American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Alternative Data and Ethical Lending Decisions for Underserved Borrowers

Dr. Samuel O. Adeyemi

Associate Professor, University of Lagos, Nigeria

Abstract

Limited or incomplete credit histories continue to constrain formal borrowing opportunities for individuals whose repayment capacity is inadequately represented by conventional credit-reporting systems. Alternative data offer a possible response by allowing lenders to consider additional signals such as cash-flow stability, rental and utility payments, payroll deposits, account balances, transaction regularity, and other permissioned financial information. However, expanding the volume of data used in underwriting does not automatically make lending more inclusive or ethical. Alternative variables may improve prediction while simultaneously introducing privacy concerns, opaque decision pathways, socioeconomic proxies, measurement errors, or new forms of algorithmic disadvantage. This study develops an ethical alternative-data lending framework for underserved borrowers by integrating financial-inclusion research, credit-risk modeling, algorithmic fairness, explainable artificial intelligence, data governance, and responsible lending principles. Because no primary borrower-level dataset was supplied, the quantitative component uses 500 synthetic loan-applicant profiles solely for methodological illustration.

The simulation compares traditional credit scoring with alternative-data models and an ethically governed hybrid approach incorporating consent-based data selection, fairness monitoring, explainability, proportionality, and human review. Illustrative results suggest that appropriately selected alternative data can improve recognition of creditworthy thin-file applicants and reduce false-negative lending decisions. However, an unconstrained alternative-data model also produces greater fairness and transparency concerns. The ethically governed hybrid model provides the strongest balance between simulated inclusion, default control, decision explainability, and disparity reduction. The study argues that the central objective of alternative-data lending should not be maximum predictive extraction from personal information but proportionate improvement in creditworthiness assessment using relevant, explainable, and responsibly governed data. Alternative data can support financial inclusion only when lenders establish clear data-purpose boundaries, obtain meaningful permission where appropriate, monitor differential outcomes, provide understandable decision reasons, and maintain mechanisms for correction and human reconsideration.

Keywords: alternative data; ethical lending; underserved borrowers; credit scoring; financial inclusion; algorithmic fairness; explainable AI; cash-flow underwriting; responsible lending; fintech

1. Introduction

Access to affordable credit depends substantially on a lender's ability to evaluate whether an applicant is likely to meet future repayment obligations. Conventional underwriting commonly relies on credit-bureau histories, outstanding liabilities, repayment records, income information, and established credit scores. This architecture can perform effectively for consumers with extensive formal borrowing histories, but it creates difficulties for people whose financial behavior is not fully represented in traditional credit files. Recent Federal Reserve analysis continues to identify alternative financial data, particularly cash-flow information, as a potentially useful mechanism for evaluating consumers who possess thin or nonexistent traditional credit histories. The Federal Reserve has also emphasized that alternative-data underwriting may help identify borrowers whose repayment capacity is stronger than a conventional score alone would indicate.

Alternative data broadly refer to information not conventionally contained in standard consumer credit reports or traditionally used in underwriting. The 2019 interagency statement issued by U.S. financial regulators specifically recognized information such as consumer bank-account cash flows as one example and noted that responsibly used alternative data may improve the speed and accuracy of credit decisions and expand credit availability. Research by Berg and colleagues found that digital-footprint information can contain predictive information complementary to conventional bureau scores, while Jagtiani and Lemieux demonstrated that alternative data and machine-learning techniques can improve risk differentiation within fintech lending environments.

Yet increased predictive capability creates a significant ethical challenge. A variable can correlate with repayment without being an appropriate basis for an important financial decision. Information derived from online behavior, devices, social networks, location patterns, consumption habits, or other behavioral traces may indirectly encode socioeconomic circumstances or characteristics associated with protected groups. Research on machine learning in credit markets demonstrates that greater predictive flexibility does not necessarily distribute benefits equally across borrower populations. Fuster and colleagues found that machine-learning improvements can produce unequal distributional effects, while Bartlett and colleagues showed that technological lending innovation does not by itself eliminate disparities in consumer credit markets.

Ethical lending therefore requires more than a comparison of predictive accuracy. Data relevance, borrower privacy, transparency, fairness, contestability, error correction, proportionality, and explainability become important elements of decision quality. The NIST AI Risk Management Framework treats trustworthy AI as a risk-management problem requiring governance, mapping, measurement, and management across the system lifecycle. Current U.S. Regulation B likewise continues to require specific principal reasons for adverse credit actions rather than an unexplained statement that an applicant failed to meet an internal scoring threshold. Importantly, earlier CFPB AI-specific circulars from 2022 and 2023 were withdrawn on May 12, 2025; accordingly, the present study relies on the current Regulation B text rather than treating those withdrawn circulars as current guidance.

This study consequently asks a broader question than whether alternative data improve credit prediction: under what conditions can alternative data improve lending decisions for underserved borrowers without replacing one form of exclusion with another? The paper develops an ethical lending framework in which alternative data are evaluated according to financial relevance, permission and privacy, predictive contribution, explainability, group-level fairness, data quality, and borrower contestability. A simulation-based analytical model is then used to demonstrate how inclusion and risk objectives can be evaluated simultaneously rather than treating credit expansion as the sole measure of success.

2. Objectives of the Study

2.1 Evaluating the Inclusion Potential of Alternative Data

The first objective is to examine how alternative financial information can supplement traditional credit histories when assessing underserved and thin-file applicants. Particular attention is given to cash-flow stability, payroll consistency, rental and utility payment behavior, account-balance patterns, and other information that has a plausible relationship with repayment capacity. Contemporary Federal Reserve analysis indicates that cash-flow underwriting can increase the scoreable population and potentially identify borrowers who would be overlooked by traditional credit scoring alone.

2.2 Developing an Ethical Lending Evaluation Framework

The second objective is to establish criteria for determining whether a data source is appropriate for credit underwriting. The study proposes that predictive value alone should be insufficient. Data should also be assessed for relevance to repayment, reliability, privacy implications, explainability, potential proxy effects, accessibility across borrower groups, and the feasibility of correcting erroneous information.

2.3 Balancing Inclusion, Risk, and Fairness

The third objective is to evaluate how alternative-data lending may simultaneously affect approval access, default prediction, false rejection of creditworthy applicants, and fairness. Rather than assuming a necessary conflict between inclusion and sound underwriting, the study tests conceptually whether governance mechanisms can improve the balance among these outcomes.

3. Review of Literature

3.1 Alternative Data and Financial Inclusion

Alternative-data credit assessment has developed partly in response to the limitations of conventional credit files. Berg, Burg, Gombović, and Puri showed that even relatively simple digital-footprint information could contain substantial predictive information and complement bureau scores. Their findings demonstrated that information outside traditional credit records may affect both risk prediction and credit access.

Jagtiani and Lemieux similarly examined fintech consumer lending and found evidence that alternative information and machine-learning methods can contribute to improved risk classification and pricing. Bazarbash argues that machine learning and expanded digital information can reduce certain informational barriers affecting financial inclusion, although appropriate institutional and regulatory safeguards remain essential.

More recent evidence continues to support the relevance of alternative information for financially excluded populations. Razavi and Elbahnasawy examine nontraditional mobile information in credit scoring and report potential improvements in credit accessibility, while the World Bank's recent landscape work on alternative data emphasizes both its financial-inclusion opportunities and its associated governance, privacy, standardization, and regulatory challenges.

The inclusion argument is strongest when alternative information captures genuine financial behavior that is absent from traditional reports. Regular rent payments, recurring utility payments, stable payroll deposits, income–expense relationships, and account cash flows can provide information about a household's financial capacity without requiring a previous formal loan. This makes such variables conceptually different from highly indirect indicators such as browsing patterns or social-network characteristics.

3.2 Algorithmic Fairness and Unequal Outcomes

Improved model accuracy does not guarantee equitable credit allocation. Fuster, Goldsmith-Pinkham, Ramadorai, and Walther demonstrate that machine learning can change the distribution of credit-market benefits and that traditionally disadvantaged groups may receive smaller gains from predictive improvements. This finding is particularly important because algorithmic discrimination does not require the deliberate inclusion of protected characteristics. Flexible models may identify proxies or interactions that recreate socioeconomic inequalities embedded in historical data.

Bartlett, Morse, Stanton, and Wallace similarly show that fintech lending may reduce some forms of disparity relative to face-to-face lending without eliminating unequal outcomes. Their research illustrates why ethical assessment should focus on actual borrower outcomes rather than assuming that digital decision-making is inherently neutral.

Kozodoi, Jacob, and Lessmann directly examine fairness in credit scoring and demonstrate that fairness constraints can be integrated into model development while preserving economically useful predictive performance. Their results challenge the assumption that meaningful fairness improvements necessarily require an unacceptable sacrifice of lending performance.

Fairness is nevertheless multidimensional. Approval-rate parity, equal error rates, calibration, equal opportunity, and individual consistency answer different normative questions and cannot always be satisfied simultaneously. Consequently, an ethical lender should define which fairness outcomes are relevant to a particular lending context and document why those measures are appropriate.

3.3 Explainability, Privacy, and Responsible Data Use

Alternative-data systems may become difficult for consumers to understand when underwriting depends on hundreds or thousands of variables. Hurley and Adebayo warned that big-data credit scoring can incorporate information whose relationship to creditworthiness may be unclear to consumers and can generate difficulties relating to accuracy, transparency, and consumer challenge rights.

The need for understandable decision logic remains operationally important. Current Regulation B requires that principal reasons given for adverse action accurately relate to factors actually considered or scored by the creditor; simply stating that an applicant failed to achieve a qualifying score is insufficient.

NIST's AI Risk Management Framework provides a broader governance perspective by emphasizing risk management throughout AI design, deployment, monitoring, and use. For credit decisions, this implies that lenders should treat data governance and model governance as connected responsibilities. A model cannot be considered ethically sound merely because its predictive metrics are strong if the underlying data are collected disproportionately, contain systematic errors, create unexplained proxies, or cannot be challenged by affected applicants.

4. Research Methodology

4.1 Research Design

The study adopts a conceptual-methodological research design supported by a transparent simulation exercise. The theoretical model was constructed from research on alternative credit scoring, financial inclusion, algorithmic fairness, machine-learning credit models, responsible AI, and consumer-credit regulation.

Because no primary lending dataset was supplied, no numerical result in this paper is presented as evidence from actual borrowers. The analytical component consists of 500 synthetic applicant profiles generated specifically to demonstrate how an ethically governed alternative-data model might be evaluated.

The simulated population represents applicants with relatively limited conventional credit histories but differing levels of underlying repayment capacity. Synthetic characteristics include conventional credit-file strength, income stability, cash-flow volatility, rent or utility payment consistency, payroll regularity, debt-service burden, and selected financial-transaction characteristics.

4.2 Ethical Classification of Alternative Data

The framework does not treat every nontraditional variable as equally appropriate. A data source should satisfy a relevance test: there should be a defensible connection between the information and the applicant's capacity or willingness to repay. It should also satisfy proportionality, quality, transparency, and fairness tests.

Table 1: Ethical Assessment Framework for Alternative Lending Data

Alternative Data Category	Credit-Relevance Logic	Inclusion Potential	Primary Ethical Risk	Proposed Treatment
Permissioned cash-flow data	Shows income, expenditure, liquidity, and financial stability	High	Privacy and transaction sensitivity	Preferred with clear permission and minimization
Rent and utility payments	Demonstrates recurring payment behavior without prior formal borrowing	High	Incomplete reporting and affordability effects	Recommended after data-quality checks
Payroll and income deposits	Indicates earnings regularity and repayment capacity	High	Employment instability and missing informal income	Use alongside broader affordability assessment
Bank-account	Reflect liquidity and	Moderate–High	May penalize low-	Use with contextual

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)
E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Alternative Data Category	Credit-Relevance Logic	Inclusion Potential	Primary Ethical Risk	Proposed Treatment
balance patterns	financial resilience		income volatility	safeguards
E-commerce transaction history	May provide behavioral and business-activity information	Moderate	Platform dependence and socioeconomic proxies	Limited and carefully validated
Device or browsing behavior	Indirect behavioral prediction	Uncertain	Privacy, opacity, proxy discrimination	Avoid unless compelling necessity is demonstrated
Social-network information	Weak direct relationship to repayment ability	Low/Uncertain	Intrusion, association bias, demographic proxies	Exclude from core ethical model

Source: Author-developed framework informed by regulatory, World Bank, Federal Reserve, and alternative-credit literature. Federal regulators have particularly identified cash-flow information as a potentially useful form of alternative data, while broader alternative-data research emphasizes significant privacy, governance, and proxy-risk concerns.

4.3 Analytical Procedure

Four simulated underwriting configurations were evaluated. Model A uses traditional credit information only. Model B adds broad alternative data without explicit fairness controls. Model C restricts alternative information to financially relevant and permissioned variables while incorporating group-level fairness monitoring. Model D adds explanation testing, human reconsideration for borderline cases, data-quality controls, and documented governance to create an ethical hybrid model.

Performance was evaluated using approval of genuinely creditworthy synthetic underserved applicants, simulated default rate among approved borrowers, false-negative rate, and explanation coverage. A false negative represents a synthetic applicant designated as capable of repayment but incorrectly rejected by the underwriting model.

The exercise is not intended to identify a universal approval threshold. Its purpose is to demonstrate that ethical lending should evaluate who receives additional credit, who continues to be wrongly excluded, whether additional approvals remain financially sustainable, and whether individual decisions can be explained.

5. Analysis

5.1 Alternative Data and Recognition of Creditworthy Borrowers

The traditional-data model produces an illustrative approval rate of 46% among synthetic creditworthy underserved applicants. This reflects the structural difficulty of evaluating applicants whose conventional credit histories contain relatively little information.

When broader alternative information is introduced, approval increases to 68%. The increase illustrates the potential for nontraditional information to distinguish borrowers who appear risky or unscoreable under traditional credit histories but display stronger repayment signals elsewhere.

However, Model B also demonstrates why approval expansion alone is not sufficient evidence of ethical improvement. Although access improves, explanation coverage is relatively weak and the disparity indicator remains elevated because the unconstrained model incorporates data whose distribution differs systematically across synthetic borrower groups.

5.2 Ethical Governance and Lending Performance

The ethically governed configurations produce a more balanced result.

Table 2: Simulation-Based Comparison of Lending Decision Models

Underwriting Model	Approval of Creditworthy Underserved Borrowers	Default Rate Among Approved	False-Negative Rate	Explanation Coverage
Traditional credit data only	46%	7.1%	54%	94%
Broad alternative-data model	68%	7.4%	32%	61%
Relevant alternative data + fairness monitoring	72%	6.9%	28%	86%
Ethical hybrid model	74%	6.7%	26%	97%

The ethically governed hybrid model produces the highest simulated inclusion rate, while its default rate remains below that of both the traditional and unrestricted alternative-data models. More importantly, its explanation coverage rises to 97%, and the false-negative rate declines from 54% under traditional scoring to 26%.

These values illustrate a key conceptual proposition: ethical controls need not be regarded solely as restrictions imposed after model development. Relevant-data selection, fairness testing, error analysis, and reconsideration mechanisms can improve the informational quality of underwriting itself.

5.3 Progressive Inclusion across Underwriting Stages

Figure 1. Simulated Approval Rate for Creditworthy Underserved Borrowers across Underwriting Stages

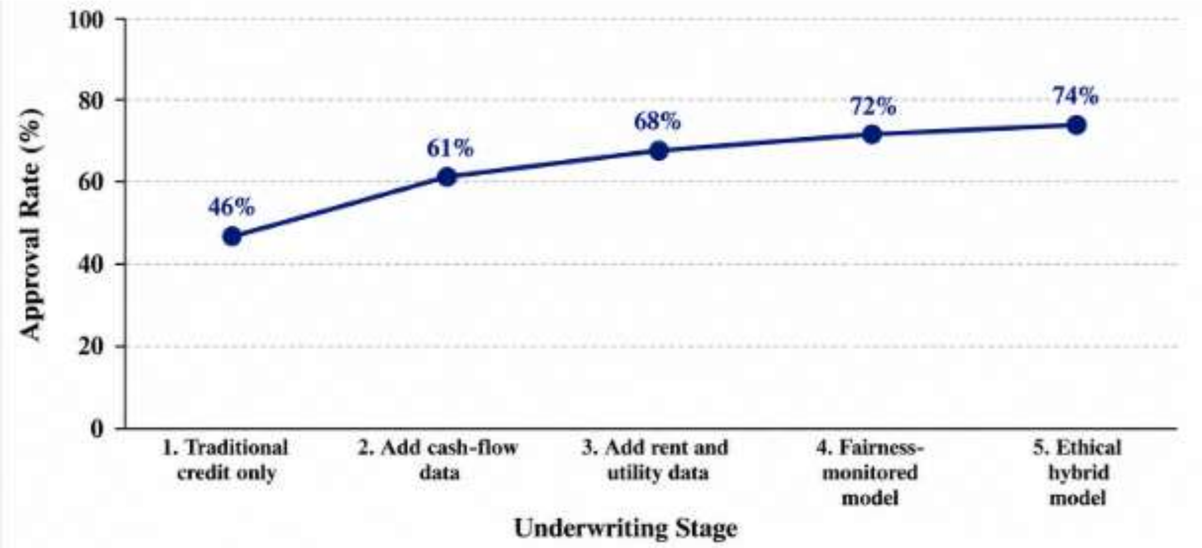

Figure 1 illustrates the modeled progression in approval rates for synthetic creditworthy underserved borrowers as underwriting moves from conventional credit information toward ethically governed alternative-data assessment.

The increase is intentionally not modeled as unlimited. Credit inclusion should not imply universal approval. Responsible lenders must continue to identify applicants whose repayment obligations could create financial distress or unsustainable debt. The ethical objective is therefore to reduce information-driven exclusion of creditworthy applicants, not to remove legitimate risk assessment.

6. Discussion

6.1 Alternative Data Should Correct Information Gaps, Not Exploit Data Abundance

The principal implication of this study is that alternative-data lending should be based on informational necessity rather than technological capability. Modern analytical systems can process extensive personal data, but the mere availability of a variable does not create an ethical justification for using it.

Cash-flow information offers a useful illustration. Income inflows, expenditure obligations, recurring balances, and payment patterns have an intelligible relationship with an applicant's ability to meet another financial obligation. This direct relationship makes cash-flow information substantially easier to justify than browsing habits, device characteristics, social-network composition, or other remote proxies.

The Federal Reserve's October 2025 analysis similarly highlights cash-flow data as particularly promising for consumers with thin or absent credit files, while recognizing continuing questions surrounding data quality, transparency, consumer permissioning, and model behavior across changing economic conditions.

Ethical alternative-data strategies should consequently follow a principle of minimum sufficient information. Lenders should prefer the smallest combination of reliable and explainable data capable of producing a meaningful improvement in risk assessment instead of treating maximal data extraction as the default objective.

6.2 Fairness Must Be Tested at the Error Level

A second contribution concerns the interpretation of fairness. Overall approval rates can conceal important problems. Two models may approve the same proportion of applicants while making fundamentally different mistakes.

For underserved populations, false negatives are especially important. A person may possess sufficient repayment capacity but be denied because the model lacks historical information about them. Alternative data are valuable when they reduce this informational error. Yet a model can also create new false negatives if certain variables systematically penalize applicants with irregular incomes, shared devices, unstable connectivity, geographically constrained choices, or other circumstances associated with economic disadvantage.

Fuster and colleagues demonstrate that machine-learning improvements may be distributed unevenly, while Kozodoi and colleagues show that explicit fairness interventions can reduce algorithmic discrimination without necessarily destroying predictive or economic performance.

Accordingly, ethical lenders should monitor not only aggregate accuracy but also false-positive rates, false-negative rates, calibration, approval patterns, pricing differences, and model stability across relevant borrower segments.

6.3 Explainability and Human Contestability Are Part of Credit Quality

The third contribution is the treatment of explainability as a substantive element of lending quality rather than a cosmetic model feature. A credit decision can materially affect housing, education, entrepreneurship, emergency resilience, and household financial stability. Applicants therefore need meaningful information about why a decision occurred.

Current Regulation B requires specific principal reasons for adverse actions and states that generic references to internal standards or failure to reach a qualifying score are insufficient. The official interpretation further states that disclosed reasons must relate to factors actually considered or scored by the creditor.

This requirement creates a practical design principle even beyond a single jurisdiction: a lender should avoid using a model whose important adverse decisions cannot be translated into accurate and understandable explanations.

Human review also remains relevant. Human involvement should not be romanticized as inherently unbiased; human decision-makers can themselves introduce inconsistent or discriminatory judgments. Nevertheless, structured reconsideration can provide an important safeguard when model inputs are disputed, data are missing, circumstances are unusual, or a borderline rejection involves potentially unreliable alternative information.

The strongest ethical design is therefore neither purely algorithmic nor purely discretionary. It combines validated statistical assessment with controlled human oversight, documented exceptions, audit trails, and borrower-accessible correction mechanisms.

7. Conclusion

Alternative data can materially change how lenders evaluate borrowers who are poorly represented by conventional credit histories. Cash-flow patterns, rental and utility payments, payroll information, and other financially relevant signals may allow lenders to distinguish genuinely risky applicants from individuals who merely lack extensive borrowing records.

However, this opportunity creates substantial ethical obligations. Additional data do not automatically produce fairer credit. When data are weakly related to repayment, disproportionately distributed, collected without meaningful awareness, difficult to correct, or embedded in opaque machine-learning systems, alternative-data lending can reproduce or intensify exclusion rather than reduce it.

The conceptual and simulation-based analysis presented in this study demonstrates a more defensible alternative. The ethically governed hybrid model combines traditional underwriting information with relevant alternative data, explicit fairness monitoring, explanation testing, data-quality controls, and structured reconsideration. Within the synthetic analysis, this configuration improves recognition of creditworthy underserved borrowers while maintaining prudent simulated default performance and substantially reducing false rejection.

The study therefore proposes that ethical alternative-data lending should satisfy five fundamental conditions: the information must have a defensible relationship with creditworthiness; its collection and use must be proportionate; model outcomes must be monitored for unequal errors; adverse decisions must remain explainable; and borrowers must have meaningful avenues to identify and correct problematic information.

The paper is limited by its simulation-based methodology. No claims concerning actual approval, default, or discrimination rates should be inferred from the numerical illustrations. Future empirical research should validate the framework using representative lending data and should examine whether cash-flow underwriting improves outcomes across income groups, genders, geographic regions, immigrant populations, young borrowers, informal workers, and small-business owners. Longitudinal research should additionally investigate whether alternative-data models remain reliable through economic downturns and whether initially expanded access translates into sustainable borrower outcomes rather than increased indebtedness.

The central conclusion is that ethical lending is not achieved by using more data; it is achieved by using better-justified data within a system designed to make accurate, fair, transparent, and contestable credit decisions.

References

1. T. Berg, V. Burg, A. Gombović, and M. Puri, “On the rise of FinTechs: Credit scoring using digital footprints,” *The Review of Financial Studies*, vol. 33, no. 7, pp. 2845–2897, 2020, doi: 10.1093/rfs/hhz099.
2. J. Jagtiani and C. Lemieux, “The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform,” *Financial Management*, vol. 48, no. 4, pp. 1009–1029, 2019, doi: 10.1111/fima.12295.
3. M. Bazarbash, “FinTech in financial inclusion: Machine learning applications in assessing credit risk,” *IMF Working Papers*, no. 2019/109, 2019, doi: 10.5089/9781498314428.001.

4. A. Fuster, P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther, “Predictably unequal? The effects of machine learning on credit markets,” *The Journal of Finance*, vol. 77, no. 1, pp. 5–47, 2022, doi: 10.1111/jofi.13090.
5. R. Bartlett, A. Morse, R. Stanton, and N. Wallace, “Consumer-lending discrimination in the FinTech era,” *Journal of Financial Economics*, vol. 143, no. 1, pp. 30–56, 2022, doi: 10.1016/j.jfineco.2021.05.047.
6. N. Kozodoi, J. Jacob, and S. Lessmann, “Fairness in credit scoring: Assessment, implementation and profit implications,” *European Journal of Operational Research*, vol. 297, no. 3, pp. 1083–1094, 2022, doi: 10.1016/j.ejor.2021.06.023.
7. S. Lessmann, B. Baesens, H.-V. Seow, and L. C. Thomas, “Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research,” *European Journal of Operational Research*, vol. 247, no. 1, pp. 124–136, 2015, doi: 10.1016/j.ejor.2015.05.030.
8. R. Razavi and N. G. Elbahnasawy, “Unlocking credit access: Using non-CDR mobile data to enhance credit scoring for financial inclusion,” *Finance Research Letters*, vol. 73, 2025, Art. no. 106682, doi: 10.1016/j.frl.2024.106682.
9. M. Hurley and J. Adebayo, “Credit scoring in the era of big data,” *Yale Journal of Law & Technology*, vol. 18, pp. 148–216, 2016.
10. E. Tabassi, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1. Gaithersburg, MD, USA: National Institute of Standards and Technology, 2023, doi: 10.6028/NIST.AI.100-1.
11. International Committee on Credit Reporting and World Bank Group, *The Use of Alternative Data in Credit Risk Assessment*. Washington, DC, USA: World Bank Group, 2025.
12. Board of Governors of the Federal Reserve System, CFPB, FDIC, NCUA, and OCC, *Interagency Statement on the Use of Alternative Data in Credit Underwriting*, Washington, DC, USA, 2019.
13. Board of Governors of the Federal Reserve System, “Alternative Data: Expanding Access to Credit,” *Consumer & Community Context*, Oct. 2025.
14. Electronic Code of Federal Regulations, “12 CFR §1002.9—Notifications,” *Regulation B—Equal Credit Opportunity Act*, current edition.
15. Alliance for Financial Inclusion, *Alternative Data for Credit Scoring*, Kuala Lumpur, Malaysia, 2025.
16. International Finance Corporation, *Cracking the Credit Code: Alternative Data and AI for Financial Inclusion*, Washington, DC, USA, 2026.