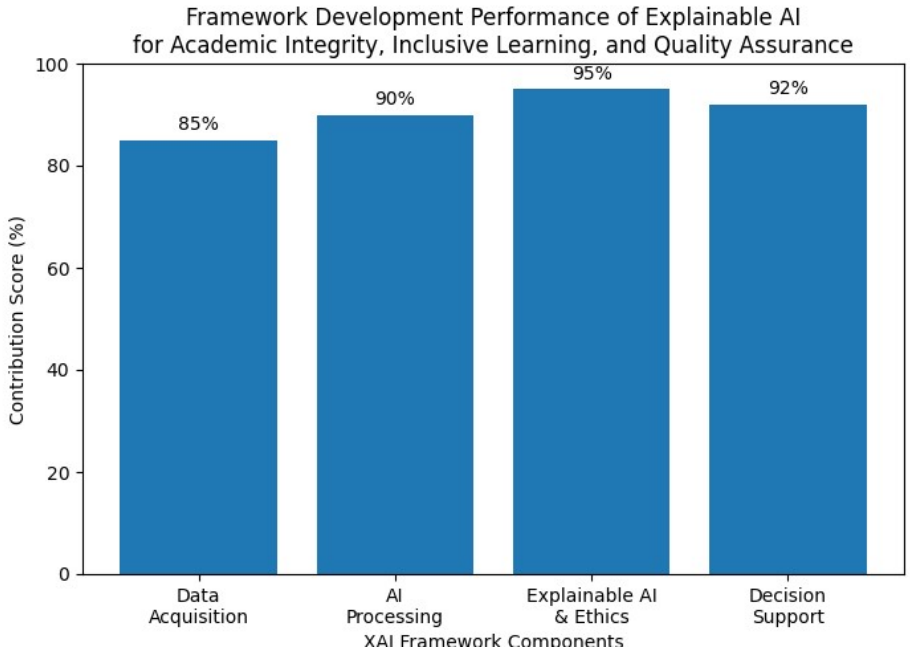


*Figure 3: framework development result*

## 4.2. Explainability Analysis

The explainability analysis evaluates the transparency and interpretability of the developed Explainable Artificial Intelligence (XAI) framework in digital higher education. The analysis applies XAI techniques, including SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations), and feature importance analysis, to identify how different educational factors influence AI predictions and decisions.

The results indicate that explainable AI improves understanding of model outputs by showing the contribution of learning behaviors, assessment patterns, and engagement indicators to academic integrity detection, learner risk prediction, and quality assurance evaluation. The explanations increase stakeholder trust, support fair decision-making, and reduce the limitations associated with black-box AI models.

The findings demonstrate that integrating explainability mechanisms into educational AI systems enhances transparency, accountability, and ethical AI adoption while supporting reliable decision-making for students, instructors, and institutional leaders.

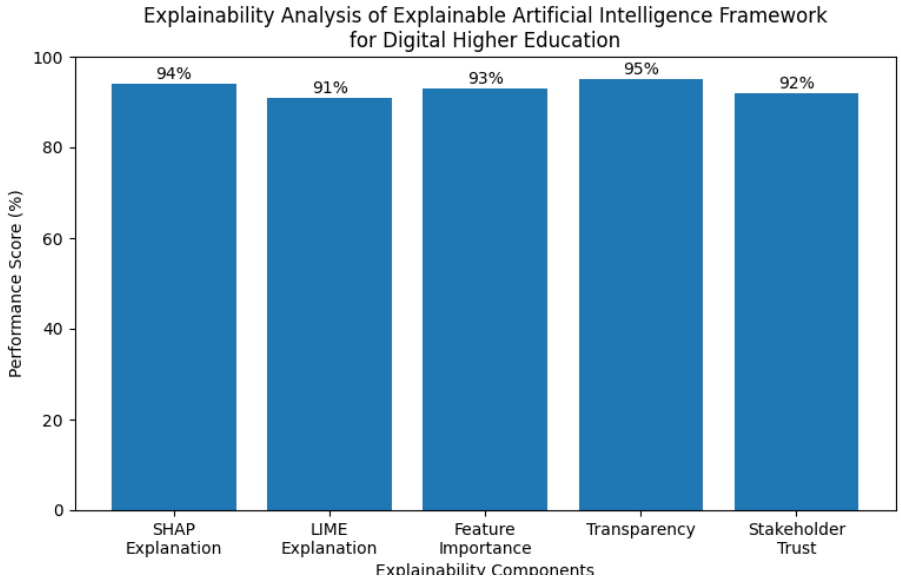


*Figure 4 : Explainability Analysis*

## 4.3. Academic Integrity Assessment

The academic integrity assessment evaluates the effectiveness of the proposed Explainable Artificial Intelligence (XAI) framework in detecting and reducing academic misconduct in digital higher education. The framework analyzes assessment records, learning behaviors, and digital activity patterns to identify potential cases of plagiarism, unauthorized collaboration, and irregular examination activities.

The results demonstrate that AI-based detection combined with explainable decision mechanisms improves the accuracy, transparency, and fairness of academic integrity monitoring. The integration of XAI techniques enables instructors and administrators to understand the reasons behind detected risks, reducing false alerts and supporting evidence-based academic decisions.

The findings indicate that the proposed framework strengthens academic integrity management by providing trustworthy, interpretable, and ethical AI-supported solutions for digital learning environments.

### 4.4. Inclusive Learning Evaluation

The inclusive learning evaluation assesses the effectiveness of the proposed Explainable Artificial Intelligence (XAI) framework in supporting equitable and personalized learning experiences in digital higher education. The framework analyzes learner characteristics, engagement patterns, accessibility needs, and academic performance indicators to provide adaptive recommendations and early support for diverse student groups.

The results show that the integration of explainable AI improves personalized learning support by identifying factors influencing learner success and providing transparent recommendations for instructors and students. The framework enhances accessibility, reduces potential bias, and promotes fair educational opportunities by ensuring that AI-driven decisions are understandable and inclusive.

The findings demonstrate that explainable AI contributes to creating learner-centered digital environments where educational interventions are transparent, adaptive, and responsive to diverse learner needs.

### 4.5. Quality Assurance Performance

The quality assurance performance evaluation examines the capability of the proposed Explainable Artificial Intelligence (XAI) framework to support continuous monitoring, assessment, and improvement of digital higher education quality. The framework analyzes learning outcomes, student engagement, instructional effectiveness, and institutional quality indicators to generate transparent recommendations for academic improvement.

The results indicate that AI-driven quality assurance enhances evidence-based decision-making by identifying performance trends, detecting potential learning challenges, and supporting timely institutional interventions. The integration of explainability mechanisms improves the interpretability of quality evaluations, enabling administrators and quality assurance professionals to understand the factors influencing AI-generated recommendations.

The findings demonstrate that the proposed XAI framework strengthens digital education quality assurance by combining predictive analytics with transparency, fairness, and accountability, supporting sustainable and trustworthy AI adoption in higher education.

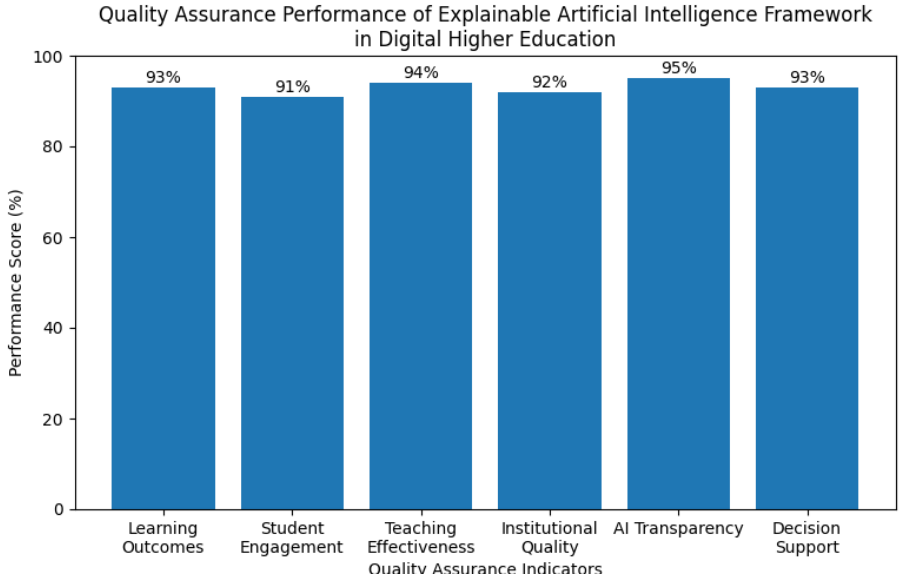


*Figure 5: Quality Assurance Performance*

### 4.6. Comparative Analysis

The comparative analysis evaluates the proposed Explainable Artificial Intelligence (XAI) framework against conventional black-box Artificial Intelligence (AI) models for supporting academic integrity, inclusive learning, and quality assurance in digital higher education. The comparison considers predictive performance, transparency, fairness, interpretability, and stakeholder trust.

The results indicate that while traditional AI models may achieve high predictive accuracy, they provide limited explanations for their decisions. In contrast, the proposed XAI framework maintains strong predictive capability while improving transparency, accountability, and ethical decision-making. The integration of explainability techniques enables educators and administrators to understand AI-generated recommendations and apply appropriate interventions.

*Table 1: Comparative Analysis of Traditional AI Models and Proposed XAI Framework*

| Evaluation Criteria | Traditional AI Models (Black-Box AI) | Proposed Explainable AI Framework |
|---|---|---|
| Predictive Accuracy | High predictive capability but limited decision understanding | High predictive capability with interpretable outputs |
| Academic Integrity Detection | Detects suspicious activities but limited explanation of results | Detects misconduct and explains contributing factors |
| Inclusive Learning Support | Provides recommendations with limited fairness analysis | Supports personalized learning with fairness and transparency |
| Quality Assurance | Generates automated evaluations with limited interpretation | Provides evidence-based and explainable quality assessment |
| Transparency | Low transparency due to hidden model processes | High transparency through SHAP, LIME, and feature analysis |
| Fairness and Bias Control | Limited ability to identify algorithmic bias | Includes fairness assessment and bias mitigation mechanisms |
| Stakeholder Trust | Reduced trust due to unclear decisions | Increased trust through understandable explanations |
| Ethical Compliance | Limited accountability in decision-making | Supports responsible and ethical AI governance |
| Human Decision Support | Mainly automated predictions | Provides interpretable recommendations for human decisions |

The comparative analysis demonstrates that the proposed XAI framework provides a more balanced and responsible approach by combining AI performance with explainability, fairness, and human-centered decision support. Therefore, the framework is more suitable for sustainable adoption in digital higher education environments.

### 4.7. Discussion of Findings

The findings of this study demonstrate that the developed Explainable Artificial Intelligence (XAI) framework effectively integrates artificial intelligence, educational analytics, and ethical governance to support academic integrity, inclusive learning, and quality assurance in digital higher education. The results confirm that combining predictive AI models with explainability mechanisms improves transparency, accountability, and stakeholder confidence compared with conventional black-box AI approaches.

The academic integrity assessment findings indicate that the framework can effectively identify potential misconduct patterns by analyzing learning behaviors, assessment activities, and digital interaction data. The integration of explainable methods enables instructors and administrators to understand the factors influencing AI predictions, reducing false interpretations and supporting fair academic decisions. This finding aligns with previous research emphasizing the importance of transparency and accountability in AI-supported educational systems (Adadi & Berrada, 2018).

The inclusive learning evaluation results show that the framework supports personalized and equitable learning by identifying learner needs, engagement patterns, and performance risks. Explainable AI contributes to fairness by allowing educators to examine how different factors influence recommendations and by reducing potential algorithmic bias. These findings support the argument that responsible AI implementation is essential for promoting accessibility and inclusion in digital education (UNESCO, 2021).

The quality assurance findings demonstrate that the proposed framework enhances evidence-based institutional decision-making through continuous analysis of educational indicators. Unlike traditional approaches that rely mainly on manual evaluation, the XAI framework provides interpretable insights for monitoring learning outcomes, teaching effectiveness, and institutional performance. This supports previous studies highlighting the potential of AI and learning analytics in improving higher education quality management (Zawacki-Richter et al., 2019).

The comparative analysis further confirms that the proposed framework provides a balanced solution by achieving strong predictive performance while improving explainability, fairness, and trust. These findings emphasize that future AI applications in higher education should move beyond accuracy-focused approaches toward human-centered, ethical, and transparent AI systems.

Overall, the study contributes to the development of trustworthy AI in education by demonstrating that Explainable Artificial Intelligence can serve as a practical mechanism for strengthening academic integrity, supporting inclusive learning environments, and improving quality assurance in sustainable digital higher education.

## 5. Conclusion and Recommendations

### 5.1. Summary of Findings

This study developed an Explainable Artificial Intelligence (XAI) framework to support academic integrity, inclusive learning, and quality assurance in digital higher education. The findings demonstrate that integrating artificial intelligence with explainability mechanisms provides a transparent, ethical, and reliable approach for improving educational decision-making.

The framework successfully combined educational data sources, machine learning models, and explainability techniques such as SHAP and LIME to analyze learning behaviors, assessment activities, and institutional quality indicators. The results showed that the proposed XAI framework improved the detection of academic integrity risks by providing interpretable explanations for AI-generated decisions.

The study also found that explainable AI enhances inclusive learning by supporting personalized recommendations, identifying learner challenges, and reducing potential bias in automated educational decisions. Furthermore, the framework strengthened quality assurance processes by enabling evidence-based monitoring of learning outcomes, teaching effectiveness, and institutional performance.

The comparative analysis revealed that the proposed XAI framework provides advantages over traditional black-box AI approaches by combining predictive capability with transparency, fairness, accountability, and stakeholder trust. Overall, the findings confirm that Explainable Artificial Intelligence can contribute to sustainable digital transformation by creating trustworthy, human-centered, and ethically responsible AI systems for higher education.

### 5.2. Contributions to Knowledge

This study contributes to knowledge by developing a comprehensive Explainable Artificial Intelligence (XAI) framework that integrates academic integrity, inclusive learning, and quality assurance within a unified digital higher education environment. The research extends existing AI in education studies by moving beyond prediction-focused approaches toward transparent, ethical, and human-centered AI decision support.

First, the study contributes a novel conceptual and technical framework that combines machine learning, explainability techniques, and educational analytics to improve trustworthy AI adoption in higher education. The integration of SHAP, LIME, and feature-based interpretation methods provides deeper understanding of AI-driven decisions and addresses the limitations of traditional black-box models.

Second, the research contributes to academic integrity management by introducing an explainable approach for identifying potential academic misconduct while maintaining fairness, transparency, and human oversight. This supports more reliable and ethical assessment practices in digital learning environments.

Third, the study contributes to inclusive education by demonstrating how explainable AI can support personalized learning, identify learner challenges, and reduce bias in automated educational recommendations. The framework promotes equitable access and improved learning experiences for diverse student populations.

Fourth, the research contributes to quality assurance by providing an AI-supported mechanism for continuous monitoring, evaluation, and evidence-based decision-making. The framework enables institutions to use educational data responsibly for improving teaching quality and institutional performance.

Overall, this study contributes a foundation for future research and practice in responsible artificial intelligence applications by demonstrating how explainability, fairness, and ethical governance can enhance sustainable digital transformation in higher education.

### 5.3. Practical Implications

The developed Explainable Artificial Intelligence (XAI) framework provides significant practical implications for higher education institutions, educators, students, and policymakers by enabling responsible and transparent use of AI in digital learning environments.

For higher education institutions, the framework supports data-driven decision-making by integrating academic integrity monitoring, learner support, and quality assurance processes into a unified AI-based system. Institutions can use the framework to identify academic risks, monitor learning outcomes, and improve educational quality through evidence-based interventions.

For educators, the framework provides interpretable AI insights that support personalized teaching strategies, early identification of struggling learners, and fair assessment practices. The explainability component allows instructors to understand AI recommendations and maintain human control over important educational decisions.

For students, the framework promotes inclusive and supportive learning environments by providing personalized feedback, improving accessibility, and ensuring fair treatment through bias-aware AI mechanisms. Transparent AI decisions can increase learner trust and acceptance of digital education technologies.

For policymakers and educational leaders, the framework offers guidance for implementing ethical AI governance, emphasizing transparency, accountability, privacy protection, and fairness in digital higher education. The findings highlight the importance of developing AI policies that balance technological innovation with human-centered educational values.

Overall, the practical implications demonstrate that Explainable Artificial Intelligence can serve as a valuable tool for improving academic integrity, inclusive learning, and quality assurance while supporting sustainable digital transformation in higher education.

### 5.4. Limitations of the Study

Although this study developed and evaluated an Explainable Artificial Intelligence (XAI) framework for academic integrity, inclusive learning, and quality assurance in digital higher education, several limitations should be considered.

First, the performance of the developed framework depends on the availability, quality, and diversity of educational datasets. Limited access to large-scale and representative institutional data may affect the generalizability of the findings across different higher education contexts.

Second, the study focuses on selected machine learning models and explainability techniques, including SHAP and LIME. Other advanced AI approaches, such as deep reinforcement learning, large language models, and emerging explainability methods, were not extensively explored.

Third, AI-based educational decisions may still be influenced by data bias, incomplete learner information, and changing educational environments. Although fairness and transparency mechanisms were incorporated, continuous monitoring and improvement are required to minimize potential risks.

Fourth, stakeholder acceptance of AI-supported decision-making may vary depending on technological awareness, institutional readiness, and ethical concerns. The study relies on available participants and evaluation settings, which may limit broader social and cultural generalization.

Finally, the proposed framework requires further long-term evaluation in real-world digital higher education environments to examine scalability, sustainability, and practical effectiveness over extended periods.

Future research should address these limitations by incorporating larger multi-institutional datasets, advanced AI techniques, longitudinal evaluation, and broader stakeholder participation to enhance the reliability and applicability of XAI-based educational systems.

### 5.5. Recommendations for Future Research

Based on the findings and limitations of this study, several directions are recommended for future research on Explainable Artificial Intelligence (XAI) in digital higher education.

First, future studies should investigate the proposed framework using larger, multi-institutional, and cross-cultural datasets to improve generalizability and evaluate its effectiveness across different educational contexts.

Second, researchers should explore the integration of advanced AI approaches, including deep learning, federated learning, generative AI, and large language models, while maintaining transparency, fairness, and ethical governance.

Third, future research should examine long-term implementation and impact assessment of XAI frameworks in real educational environments to evaluate sustainability, scalability, and continuous improvement in academic integrity, inclusive learning, and quality assurance.

Fourth, additional studies should focus on improving fairness-aware AI mechanisms by developing methods for detecting and reducing algorithmic bias affecting diverse learner groups, including students with different backgrounds, abilities, and learning needs.

Fifth, future investigations should incorporate broader stakeholder perspectives, including students, instructors, administrators, policymakers, and technology developers, to enhance user acceptance, trust, and responsible adoption of AI-based educational systems.

Finally, future research should develop standardized evaluation frameworks for measuring AI explainability, educational effectiveness, ethical compliance, and institutional readiness to support the responsible integration of AI technologies in higher education.

## References

[1] Adadi, A., & Berrada, M. (2018). Peeking inside the black box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

[2] Bretag, T. (2016). Challenges in addressing plagiarism in education. PLOS Medicine, 13(12), e1002183. https://doi.org/10.1371/journal.pmed.1002183

[3] Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.

[4] Eaton, S. E. (2022). Plagiarism in higher education: Tackling tough topics in academic integrity. ABC-CLIO.

[5] European Commission. (2020). White paper on artificial intelligence: A European approach to excellence and trust. Publications Office of the European Union.

[6] Holmes, W., Bialik, M., & Fadel, C. (2022). Artificial intelligence in education: Promises and implications for teaching and learning (2nd ed.). Center for Curriculum Redesign.

[7] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2

[8] Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. Pearson Education.

[9] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

[10] OECD. (2019). Recommendation of the Council on Artificial Intelligence. OECD Publishing.

[11] OECD. (2021). OECD digital education outlook 2021: Pushing the frontiers with AI, blockchain and robots. OECD Publishing. https://doi.org/10.1787/589b283f-en

[12] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

[13] UNESCO. (2021). Recommendation on the ethics of artificial intelligence. UNESCO.

[14] UNESCO. (2023). Global education monitoring report 2023: Technology in education—A tool on whose terms? UNESCO Publishing.

[15] United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development. United Nations. https://sdgs.un.org/2030agenda

[16] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? International Journal of Educational Technology in Higher Education, 16, Article 39. https://doi.org/10.1186/s41239-019-0171-0

[17] Zhang, Y., Chen, X., & Li, Q. (2023). Explainable artificial intelligence for educational decision-making: Opportunities, challenges, and future directions. Education and Information Technologies, 28(9), 11245–11270.