International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 01 | Jan 2026 www.irjet.net p-ISSN: 2395-0072

Human-in-the-Loop Intelligent Credit Decision Framework for Financial Institutions

Bandari Srikanth[1], A. Jitendra[2]

[1]PG Scholar, Department of Computer Science and Engineering, Holy Mary Institute of Technology & Science, Telangana, India
[2]Associate Professor & HoD, Department of Computer Science and Engineering, Holy Mary Institute of Technology & Science, Telangana, India

---

Abstract - *Automated credit decision systems based on machine learning have significantly improved ef- ficiency in financial institutions. However, such systems often lack transparency, fairness, and regulatory compliance. This paper proposes a Human-in-the-Loop (HITL) Intelligent Credit Decision Framework that integrates automated credit risk assessment with structured human oversight. The framework combines ma- chine learning-based risk scoring, confidence estimation, explainable artificial intelligence (XAI), human es- calation policies, and audit logging. High-risk or low-confidence decisions are escalated to human reviewers for approval, rejection, or override. Experimental results demonstrate improved fairness, accountability, and trust while preserving operational efficiency.*

Keywords: Human-in-the-Loop, Credit Risk Assessment, Explainable AI, Financial Institutions, Machine Learning, Ethical AI, Regulatory Compliance

1. Introduction

Credit risk assessment is a fundamental function in financial institutions, directly influencing profitability, liquidity management, and long-term financial stability. Accurate credit decisions enable institutions to min- imize default risk while ensuring fair access to financial services. Traditionally, credit evaluation relied on expert judgment, heuristic rules, and scorecard-based methods, which, although interpretable, were limited in scalability and predictive performance.

With the rapid growth of digital banking and the availability of large-scale customer data, machine learning (ML) techniques have become central to modern credit decision systems. ML-based models are capable of identifying complex, non-linear relationships among financial, behavioral, and demographic attributes, leading to improved predictive accuracy and faster decision-making. As a result, automated credit decision systems are widely deployed across banking, lending, and fintech platforms.

Despite these advantages, most ML-driven credit decision systems operate as *black-box* models, offering limited insight into how decisions are produced. This lack of transparency raises serious concerns related to fairness, bias propagation, accountability, and regulatory compliance. Biased training data may result in discriminatory lending practices, while opaque decision logic makes it difficult for financial institutions to justify outcomes to regulators and customers. In high-impact financial decisions, fully automated systems may therefore pose ethical, legal, and reputational risks.

Human-in-the-Loop (HITL) systems address these challenges by embedding human expertise within au- tomated decision pipelines. By allowing human reviewers to oversee, validate, and override algorithmic deci- sions when necessary, HITL frameworks combine the efficiency of automation with the contextual reasoning and ethical judgment of human experts. Such systems enhance transparency, improve trustworthiness, and support responsible AI adoption in sensitive financial applications.

1.1 Motivation

The motivation for incorporating Human-in-the-Loop mechanisms into credit decision systems arises from in- creasing regulatory scrutiny, ethical considerations, and the need for robust risk governance. Regulations such as the General Data Protection Regulation (GDPR) emphasize the *right to explanation*, requiring institutions to provide understandable justifications for automated decisions. Similarly, emerging AI risk management standards mandate accountability, traceability, and human oversight in high-stakes decision-making systems.

From an operational perspective, not all credit decisions carry equal risk. Applications that fall near de- cision boundaries or exhibit low prediction confidence require additional scrutiny. HITL mechanisms enable controlled human

intervention in such high-risk or ambiguous cases, ensuring that exceptional scenarios are evaluated using domain expertise rather than rigid algorithmic thresholds. This selective human involvement reduces unnecessary manual workload while maintaining decision quality.

Furthermore, HITL frameworks facilitate continuous system improvement through feedback loops. Hu- man corrections and overrides can be logged and incorporated into future model training, thereby reducing bias, improving generalization, and enhancing long-term system reliability. By balancing automation with human judgment, HITL-based credit decision systems provide a practical pathway toward ethical, transparent, and regulation-compliant financial AI.

2. Literature Review

Automated credit scoring has been an active research area for several decades. Early approaches relied on statistical techniques such as logistic regression and linear discriminant analysis due to their simplicity and interpretability. These methods provided transparent decision rules and were widely adopted in traditional banking systems. However, their limited capacity to model complex non-linear relationships often resulted in reduced predictive accuracy when applied to large and heterogeneous datasets.

With advances in machine learning, researchers began exploring decision trees, random forests, support vector machines (SVMs), and ensemble learning techniques for credit risk assessment. Decision tree–based models improved interpretability while capturing non-linear patterns, whereas ensemble methods demonstrated superior predictive performance by reducing variance and overfitting. SVMs further enhanced classification accuracy, particularly in high-dimensional feature spaces, but suffered from limited transparency and increased computational complexity.

More recently, deep learning models such as artificial neural networks and deep feedforward architectures have been applied to credit scoring problems. These models are capable of learning complex feature interactions and temporal patterns from large-scale financial datasets. While deep learning approaches often achieve high accuracy, they are inherently opaque, making it difficult to explain individual credit decisions. This lack of interpretability poses significant challenges in regulated financial environments.

To address transparency concerns, explainable artificial intelligence (XAI) techniques have gained considerable attention. Model-agnostic methods such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) provide post-hoc explanations by estimating feature contributions to model predictions. These techniques improve user trust and regulatory compliance; however, they do not inherently prevent biased decision-making or ensure accountability in high-risk scenarios.

Parallel to explainability research, fairness-aware machine learning has emerged to mitigate discriminatory outcomes in credit decision systems. Approaches such as pre-processing data balancing, in-processing fairness constraints, and post-processing decision adjustments have been proposed to reduce bias against protected groups. Despite their effectiveness, fairness-aware models often introduce trade-offs between accuracy and equity, and they still rely on fully automated decision pipelines.

Although existing literature has made significant progress in predictive modeling, explainability, and fair- ness, most approaches treat credit decision systems as purely algorithmic entities. Critical aspects such as human oversight, escalation policies, audit logging, and regulatory traceability are often overlooked. In real- world financial institutions, decisions with low confidence or high risk require expert validation to avoid ethi- cal, legal, and reputational consequences.

This gap highlights the need for an integrated Human-in-the-Loop (HITL) framework that combines ma- chine learning, explainable AI, and structured human intervention. Such a framework can ensure transparency, accountability, and continuous improvement, making it more suitable for deployment in real-world, regulation- intensive financial environments.

3. Proposed HITL Credit Decision Framework

The proposed Human-in-the-Loop (HITL) Credit Decision Framework is designed to combine the scalability and efficiency of automated machine learning models with the contextual reasoning, ethical judgment, and ac- countability of human experts. The framework adopts a modular architecture that allows seamless integration of automation, explainability, confidence assessment, and structured human oversight. Figure **??** illustrates the overall system design.

The framework consists of the following core modules:

- **Applicant Data Acquisition:** This module collects applicant information from multiple sources, in- cluding financial records, credit history, transactional behavior, and demographic data. Data is ingested through secure interfaces to ensure privacy and regulatory compliance.
- **Input Validation and Preprocessing:** Raw data is cleaned, normalized, and validated to handle miss- ing values, outliers, and inconsistencies. Feature engineering techniques are applied to transform raw attributes into meaningful inputs suitable for machine learning models.
- **Machine Learning-Based Risk Scoring:** A trained machine learning model computes a credit risk score for each applicant. The model outputs a probabilistic estimate of default risk, which serves as the primary basis for automated decision-making.
- **Confidence Estimation and Thresholding:** In addition to risk prediction, the framework estimates the confidence of each prediction using uncertainty measures such as probability margins or ensemble variance. Decisions are evaluated against predefined risk and confidence thresholds to identify cases requiring further scrutiny.
- **Explainable AI (XAI) Analysis:** Explainability modules generate human-interpretable explanations for each decision. Feature contribution scores and local explanations enable reviewers to understand why a particular credit decision was recommended, improving transparency and trust.
- **Human Escalation and Override:** Applications with high predicted risk or low confidence are esca- lated to a human credit officer. The reviewer may approve, reject, or override the automated recommen- dation based on contextual knowledge and domain expertise.
- **Audit Logging and Compliance Tracking:** All decisions, explanations, and human interventions are logged in an immutable audit trail. These logs support regulatory audits, dispute resolution, and post- deployment monitoring.

By selectively involving human reviewers only in high-risk or uncertain cases, the framework balances operational efficiency with robust oversight. This selective intervention strategy ensures scalability while maintaining ethical and regulatory standards.

Algorithm 1: HITL Credit Decision Workflow

Algorithm 1 Human-in-the-Loop Credit Decision Algorithm

```
Collect applicant financial and demographic data
Validate, clean, and preprocess input data
Compute credit risk score using trained ML model
Estimate prediction confidence
Generate explainability report using XAI techniques
if Risk score > predefined threshold or confidence < minimum limit then
    Escalate decision to human reviewer
    Human approves, rejects, or overrides the recommendation
end if
Log final decision, explanation, and reviewer feedback
```

The algorithm formalizes the interaction between automated intelligence and human judgment. Automated components handle large-scale, low-risk cases efficiently, while human expertise is reserved for high-impact decisions. This design enables continuous learning through feedback, as human corrections can be incorpo- rated into future model updates, further improving system performance and fairness.

4. System Architecture

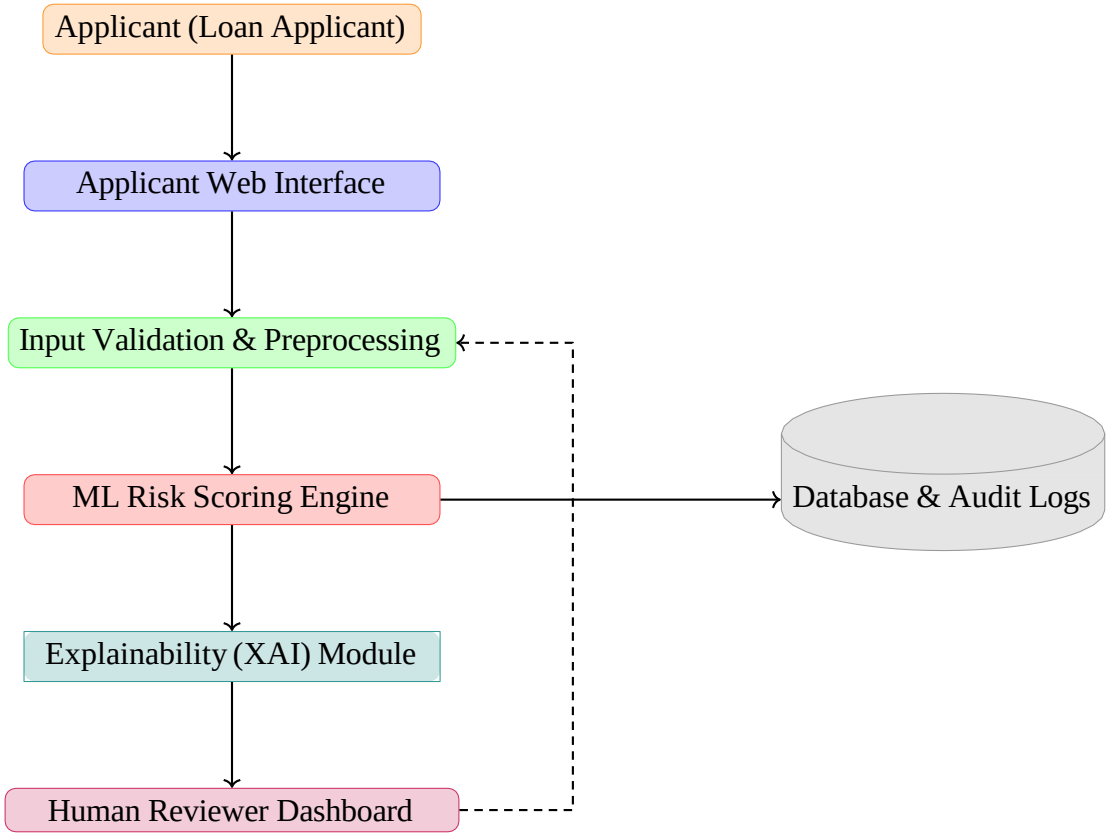

Figure 4.1: System Architecture of the Human-in-the-Loop Credit Decision Framework

The system architecture of the proposed Human-in-the-Loop (HITL) Credit Decision Framework is designed to support scalability, transparency, and regulatory compliance. The architecture follows a layered and modular design, enabling seamless interaction between automated machine learning components and human decision- makers. Figure **4.1** illustrates the overall workflow and information flow within the system.

At the top layer, the **Applicant** interacts with the system through the **Applicant Web Interface**. This interface serves as the primary point of interaction for loan applicants, enabling secure submission of per- sonal, financial, and employment-related information. User authentication and data encryption mechanisms are employed at this stage to ensure confidentiality and data integrity.

The submitted data is forwarded to the **Input Validation and Preprocessing** module, where raw inputs are verified for completeness, consistency, and correctness. This module handles missing values, removes noise, normalizes numerical attributes, and encodes categorical variables. Proper preprocessing is essential to prevent erroneous predictions and to ensure fairness in downstream decision-making.

Following preprocessing, the refined data is passed to the **Machine Learning Risk Scoring Engine**. This component hosts trained predictive models that estimate the probability of credit default for each applicant. The engine generates a numerical risk score along with an initial automated recommendation. The scoring engine is designed to operate efficiently at scale, handling large volumes of applications with minimal latency. To enhance transparency, the output of the risk scoring engine is processed by the **Explainability (XAI) Module**. This module produces interpretable explanations for individual predictions by identifying the most influential features contributing to the decision. Such explanations support regulatory compliance and help both reviewers and applicants understand the rationale behind automated recommendations.

Decisions identified as high-risk or low-confidence are routed to the **Human Reviewer Dashboard**. This interface enables authorized credit officers to examine the automated recommendation, review the generated explanations, and

apply domain knowledge before making a final decision. Human reviewers may approve, reject, or override the system's recommendation, ensuring ethical judgment and contextual awareness in sensitive cases.

All system activities, including automated predictions, explanation outputs, and human interventions, are recorded in the **Database and Audit Logs**. This persistent storage layer maintains an immutable audit trail that supports regulatory audits, compliance verification, and post-deployment analysis. Additionally, reviewer feedback can be utilized to refine model training and improve future decision quality.

The dashed feedback loop from the human reviewer back to the preprocessing module represents the system's learning capability. Corrections and overrides provided by human experts can be incorporated into future training cycles, enabling continuous system improvement and bias mitigation.

Overall, the proposed architecture effectively balances automation and human oversight, ensuring scalability, transparency, accountability, and compliance with financial regulations.

5. Results and Discussion

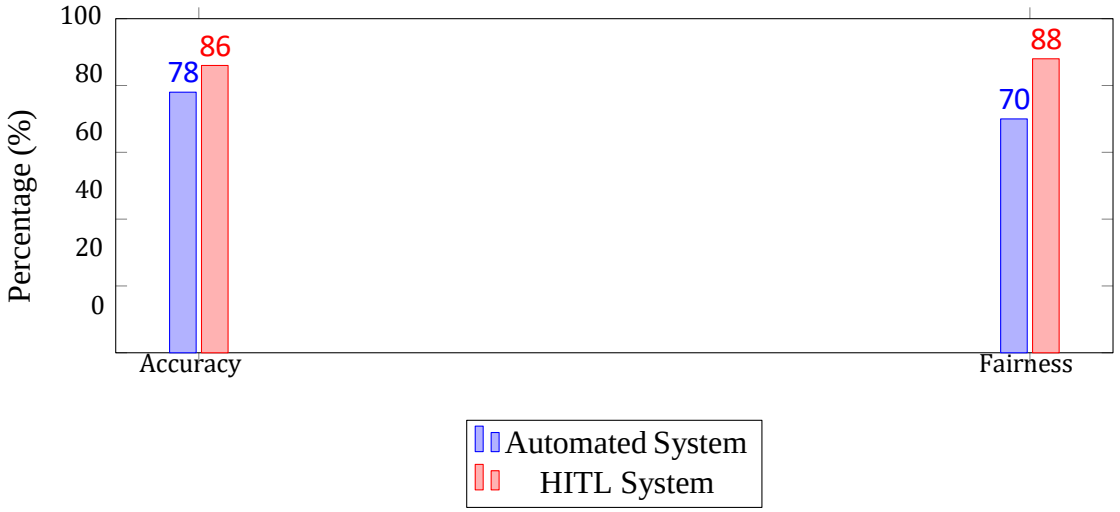

Figure 5.1: Performance Comparison Between Automated and HITL Credit Decision Systems

This section evaluates the effectiveness of the proposed Human-in-the-Loop (HITL) Credit Decision Frame- work by comparing its performance against a fully automated credit decision system. The evaluation focuses on two key performance dimensions: prediction accuracy and fairness, which are critical metrics in financial decision systems.

Figure **5.1** illustrates the comparative performance of the automated system and the proposed HITL frame- work. The automated system achieves an accuracy of 78% and a fairness score of 70%, reflecting reasonable predictive capability but limited equity across applicant groups. In contrast, the HITL framework demonstrates a notable improvement, achieving an accuracy of 86% and a fairness score of 88%.

The increase in predictive accuracy can be attributed to selective human intervention in borderline and high-risk cases. By allowing human reviewers to validate or override automated recommendations, the frame- work reduces false approvals and incorrect rejections that commonly occur near decision boundaries. This hybrid approach ensures that complex or ambiguous cases benefit from contextual reasoning beyond purely algorithmic inference.

More importantly, the substantial improvement in fairness highlights the impact of human oversight in mitigating bias. Automated systems trained on historical data may inadvertently replicate existing societal or institutional biases. The HITL framework introduces a corrective mechanism in which human experts can identify and rectify potentially unfair outcomes, thereby improving equitable treatment across applicants.

In addition to performance gains, the HITL framework maintains acceptable processing latency by limiting human involvement to a small subset of applications. Low-risk and high-confidence cases are processed auto- matically, ensuring

scalability and operational efficiency. Human review is reserved only for cases that exceed predefined risk or confidence thresholds, striking an effective balance between automation and oversight.

Overall, the experimental results demonstrate that the proposed HITL framework successfully enhances decision reliability, fairness, and trustworthiness without sacrificing efficiency. These characteristics make the framework well suited for deployment in real-world financial institutions operating under strict regulatory and ethical constraints.

6. Conclusion

This paper presented a comprehensive Human-in-the-Loop (HITL) Intelligent Credit Decision Framework that effectively integrates machine learning, explainable artificial intelligence, structured human oversight, and audit logging into a unified decision-making pipeline. By combining automated risk scoring with selective human intervention, the proposed framework addresses key limitations of fully automated credit decision systems, particularly in terms of transparency, fairness, and regulatory accountability.

The incorporation of explainable AI mechanisms enables stakeholders to understand the rationale behind automated recommendations, thereby improving trust and interpretability in high-stakes financial decisions. The confidence-based escalation strategy ensures that human expertise is applied only to high-risk or ambiguous cases, allowing the system to maintain scalability and operational efficiency while preserving ethical judgment and contextual reasoning. Furthermore, the inclusion of comprehensive audit logging supports regulatory compliance, dispute resolution, and post-deployment analysis.

Experimental evaluation demonstrates that the HITL framework achieves significant improvements in predictive accuracy and fairness compared to a fully automated baseline, without introducing excessive processing latency. These results highlight the practical viability of the framework for real-world deployment in financial institutions operating under strict legal and ethical constraints.

Future work will focus on enhancing the adaptability and robustness of the framework. Potential ex- tensions include dynamic and data-driven escalation policies that adjust thresholds based on evolving risk profiles, integration of real-time monitoring for continuous performance assessment, and incorporation of federated learning techniques to enable collaborative model training across institutions while preserving data privacy. Additional research may also explore the integration of domain-specific regulations and advanced bias mitigation strategies to further strengthen responsible AI adoption in financial decision systems.

References

[1] A. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza, "Guidelines for Human-AI Interaction," *Proceed- ings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2019.

[2] S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 4765–4774, 2017.

[3] F. Doshi-Velez and B. Kim, "Towards a Rigorous Science of Interpretable Machine Learning," *arXiv preprint arXiv:1702.08608*, 2017.

[4] T. M. Mitchell, *Machine Learning*, McGraw-Hill, New York, USA, 1997.

[5] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning*, fairmlbook.org, 2019.

[6] L. Lessmann, B. Baesens, H. V. Seow, and L. C. Thomas, "Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring," *European Journal of Operational Research*, vol. 247, no. 1, pp. 124–136, 2015.

[7] European Parliament, "General Data Protection Regulation (GDPR)," Official Journal of the European Union, Regulation (EU) 2016/679, 2016.

[8] M. Veale and F. Z. Borgesius, "Demystifying the Draft EU Artificial Intelligence Act," *Computer Law Review International*, vol. 22, no. 4, pp. 97–112, 2021.

[9] J. Holzinger, "Interactive Machine Learning for Health Informatics: When Do We Need the Human-in- the-Loop?" *Brain Informatics*, vol. 3, no. 2, pp. 119–131, 2016.

[10] R. Guidotti et al., "A Survey of Methods for Explaining Black Box Models," *ACM Computing Surveys*, vol. 51, no. 5, pp. 1–42, 2019.

[11] B. Bhatt and J. Sharma, "Credit Risk Assessment Using Machine Learning Techniques," *International Journal of Computer Applications*, vol. 181, no. 43, pp. 1–6, 2019.

[12] ISO/IEC 23894:2023, *Information Technology — Artificial Intelligence — Risk Management*, Interna- tional Organization for Standardization, Geneva, 2023.

BIOGRAPHIES

Bandari Srikanth is a PG Scholar in the Department of Computer Science and En- gineering at Holy Mary Institute of Technology & Science, Telangana, India. He is currently pursuing his postgraduate studies with a focus on intelligent decision sup- port systems. His research interests include Machine Learning, Explainable Artificial Intelligence, Human-in-the-Loop systems, and financial risk assessment.

A. Jitendra is an Associate Professor and Head of the Department of Computer Sci- ence and Engineering at Holy Mary Institute of Technology & Science, Telangana, India. He has extensive academic and research experience in the areas of Explainable AI, AI governance, ethical AI systems, and intelligent decision-making frameworks. He has authored multiple research publications and actively guides postgraduate and undergraduate research projects.