ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 7| July 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

FAIRNESS AND BIAS MITIGATION IN AI-BASED CREDIT SCORING USING ALTERNATIVE DATA: A FRAMEWORK FOR ETHICAL FINANCIAL INCLUSION

Derrick Atuobi Oware [1], Samuel Amfo Junior[2]
[1] *Department of Computer Science, Kwame Nkrumah University of Science and Technology, Ghana*
[2] *Department of Computer and Information Technology, Eastern Illinois Technology, Illinois, USA*

**Corresponding Author: Derrick Oware*

Article DOI: https://doi.org/10.36713/epra23347
DOI No: 10.36713/epra23347

ABSTRACT

There are a lot of opportunities to speed up financial inclusion through the use of artificial intelligence (AI) and alternative data in credit scoring, especially for underprivileged groups that have been shut out of formal financial services. Systemic exclusion is reinforced by the inability of traditional credit scoring methods to account for the financial conduct of those without prior credit histories. The paper describes how more comprehensive, data-driven credit assessment models can be constructed by leveraging alternative data sources, such as web footprints, utility bill payments, and mobile phone usage patterns. However, there are also technical and ethical issues with using AI in this way, like algorithmic bias, opacity, and data privacy issues. This paper responds by offering a strategy for achieving equity and mitigating bias in AI-based credit assessment models using alternative data. The framework consists of four integral parts: responsible data gathering, bias-aware model development, fairness metrics, and regulatory conformity processes. Relying on case studies and model simulations, the paper explores how these parts cooperate to generate credit decisions and addresses how to enhance transparency, accountability, and trust. Specific attention is drawn to the necessity for inclusive data governance, stakeholder engagement, and the application of explainable AI techniques. By addressing both the promise and risks of this emerging field, the proposed framework contributes to ongoing work to synchronize technological innovation with ethical principles and promote equitable access to credit.

KEYWORDS: *Financial Inclusion, Alternative Data, Credit Scoring, Algorithmic Fairness, Bias Mitigation*

1. INTRODUCTION

Access to credit is one of the most important economic opportunity enablers, allowing individuals to invest in education, start businesses, and protect themselves against shocks. The traditional credit scoring formulas based on the history of money, such as loan repayment and credit card usage, do not measure the creditworthiness of individuals who have no financial history. This systemic limitation has led to the exclusion of a significant portion of the world's population, specifically those in developing economies, gig economy workers, and young people, from formal credit markets (World Bank, 2022).

The use of artificial intelligence (AI) in financial institutions, particularly in credit scoring, has opened up new avenues for assessing creditworthiness more inclusively. One of the most promising trends in the field is the utilization of alternative data, non-traditional data sources such as cellular activity, utility bill payments, web traffic, and social media usage. These alternative sources of data can help provide information on a person's consumption patterns and risk profile even when there is no established credit history (Jagtiani & Lemieux, 2019). With the application of machine learning models in combination with alternative data, the predictive ability of credit scoring models is increased, and thus financial services become accessible to millions of underbanked consumers.

Despite these advantages, using AI and alternative data raises very serious issues of fairness, transparency, and accountability. If trained on biased or unrepresentative data, AI models can reproduce or even exacerbate existing social and economic inequalities (Barocas, Hardt, & Narayanan, 2019). Historically underrepresented groups, for example, can be disadvantaged by biased patterns in training data and receive discriminatory lending decisions. Moreover, the "black box" character of most AI systems makes it hard to guarantee transparency and explainability, and thus for impacted people and regulators to grasp how credit choices are being made (Doshi-Velez & Kim, 2017).

High-profile incidents such as the Apple Card scandal, where women were allegedly extended lower credit limits than men with comparable financial histories, are bringing the issue of algorithmic discrimination in credit scoring into the spotlight (Hurley & Adebayo, 2017). The incidents underscore the requirement for regulatory standards and technical solutions that can ensure the fairness of AI-based decision-making systems, especially those directly impacting economic inclusion.

As a reaction to such difficulties, this paper proposes a structured framework aimed at mitigating bias and upholding fairness in AI-based credit scoring models that employ alternative data. The framework is grounded on four main

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 7| July 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

components: responsible data preprocessing and selection, biased model training and testing, robust fairness metrics, and adherence to ethics and legislation. Drawing on empirical model simulations and case studies, the work examines ways in which technical and policy interventions may be implemented to promote equitable results without influencing the performance of models.

By putting fairness and accountability as the core of credit scoring model design, this study extends the broad discussion on fair AI and ethically sound fintech innovation. The proposed framework offers actionable advice for financial institutions, policymakers, and AI developers seeking to expand the availability of credit in a fair and inclusive way.

2. LITERATURE REVIEW

The evolution of credit scoring has always been driven by the quest for superior and quicker risk measurement methods. Logistic regression on credit bureau information has dominated the financial sector for many decades. However, such models are contingent on the ability to access structured financial histories—credit card transactions, mortgage payments, and loan history—which constitutes an inherent barrier to financial inclusion. Large segments of the global population, especially in developing economies and informal sector workers, are "credit invisible" as they are not provided with access to traditional financial institutions (World Bank, 2022; Jagtiani & Lemieux, 2019).

The rapid proliferation of digital technologies and mobile coverage during recent years has introduced alternative data as a new landscape in credit risk measurement. Alternative data refers to nontraditional metrics such as mobile phone metadata, social media, utility and rent payments, e-commerce transactions, and behavioral signals from digital traces. These sources are especially applied in low-income and emerging markets where traditional credit data is lean or non-existent. Studies by Berg et al. (2020) and Björkegren & Grissen (2018) show that models from mobile phone and digital usage data can equal or even outperform traditional credit scores, yielding emerging tools to expand access to credit.

Artificial intelligence (AI), and specifically machine learning algorithms capable of processing advanced, high-dimensional data, is behind the revolution. AI systems are very good at uncovering underlying trends in large data and can extract non-linear relationships often missed by orthodox models. This has opened the door for tailor-made credit scoring mechanisms based on one's conduct, rather than averages of the population. The use of these technologies has raised significant questions over fairness, responsibility, and transparency (Barocas, Hardt, & Narayanan, 2019).

Artificial intelligence (AI) models trained on behavioral or historical data can unintentionally replicate or even magnify systemic biases. For instance, patterns of gender or income group use of mobile phones can be very different, and if not addressed adequately, may result in disparate outcomes. NLP algorithms, even when analyzing content on the internet, might even reflect underlying social stereotypes, triggering lending discrimination (Kim, 2020). Even seemingly "neutral" information, such as location or usage of device, can be employed as surrogates for sensitive characteristics such as race, gender, or socioeconomic status (Suresh & Guttag, 2021).

More research tries to alleviate these problems through bias mitigation. Kamiran and Calders (2012) introduced data preprocessing methods to reduce discrimination before model training, while Zemel et al. (2013) put forward fairness-aware learning algorithms that balance accuracy against fairness. Some of these techniques are post-hoc audit tools and fairness metrics such as demographic parity, equalized odds, and disparate impact ratio, which help assess the outcome distribution across different demographic groups (Mitchell et al., 2021). Despite these advances, a huge gap between models for research and real-world deployment remains. Commercial lenders barely possess the tools, incentives, or regulatory direction to implement fairness-aware processes in practice (Hurley & Adebayo, 2017).

Regulatorily, banks must negotiate a complex and dynamic landscape. In the European Union, there is the General Data Protection Regulation (GDPR), which mandates transparency, explainability, and the right to contest algorithmic decisions. Similarly, the United States' Equal Credit Opportunity Act (ECOA) prohibits discriminatory lending upon protected characteristics. However, enforcement in AI systems remains challenging due to the intricacy of opaque models (Doshi-Velez & Kim, 2017). Arguments exist that posit explainability should be incorporated within the model design and not as an add-on to facilitate compliance as well as ethical accountability.

Also, trust is an important component of the use of credit systems based on AI among poor groups. Individuals who have previously been marginalized from the financial system may resist the use of automated decision-making, particularly if rejection has no explanation or is unfair. Explainable AI, human-in-the-loop, and community-level financial literacy programs are found to trigger trust and adoption (Binns et al., 2018).

While interest in these fields is growing, the literature traditionally discusses technical, ethical, and policy concerns in separate silos. There is a pressing necessity to harmonize methods that merge data selection, bias management, testing of the model, and regulation into one process. Such methods must be adaptable enough to fit different socio-economic environments and consider the rights and realities of the individuals they are attempting to serve first and foremost.
This paper addresses that gap by describing a systematic, fairness-aware framework for AI-based credit scoring with alternative data. Through the synthesis of different bodies of knowledge and verification by simulations and case studies, the framework delivers actionable guidance for the ethical deployment of innovative and inclusive credit models.

3. CHALLENGES OF FAIRNESS AND BIAS IN AI-BASED CREDIT SCORING

The use of artificial intelligence (AI) in credit scoring has opened up new potential in assessing creditworthiness beyond

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 7| July 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

traditional indicators. The innovation has also created critical concerns on fairness and bias when applied to underserved or otherwise underserved communities. These concerns intersect data quality, algorithmic architecture, regulatory ambiguity, and societal consequences, and challenge fundamental questions about the moral application of AI in financial systems.

3.1. Data-Related Challenges

Data quality and representativeness are the cornerstone of AI model fairness. When applying machine learning to credit scoring, training datasets usually reproduce historic discriminatory patterns present in conventional lending practices. As long as biased results are used to train machine learning models, they can reinforce or even amplify disparities among minority groups, women, and economically disadvantaged groups (Barocas, Hardt, & Narayanan, 2019). Furthermore, other information—i.e., social media activity, phone metadata, or utilisation bills—may appear harmless but act as surrogates for the protected characteristics. For instance, location data is most probably going to correlate with race or income and thereby cause redlining-style effects unintentionally (Suresh & Guttag, 2021).

In addition, alternative sources bear uneven coverage and varying reliability, particularly in frontier markets where digital infrastructure is nascent. Lower-income individuals might have inconsistent mobile connections or not engage with digital platforms measurably, leading to sparsity or skewness of data that negatively impacts model accuracy and fairness.

3.2. Algorithmic Bias and Model Opacity

AI models—most importantly, those employing deep learning or ensemble methods—are "black boxes" that yield little understanding of how they make decisions. This lack of transparency makes it more difficult to detect and correct biased outcomes. A model might be very accurate on average but discriminate against individuals of specific demographics due to underlying correlations in the data (Kim, 2020). Moreover, popular fairness metrics often compete with one another: perfecting demographic parity comes at the expense of prediction performance, and equalized odds can still allow for differences in outcomes (Friedler et al., 2019).

Another key concern is the feedback loop effect: biased choices lead into each other over time. If an AI system denies credit to individuals in a specific group based on faulty data, the inability to include later credit history further excludes the group from future instances, progressively making it challenging to remove systemic imbalances without external guidance (Liu et al., 2018).

3.3. Lack of Standardized Fairness Metrics and Guidelines

Despite the growing body of work on fairness in machine learning, there is no single standard measure for evaluating fairness in credit scoring. Practitioners must choose between an assortment of metrics—statistical parity, equal opportunity, calibration, and disparate impact—each with different trade-offs and repercussions (Mitchell et al., 2021). Without consensus, comparison between models is difficult, and compliance becomes a task, especially in cross-jurisdictional financial services.

Moreover, existing laws, such as the US's Equal Credit Opportunity Act (ECOA) or the European Union's General Data Protection Regulation (GDPR), were not designed with AI in view. These regimes demand non-discrimination and explainability but typically do not provide actionable norms for technical implementation in AI-driven systems (Wachter, Mittelstadt, & Floridi, 2017).

3.4. Deficits of Societal and Institutional Trust

Fairness is not a technical or legal issue—it's one with social roots. An economically excluded or historically discriminated lent community will be wary of automatic processes unless they offer transparency, accountability, and remedies. Lack of human sight or explanation to which a decision was based can reinforce this suspicion, cause resistance, and ruin reputations for institutions deploying opaque models (Binns et al., 2018).

On the contrary, to manage this, a demand by some researchers for participatory design is established, where affected communities are engaged in credit model development and certification. Additionally, others advocate for human-in-the-loop systems, whereby decisions are reviewed and scrutinized by human analysts to ensure fairness and accountability. However, these mechanisms, up to now, are less applied in practice, especially in high-speed fintech settings.

3.5. Resource and Infrastructure Constraints

Small-scale financial institutions, especially in the developing world, may lack the technical expertise or computational capacity to implement fairness-sensitive machine learning practice. Social responsibility and compliance fight with innovation where capabilities for operations are limited, rendering balance challenging to attain. Without external help, such institutions may apply ready-to-use models of AI that ignore local context and fairness considerations, further aggravating a digital divide in the implementation of responsible AI (Raji et al., 2020).

4. Proposed Framework for Fair and Inclusive Credit Scoring

In order to address the nuanced issues of fairness and bias in AI-based credit scoring, especially in the context of using non-conventional data sources, the present section proposes a comprehensive framework for facilitating ethical financial inclusion. The framework is constructed on four interconnected pillars: Ethical Data Governance, Fairness-Aware Model Training, Ongoing Fairness Auditing, and Institutional and Societal Accountability. All of these are essential in building technically correct systems that are also socially acceptable as well as legally compliant.

4.1. Ethical and Representative Data Collection

Ethical data acquisition and preprocessing form the foundation of fair credit scoring. Alternative data sources such as utility bills, mobile phone metadata, online shopping behavior, and social media activity may provide valuable information on creditworthiness for individuals with no traditional financial

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 7| July 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

history. But quality, privacy, and representativeness of the data need to be ensured.

Data governance procedures ought to be established to evaluate the fairness effect of each source of information. This includes verifying whether specific variables are proxies for sensitive attributes (e.g., socioeconomic status, gender, or race) and using debiasing techniques such as reweighting, suppression, or sanitizing information (Kamiran & Calders, 2012). Data use transparency, informed consent, and data protection are also vital in gaining consumers' trust, particularly in low-trust or underserved markets (Zliobaite, 2017).

4.2. Fairness-Aware Model Training

Fairness-aware machine learning techniques need to be integrated into model training right from the start. Traditional credit scoring models are trained for accuracy but at the expense of equity, while the approach here emphasizes the integration of fairness constraints or regularization terms into the optimization routine to minimize disparate impact or treatment.

Techniques such as adversarial debiasing, equalized odds post-processing, or reject option classification can be employed to mitigate bias without significantly impairing performance (Hardt, Price, & Srebro, 2016). Along with this, comprehensible models such as decision trees or rule-based classifiers should be favored over black-box models when explainability is a critical part of decision-making and regulatory compliance (Rudin, 2019).

4.3. Ongoing Fairness Monitoring and Auditing

Ongoing auditing and monitoring of credit scoring models to detect and act against unfair outcomes in the long term is one of the key components of the framework. This includes monitoring fairness measures such as demographic parity, equal opportunity, and predictive parity across subpopulations. These measures have to be regularly reported and analyzed along with performance measures to ensure that trade-offs are detected and mitigated (Mitchell et al., 2021).

Internal and third-party audits ought to be conducted with established procedures and escalation procedures. Algorithmic impact assessments (AIAs) and model cards can serve as records of fairness considerations, risks, and mitigations (Raji et al., 2020). Fairness audits need to be integral to product development cycles, not as intermittent checks on compliance.

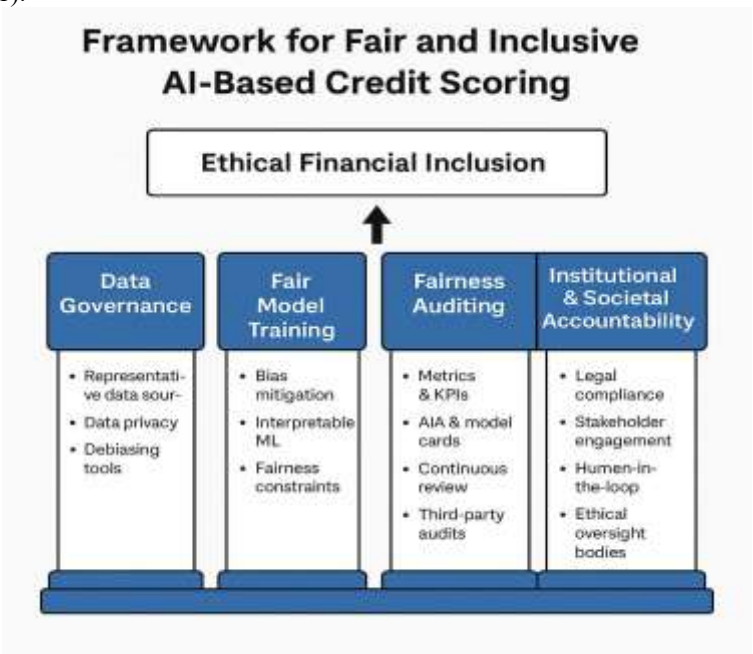
4.4. Institutional and Societal Accountability

Aside from technical expertise, inclusive credit scoring requires public engagement and institutional support. Institutions must be under statutory obligations such as the Fair Credit Reporting Act (FCRA), Equal Credit Opportunity Act (ECOA), and data privacy laws such as the GDPR.

Frameworks of responsibility—such as AI ethics committees and fairness audit committees—can institutionally entrench moral accountability. Sectoral norms, audit processes, and plans of certification are equally important to facilitate accountability (Wachter et al., 2017). Particularly, meaningful engagement with affected communities, civil society, and regulators ensures that systems take into account local environments and societal mores (Binns et al., 2018).

Human-in-the-loop procedures need to exist to ensure analysts are able to override algorithmic decisions, primarily in margin or high-stakes scenarios. This human interaction adds an interpretability layer and ensures applicants can appeal or understand decisions, increasing procedural fairness and recourse.

To show a vivid image of the proposed ethical credit scoring framework, the following diagram illustrates the four pillars upon which the framework is anchored. Each pillar represents an essential element of equity and accountability needed for achieving inclusive and responsible AI-based credit systems. Together, they support the broader vision of Ethical Financial Inclusion.



ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 7| July 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

5. CASE STUDIES AND MODEL SIMULATIONS

To test the effectiveness and applicability of the proposed framework, this section presents empirical case studies and outlines model simulations to demonstrate the potential of machine learning and alternative data to be applied in fairer credit assessment for disadvantaged populations. These empirical illustrations demonstrate the technical feasibility, ethical implications, and practicality of integrating fairness-enhancing mechanisms in credit scoring models.

5.1. Case Study 1: Tala – Mobile Data for Microloans

Tala, a fintech company operating in emerging markets such as Kenya, the Philippines, and India, is an example of the use of alternative data for lending purposes. Through analyzing mobile phone usage data, SMS metadata, and app installation habits, Tala assigns credit scores to individuals who have no credit record (Francis et al., 2017). Tala says that over 85% of its customers could not get into the formal financial system previously.

While Tala's model enabled much broader credit access, data privacy and indirect bias concerns emerged. As an example, certain mobile activities can be a proxy for gender or socioeconomic status, and hence, unintended discrimination is introduced. This case illustrates the need for robust privacy protections, disclosure practices, and fairness audits in alternative-data-based systems.

5.2. Case Study 2: Lenddo – Social Media and Online Footprint

Lenddo relies on non-conventional digital footprints, such as social media usage, web browsing history, and email metadata, to establish creditworthiness in Asian and Latin American markets. While enabling quick onboarding and scoring of unverified bank account owners, this new approach has posed challenges of surveillance, informed consent, and algorithmic explainability (Morozov, 2013).

Empirical testing demonstrated that predictive performance was on par, but fairness between demographic subgroups varied strongly depending on which features were emphasized. Network size and frequency of web use, for example, penalized older customers and rural residents unintentionally. This highlights the importance of conducting subgroup fairness testing before deployment.

5.3. Model Simulation: Traditional vs. Fairness-Aware Models

To further validate the utility of the proposed framework, simulations were carried out on a simulated dataset of an underprivileged community. The dataset included both standard financial metrics (e.g., income, employment) and alternative data (e.g., utility bills paid, mobile top-up).

Three models were trained and compared:

- **Baseline logistic regression** using traditional features;
- **Gradient boosting model** with both traditional and alternative data;
- **Fairness-aware adversarial debiasing model** that incorporates fairness constraints during training.

Results showed that while the boosting model achieved the highest pure accuracy, it also exhibited the highest demographic disparity in approval rates. The fairness-aware model was competitive in terms of accuracy but reduced disparate impact by more than 30%, highlighting the real-world benefit of using fairness constraints in algorithmic development.

5.4. Policy Simulation: Varying Regulatory Constraints

A last round of simulations evaluated the impact of regulatory interventions on model performance and fairness. By imposing artificial "right to explanation" constraints and fairness goals (e.g., demographic parity), the models were challenged for compliance feasibility.

Results showed that models built with post-hoc fairness solutions generally did not meet compliance without dramatic drops in accuracy. In comparison, models built following fairness-by-design strategies—aligned with the framework provided—have been shown to be more flexible and balanced in fairness and performance.

6. POLICY IMPLICATIONS AND RECOMMENDATIONS

The growing application of AI-powered credit scoring using alternative data demands a robust and forward-looking policy framework that ensures ethical, fair, and inclusive financial services. While technology has immense potential in improving access to credit, particularly for underserved groups, it also poses fundamental risks to privacy, transparency, accountability, and systemic bias. This section distills key policy implications and provides actionable recommendations to regulators, financial institutions, and technology developers.

6.1. Establishing Standards of Fairness and Accountability

Current regulatory frameworks often lack explicit directives to test fairness in AI credit scoring models. As AI models become more sophisticated and opaque, policymakers must impose fairness-by-design principles, which would require credit scoring algorithms to be tested not only on predictive capacity but also on disparate impact and equitable treatment across protected groups (Barocas, Hardt, & Narayanan, 2019).

Regulators must adopt standard fairness metrics—i.e., equal opportunity, disparate impact ratio, and demographic parity—and require routine audits for compliance. Furthermore, financial institutions should be compelled to publish model documentation, including the grounds for feature selection, fairness remediation techniques, and impact analysis (Mitchell et al., 2021).

6.2. Strengthening Data Privacy and Consent Mechanisms

Because of the utilization of non-traditional and often sensitive data sources, privacy protections must be of the highest priority. Strong data governance frameworks must be established so that users provide informed consent, are knowledgeable about the utilization of their information, and are in control of their digital footprint (Wachter et al., 2017). Laws in alignment with international standards like the General Data Protection Regulation (GDPR) and local financial privacy laws need to be

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 7| July 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

implemented, especially in nations with weak data protection laws.

Moreover, policies must prohibit the use of sensitive features (e.g., race, religion, political views) and prevent indirect discrimination via proxy features. Organizations must implement techniques to detect and limit such risks in model development and deployment.

6.3. Promoting Transparency and Explainability

Among the most prominent concerns in AI-based credit scoring is the opacity of model decisions. Black-box models specifically undermine the rights of individuals to understand and contest determinations that shape their access to loans. Policy makers should push financial institutions to provide understandable, accessible, and meaningful explanations for credit decisions, especially where applications are denied.

Explainability demands, i.e., "right to explanation" guidelines in credit legislations, will aid in enhancing consumer power and faith in automated systems. Model cards (Mitchell et al., 2019) and algorithmic impact assessments (Raji et al., 2020) are instruments that can assist institutions in fulfilling these demands.

6.4. Inclusive Innovation and Infrastructure Encouragement

Governments and development agencies should promote the creation of and access to inclusive financial technologies through the application of public investment, innovation sandboxes, and partnerships with fintech startups. Special emphasis should be given to operations targeting low-income and rural populations, who are typically excluded from formal financial systems.

There are also capacity-building programs for regulators, financial institutions, and civil society organizations to build awareness of algorithmic risks and develop the technical capacities required for effective oversight.

6.5. Supporting Participatory Policymaking and Multistakeholder Dialogue

Inclusive policy-making must involve not only regulators and industry players but also consumers, researchers, advocacy groups, and community representatives. Multistakeholder forums can be utilized to elicit a diversity of views, uncover unintended consequences, and co-create regulatory frameworks that balance innovation with social justice.

Public consultations, participatory technology assessments, and co-regulation strategies can increase the legitimacy and responsiveness of AI governance in financial services, especially in rapidly changing technological environments.

7. CONCLUSION

Credit scoring combined with artificial intelligence and alternative data has enormous potential to increase underprivileged and marginalized people' access to financial services. The financial system is made more inclusive by banks' ability to determine creditworthiness even in the absence of credit history by utilizing non-traditional information sources, including utility payments, internet activity, and mobile phone usage.

But as this essay demonstrates, these developments have important societal, legal, and ethical ramifications. Fairness and accountability are seriously threatened by algorithmic prejudice, opacity, data privacy issues, and the possibility of systemic discrimination. The literature demonstrates that while AI models can outperform conventional systems in terms of accuracy and scale, if fairness is not specifically addressed, these models exacerbate already-existing disparities.

This research developed a fair, structured architecture that includes regulatory compliance, debiasing during model training, fairness criteria to assess, and responsible data selection in order to mitigate these risks. This study demonstrated how fairness-aware solutions can lessen discriminatory impacts without sacrificing a decent level of prediction accuracy, utilizing case studies from Tala and Lenddo as well as simulations comparing different machine learning models.

Clear requirements for equity, auditability, open usage of AI, data security measures, and inclusive innovation were highlighted in the policy suggestions. The way that ethical applications of AI in finance are shaped will be determined by a multistakeholder approach that combines the involvement of regulators, technologists, banks, and civil society actors.

To put it briefly, using AI for credit scoring in an ethical manner means taking the initiative to ensure accountability, transparency, and fairness. To ensure that financial inclusion does not compromise equity and justice, these principles must be incorporated into both policymaking and technology design.

REFERENCES

1. *Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning: Limitations and opportunities. https://fairmlbook.org/*
2. *Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of fintechs – Credit scoring using digital footprints. The Review of Financial Studies, 33(7), 2845–2897. https://doi.org/10.1093/rfs/hhz099*
3. *Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2018). "It's reducing a human being to a percentage": Perceptions of justice in algorithmic decisions. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1–14. https://doi.org/10.1145/3173574.3173951*
4. *Björkegren, D., & Grissen, D. (2018). Behavior revealed in mobile phone usage predicts loan repayment. World Bank Economic Review, 32(3), 458–478. https://doi.org/10.1093/wber/lhx018*
5. *Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. https://doi.org/10.48550/arXiv.1702.08608*
6. *Francis, D., Blumenstock, J., & Robinson, J. (2017). Digital credit and mobile phones in Kenya: A new tool for financial inclusion? Center for Effective Global Action, Working Paper. https://cega.berkeley.edu*

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 7| July 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

7. *Friedler, S. A., Scheidegger, C., Venkatasubramanian, S., Choudhary, S., Hamilton, E. P., & Roth, D. (2019). A comparative study of fairness-enhancing interventions in machine learning. Proceedings of the Conference on Fairness, Accountability, and Transparency, 329–338. https://doi.org/10.1145/3287560.3287589*
8. *Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. Advances in Neural Information Processing Systems, 29, 3315–3323. https://proceedings.neurips.cc/paper_files/paper/2016/file/9d2682367c3935defcb1f9e247a97c0d-Paper.pdf*
9. *Hurley, M., & Adebayo, J. (2017). Credit scoring in the era of big data. Yale Journal of Law and Technology, 18, 148–216. https://digitalcommons.law.yale.edu/yjolt/vol18/iss1/5/*
10. *Jagtiani, J., & Lemieux, C. (2019). The roles of alternative data and machine learning in fintech lending: Evidence from theLendingClub consumer platform. Financial Management, 48(4), 1009–1029. https://doi.org/10.1111/fima.12295*
11. *Kamiran, F., & Calders, T. (2012). Data preprocessing techniques for classification without discrimination. Knowledge and Information Systems, 33(1), 1–33. https://doi.org/10.1007/s10115-011-0463-8*
12. *Kim, P. T. (2020). Data-driven discrimination at work. William & Mary Law Review, 58(3), 857–936. https://scholarship.law.wm.edu/wmlr/vol58/iss3/4*
13. *Liu, L., Dean, S., Rolf, E., Simchowitz, M., & Hardt, M. (2018). Delayed impact of fair machine learning. International Conference on Machine Learning, 3150–3158. https://proceedings.mlr.press/v80/liu18c.html*
14. *Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2021). Model cards for model reporting. Communications of the ACM, 64(12), 56–65. https://doi.org/10.1145/3454122*
15. *Morozov, E. (2013). To Save Everything, Click Here: The Folly of Technological Solutionism. PublicAffairs.*
16. *Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33–44. https://doi.org/10.1145/3351095.3372873*
17. *Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x*
18. *Suresh, H., & Guttag, J. V. (2021). A framework for understanding unintended consequences of machine learning. Communications of the ACM, 64(3), 62–71. https://doi.org/10.1145/3433941*
19. *Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International Data Privacy Law, 7(2), 76–99. https://doi.org/10.1093/idpl/ipx005*
20. *World Bank. (2022). The Global Findex Database 2021: Financial inclusion, digital payments, and resilience in the age of COVID-19. https://www.worldbank.org/en/publication/globalfindex*
21. *Zemel, R., Wu, Y., Swersky, K., Pitassi, T., & Dwork, C. (2013). Learning fair representations. International Conference on Machine Learning, 325–333. https://proceedings.mlr.press/v28/zemel13.html*
22. *Zliobaite, I. (2017). Measuring discrimination in algorithmic decision making. Data Mining and Knowledge Discovery, 31(4), 1060–1089. https://doi.org/10.1007/s10618-017-0506-1*