



A trustworthy explainable generative AI framework for credit risk assessment in digital lending platforms: Evidence from emerging economies

Dr. A R Annadurai¹, Dr. R Ilavenil²

¹ Assistant Professor, Department of Commerce, Alagappa University College of Arts and Science for Women, Satellite Campus, Thondi, Ramanathapuram, Tamil Nadu, India

² Assistant Professor, Alagappa University College of Arts and Science for Women, Satellite Campus, Thondi, Ramanathapuram, Tamil Nadu, India

Abstract

The expansion of digital lending in emerging economies has improved access to credit through faster and more efficient loan processing. Artificial Intelligence (AI) plays a central role by enabling lenders to assess borrowers using financial and behavioural data. However, concerns about transparency, fairness, accountability, and ethical decision-making continue to limit trust in AI-driven credit assessment, as many predictive models remain difficult to interpret.

This study proposes a Trustworthy Explainable Generative AI (TEGAI) framework for credit risk assessment in digital lending platforms. The framework integrates Generative AI (GenAI) and Large Language Models (LLMs) with Explainable Artificial Intelligence (XAI) to enhance predictive performance while ensuring transparent and fair lending decisions. It also incorporates responsible AI principles, including bias mitigation, privacy protection, regulatory compliance, accountability, and human oversight to promote trustworthy AI adoption.

A conceptual model is developed to explain the relationships among explainability, trustworthiness, fairness perceptions, decision quality, and lending effectiveness in AI-driven lending environments. The proposed framework contributes to the FinTech literature by providing an ethical, transparent, and explainable approach to credit risk management. It also supports financial inclusion and the development of sustainable digital lending ecosystems in emerging economies by enabling more responsible and trustworthy AI-based lending decisions.

Keywords: Generative Artificial Intelligence, Explainable Artificial Intelligence, credit risk assessment, digital lending, fintech, trustworthy AI, emerging economies

Introduction

The financial services industry has undergone substantial transformation due to the rapid growth of Financial Technology (FinTech). Among the most influential developments is digital lending, which has expanded access to financial services for individuals and businesses that traditionally faced barriers to obtaining formal credit. This development is particularly significant in emerging economies, where a large portion of the population remains underserved by conventional banking institutions.

Artificial Intelligence has played an important role in improving digital lending operations. AI-based systems can process extensive datasets, automate lending decisions, identify risk patterns, and reduce the time required for loan approvals. By utilizing both traditional and alternative data sources, lenders can evaluate applicants who lack formal credit histories, thereby promoting financial inclusion.

Although AI offers considerable benefits, its application in lending has raised important concerns. Many advanced machine learning models provide highly accurate predictions but offer limited insight into how those predictions are generated. Such opacity may create difficulties in explaining lending outcomes, identifying bias, and complying with regulatory requirements. Consequently, stakeholders often question the reliability and fairness of AI-driven decisions.

The emergence of Generative AI and Large Language Models has created opportunities to improve decision-making through advanced reasoning, data interpretation, and

scenario analysis. However, these technologies also introduce challenges related to ethics, privacy, transparency, and governance. As a result, there is an increasing need for AI systems that not only perform effectively but also operate in a trustworthy and explainable manner.

Existing studies have explored AI applications in lending and credit scoring, yet limited attention has been devoted to integrating GenAI and Explainable AI within a single framework, particularly in emerging-market contexts. To address this issue, the present study introduces a Trustworthy Explainable Generative AI (TEGAI) framework that combines predictive intelligence with transparency and responsible governance. The framework aims to enhance decision quality while ensuring fairness, accountability, and regulatory compliance.

Review of Literature

Al-Baity (2023): Examined the growing influence of Artificial Intelligence in the digital finance sector, particularly within the Saudi Arabian context. The study demonstrated that AI technologies contribute significantly to financial services by improving credit assessment, fraud management, customer engagement, and lending operations. The author also identified several barriers to successful implementation, including regulatory challenges, concerns regarding data security, and shortages of specialized expertise. The research emphasized the need for a structured adoption framework that aligns AI innovation with

governance requirements and long-term sustainability objectives.

Dura and Cazacu (2024): Investigated the impact of Artificial Intelligence on credit risk evaluation and financial business transformation. Their findings suggest that AI-driven systems enhance lending processes by increasing analytical efficiency, improving risk prediction capabilities, and enabling the use of non-traditional data sources. At the same time, the study highlighted issues such as algorithmic discrimination, limited model interpretability, and evolving regulatory expectations. The authors argued that financial institutions should adopt balanced strategies that combine technological innovation with ethical and regulatory responsibilities.

Bahlool, Hewahi, and Elmedany (2025): Explored the role of responsible AI practices in digital lending environments. Their review indicated that effective credit assessment requires more than accurate predictive models; it also depends on fairness, transparency, and compliance with regulatory standards. The authors emphasized that explainable and unbiased lending systems can improve stakeholder confidence and support responsible financial decision-making. They further noted that integrating explainability mechanisms and bias-control measures is essential for the sustainable development of digital lending platforms.

Bahadori and Hassanzadeh (2026): Conducted a systematic review of AI applications in digital lending and credit risk management. Their analysis revealed that machine learning technologies contribute to more efficient and accurate credit evaluations, but several concerns remain unresolved. These include insufficient transparency, privacy-related risks, algorithmic bias, and challenges in explaining automated decisions. The study stressed the importance of developing trustworthy AI systems capable of delivering interpretable and accountable lending decisions. The authors recommended that future research focus on frameworks that integrate transparency, fairness, accountability, and explainability, particularly in emerging economies where digital lending plays a vital role in advancing financial inclusion.

Research Gap

Despite substantial progress in AI-based credit assessment, several important limitations remain.

First, existing studies largely emphasize predictive performance while paying comparatively less attention to transparency and interpretability. As a result, many lending systems continue to function as opaque decision-making tools.

Second, although XAI methods such as SHAP and LIME have been adopted to improve interpretability, they are generally implemented separately from Generative AI technologies. Research examining the combined application of GenAI and XAI in credit risk assessment remains limited. Third, ethical dimensions such as fairness, bias reduction, privacy protection, accountability, and human oversight are often treated independently rather than incorporated into a comprehensive framework.

Fourth, relatively few studies specifically address the realities of emerging economies, where alternative data sources, limited credit histories, and financial inclusion challenges significantly influence lending practices.

These gaps indicate the need for a holistic framework that integrates GenAI, XAI, and responsible AI principles to support transparent and trustworthy credit risk assessment in digital lending environments.

Objectives of the Study

1. To Study the role of Artificial Intelligence in strengthening credit risk assessment within digital lending platforms.
2. To evaluate the contribution of Explainable Artificial Intelligence in improving transparency, accountability, and interpretability of credit risk models.
3. To develop a Trustworthy Explainable Generative AI (TEGAI) framework that combines GenAI and XAI for responsible credit decision-making.

Research Questions

1. In what ways does Artificial Intelligence improve credit risk assessment in digital lending platforms operating in emerging economies?
2. How can the integration of Generative AI and Explainable AI support the development of a trustworthy framework for credit risk assessment?

Research Methodology

This research adopts a conceptual and exploratory approach based on an extensive review of academic publications, industry reports, policy documents, and regulatory frameworks related to AI, GenAI, XAI, FinTech, digital lending, and responsible AI practices.

The literature was analysed to identify the major factors influencing trustworthy AI-driven lending decisions. Insights derived from the review were used to construct the Trustworthy Explainable Generative AI (TEGAI) framework. The framework combines the analytical capabilities of Generative AI with the interpretability provided by XAI while embedding principles such as fairness, privacy, accountability, transparency, and human oversight.

A conceptual model was subsequently developed to illustrate the relationships among explainability, trustworthiness, perceived fairness, decision quality, and lending effectiveness. As a theoretical study, the research provides a foundation for future empirical investigations and practical implementation in digital lending environments.

1. Role of AI in Credit Risk Assessment:

Artificial Intelligence enhances credit risk assessment by converting diverse borrower information into actionable lending insights. Modern lending platforms utilize both traditional and alternative data sources, including credit histories, banking transactions, utility payments, mobile usage records, and digital behaviour indicators.

AI algorithms process these datasets through data preparation, feature extraction, and pattern recognition techniques. Machine learning models such as Random Forests, Neural Networks, and Gradient Boosting methods can estimate default probabilities, classify borrower risk levels, and identify potentially fraudulent activities.

The outputs generated by these models support lending decisions, including loan approvals, credit limit allocation, pricing strategies, and portfolio management. As a result, lenders can improve operational efficiency, reduce credit losses, and expand access to credit among underserved populations.



Fig 1: Credit Risk Assessment in Digital Lending

2. Contribution of Explainable AI

Explainable Artificial Intelligence enhances the transparency and accountability of AI-driven lending systems. Traditional predictive models often provide outcomes without clearly explaining the factors influencing those outcomes. XAI addresses this challenge by offering interpretable explanations that help stakeholders understand model behaviour.

Techniques such as SHAP, LIME, feature-importance analysis, and rule extraction reveal the variables that contribute most significantly to lending decisions. These explanations enable lenders to validate model outputs, borrowers to understand approval or rejection decisions, and regulators to assess compliance with fairness requirements. Consequently, XAI strengthens stakeholder confidence and supports responsible implementation of AI within financial services.

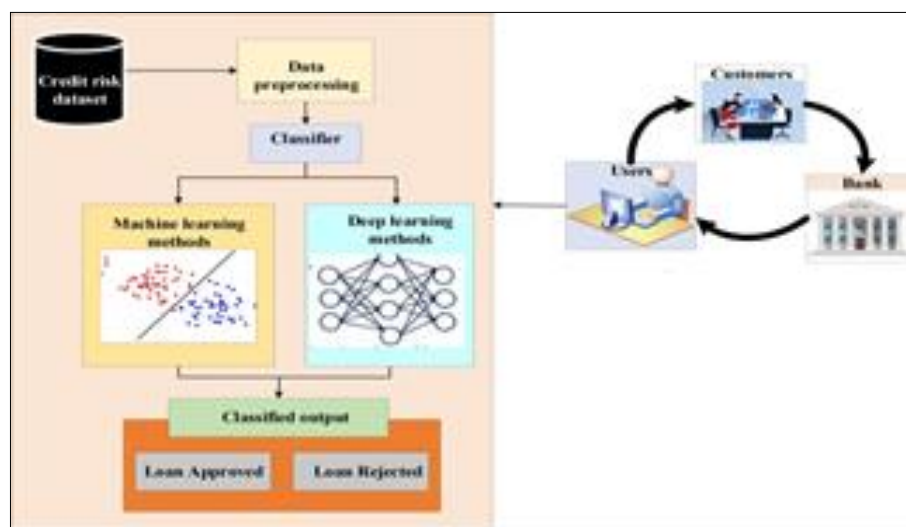


Fig 2: Contribution of XAI in Credit Risk strategy

3. Trustworthy Explainable Generative AI (TEGAI) Framework

The proposed TEGAI framework integrates Generative AI, Explainable AI, and responsible governance mechanisms into a unified credit risk assessment system.

Input Layer

The framework begins by collecting information from multiple sources, including credit bureau records, banking transactions, digital behaviour data, mobile activity, utility payments, and e-commerce interactions.

Data Intelligence Layer

Collected data undergo cleaning, enrichment, and feature engineering processes. Generative AI techniques can also

generate synthetic data to address data scarcity and improve model robustness.

GenAI Credit Risk Engine

The core analytical component utilizes Generative AI models and Large Language Models to identify complex patterns, simulate risk scenarios, conduct stress testing, and support contextual borrower analysis.

Explainability Layer

The XAI component generates transparent explanations using methods such as SHAP, LIME, rule extraction, and natural-language reasoning. These explanations provide understandable justifications for model predictions.

Trust and Governance Layer

This layer incorporates fairness assessment, bias detection, privacy safeguards, compliance monitoring, audit trails, and accountability mechanisms. Human oversight is maintained to ensure ethical and regulatory alignment.

Decision Layer

Based on the generated insights and explanations, the

system supports loan approval, rejection, conditional approval, dynamic pricing, and credit limit allocation decisions.

Outcomes

The framework is expected to improve predictive accuracy, transparency, fairness, stakeholder trust, and regulatory compliance while reducing model-related risks.



Fig 3: Explainable Generative AI (TEGAI)

Findings

The analysis indicates that Artificial Intelligence substantially improves credit risk assessment by enhancing predictive accuracy, reducing processing time, lowering operational costs, and supporting broader financial inclusion. Explainable AI contributes by increasing transparency and accountability, thereby making lending decisions more understandable to stakeholders.

The proposed TEGAI framework demonstrates that combining GenAI and XAI can improve both analytical performance and interpretability. Furthermore, the integration of fairness controls, privacy protections, governance mechanisms, and human oversight strengthens trust in AI-enabled lending systems. These elements collectively support responsible innovation within digital lending ecosystems.

Conclusion

As digital lending continues to expand across emerging economies, there is a growing need for AI systems that balance performance with ethical responsibility. The Trustworthy Explainable Generative AI (TEGAI) framework proposed in this study addresses this need by integrating Generative AI, Explainable AI, and responsible AI principles within a comprehensive credit risk assessment architecture.

The framework supports accurate and transparent lending decisions while promoting fairness, accountability, privacy protection, and regulatory compliance. By fostering trust among borrowers, lenders, and regulators, TEGAI contributes to sustainable digital finance and inclusive economic development. The framework also establishes a theoretical foundation for future empirical studies and practical implementation of trustworthy AI-based credit risk management systems.

References

1. Al-Baity HH. Artificial intelligence in digital finance: Opportunities, challenges, and future directions. *Journal of Risk and Financial Management*, 2023;16(8):357–372.
2. Bahadori M, Hassanzadeh A. Artificial intelligence in digital lending and credit risk assessment: A systematic review of challenges and opportunities. *International Journal of Financial Innovation and Technology*, 2026;11(1):45–68.
3. Bahloul A, Hewahi N, Elmedany W. Ethical and explainable AI for digital lending: A review of responsible credit scoring frameworks. *Journal of Financial Technology and Data Science*, 2025;7(2):101–118.
4. Dura C, Cazacu S. Artificial intelligence and credit risk assessment: Transforming lending decisions and financial business models. *Economics*, 2024;12(4):89–105.
5. OECD. OECD framework for trustworthy artificial intelligence and responsible innovation. OECD Publishing, 2024.
6. United Nations Development Programme (UNDP). Artificial intelligence for financial inclusion in emerging economies. UNDP Publications, 2023.
7. World Bank. Digital financial services and financial inclusion in developing economies. World Bank Publications, 2022.
8. World Economic Forum. The future of AI in financial services: Governance, transparency and trust. World Economic Forum, 2024.