Journal of Science Research International (JSRI)
ISSN: 2456-6365 (Online)

OPEN ACCESS

ALGORITHMIC BIAS IN ALTERNATIVE CREDIT SCORING: A COMPARATIVE ANALYSIS OF SME FINANCIAL INCLUSION IN INDIA AND THE UNITED STATES

Abhijeet Parshuram Salvi*, Megha Mangesh Chavan, Mitali Devendra Deshmukh, Aarya Yashwant Gawade and Amisha Shivdas Kale

Department of Computer Science,

Pillai College of Arts, Commerce and Science (Empowered Autonomous), New Panvel

**Corresponding author E-mail: abhijeetsalvi@mes.ac.in*

Received: 21 January 2026	Revised: 17 February 2026	Accepted: 20 March 2026	Published: 13 April 2026

DOI: **https://doi.org/10.5281/zenodo.19554224**

Abstract:

AI-driven alternative credit scoring has expanded financial access for underserved Small and Medium Enterprises (SMEs), yet simultaneously risks amplifying structural inequalities through algorithmic bias embedded in training data, proxy variables, and model architecture. This paper makes three contributions: (i) a cross-national comparative analysis of alternative credit scoring ecosystems in India and the United States, mapping data sources, lending models, and bias typologies in each jurisdiction; (ii) a Bias-Source-to-Impact taxonomy that operationalises five bias forms—historical, selection, measurement, automation, and emergent—against identifiable SME borrower groups; and (iii) a comparative regulatory framework analysis contrasting India's RBI-led borrower-centric model with the US's multi-agency fair-lending regime. Findings confirm the "paradox of inclusion": alternative data improves credit access for thin-file SMEs while proxy variables encoding race, geography, and gender can deepen discrimination. The paper concludes that responsible AI governance—explainability mandates, continuous bias monitoring, and cross-jurisdictional regulatory coordination—is essential to realise equitable SME financing at scale.

Keywords: *Algorithmic Bias, Alternative Credit Scoring, SME Financial Inclusion, AI Fairness, Fintech, India, United States, Regulatory Frameworks, Explainable AI.*

1. Introduction

Small and Medium Enterprises (SMEs) constitute nearly 99% of global businesses and contribute 30% of India's GDP, yet face persistent exclusion from formal credit markets (1). Traditional credit scoring—built on credit histories, audited financials, and collateral—systematically disadvantages "thin-file" SMEs: new firms, informal-

sector enterprises, and women-led or rural businesses lacking the documented track records these models require (2, 6).

AI-driven alternative credit scoring has emerged as a disruptive solution, leveraging machine learning on non-traditional data—UPI transactions, GST filings, e-commerce behavior, psychometric signals—to extend credit access beyond conventional scoring's reach (3). Yet this technology embeds a structural paradox: the same data advantage that enables inclusion can encode and amplify existing socioeconomic inequalities when models learn from historically biased lending decisions or use proxy variables that indirectly represent protected characteristics (4, 5).

Despite a growing literature on both algorithmic bias and SME finance, cross-national comparative analysis remains sparse. This paper fills that gap by examining how algorithmic bias in alternative credit scoring affects SME financial inclusion differently across India and the United States—two economies that are large, digitally dynamic, yet structurally distinct in financial infrastructure, data availability, and regulatory philosophy. The study aims to: (a) map alternative credit scoring practices in each jurisdiction; (b) document identifiable bias patterns and their SME impact; (c) compare regulatory responses; and (d) derive governance recommendations for equitable, responsible AI in lending.

2. Literature review and bias taxonomy

Traditional credit scoring models treat SMEs as "opaque borrowers"—entities that lack the audited financial statements and credit histories required by models designed for large corporates (6). The consequence is a global SME finance gap exceeding USD 5 trillion, forcing small businesses into high-cost informal credit (7). Structural barriers compound model limitations: collateral requirements exclude asset-scarce firms; documentation gaps disqualify informal businesses; and self-exclusion removes owner-demographic segments before applications are even filed (8, 9).

Table 1: Bias Taxonomy: Source, Mechanism, and SME Impact

Bias Type	Source	Mechanism	Documented SME Impact
Historical	Legacy lending datasets	Models replicate past denial patterns	Women-led, minority-owned, rural SMEs under-scored
Selection	Unrepresentative training samples	Marginalised groups under-sampled	Gig workers, informal-sector SMEs excluded
Measurement	Proxy variables (ZIP code, device type)	Neutral vars encode protected traits	Geographic and demographic discrimination
Automation	Score over-reliance	Human oversight removed from decisions	Borderline applicants denied without appeal
Emergent	Model drift over time	New discriminatory patterns develop post-deployment	Older adults (55+), immigrants, seasonal businesses

AI-based alternative credit scoring addresses data scarcity by incorporating high-frequency transactional signals, behavioral data, and external environmental data (11). Machine learning ensembles such as XGBoost and Gradient Boosting demonstrate superior predictive power over logistic regression, particularly during economic stress periods (12). However, these performance gains mask a critical risk: when models train on historical lending data that reflects past discrimination, they learn to reproduce those patterns at scale (13). Five bias types have been

identified in the literature: (i) historical bias—legacy discrimination encoded in training data; (ii) selection bias—unrepresentative samples under-counting marginalized groups; (iii) measurement bias—variables that proxy protected characteristics; (iv) automation bias—over-reliance on algorithmic scores over human judgment; and (v) emergent bias—discriminatory patterns arising over time through model drift (14).

The regulatory response to these risks diverges sharply between India and the US. India's RBI has adopted a centralized, borrower-centric approach; the US operates a fragmented multi-agency model anchored in civil-rights-era anti-discrimination statutes. Both demand explainability, grievance redress, and third-party accountability, but enforcement mechanisms and coverage differ materially (15, 16).

3. Methodology

This study employs a qualitative, descriptive comparative design. India and the United States were selected for their large SME sectors, contrasting levels of digital financial infrastructure, and structurally different regulatory regimes—conditions that maximize analytical leverage in a cross-national comparison (17). The exclusive reliance on secondary data reflects the proprietary nature of credit-scoring algorithms and lending datasets: financial institutions and fintech firms do not disclose model architecture, training data, or decisioning logic, making primary data collection impractical at this scope (18).

Secondary sources were retrieved from Google Scholar, JSTOR, SSRN, and official publications of the RBI, CFPB, FTC, World Bank, OECD, and major credit bureaus (Experian, Equifax, TransUnion) covering 2019–mid-2026. Purposive sampling selected literature addressing alternative credit scoring, algorithmic bias, SME financial inclusion, or AI regulatory governance (19, 20). Thematic content analysis organised findings around six dimensions: data source types, model architectures, bias typologies, regulatory instruments, mitigation tools, and cross-jurisdictional contrasts. Triangulation across academic, regulatory, and industry sources strengthens validity; consistent category application across both country cases enhances reliability.

4. Results: Comparative credit scoring landscape

Table 2: Alternative Credit Scoring Practices: India vs. United States (2026)

Dimension	India	United States
Traditional base	CIBIL credit bureau score	FICO, VantageScore
Primary alt. data	UPI transactions, GST returns, EPFO/PF contributions, utility & telecom payments, mobile wallet activity	Bank cash-flow via Plaid, rent/utility payments, gig-economy income (Uber, DoorDash), BNPL history, paycheck data
Digital footprints	E-commerce transactions, smartphone behavioural data, psychometric test tokens	Device intelligence, social media signals (cautious use), app-usage behaviour
ML models used	XGBoost, Gradient Boosting, cash-flow models, activity-based underwriting, co-lending (bank + fintech)	FICO Score XD, hybrid scoring, cash-flow models, psychometric models, fraud-detection DL models
Key platform	Unified Lending Interface (ULI), AA (Account Aggregator) framework	Open-banking via Plaid; CFPB Personal Financial Data Rights Rule
Target segment	Credit-invisible & thin-file SMEs, gig workers, micro-entrepreneurs	~49 million underserved adults lacking conventional credit data

4.1. Evidence of algorithmic bias

Across both jurisdictions, models trained on legacy lending data perpetuate historical discrimination against women-led, minority-owned, and informal enterprises. Proxy variable bias is pervasive: ZIP code, smartphone make, and digital behaviour patterns serve as indirect encoders of race, age, and socioeconomic status—legally neutral variables that produce disparate outcomes in violation of the spirit of anti-discrimination statutes (14). Intersectional bias—where race and gender compound disadvantage—is systematically under-detected by single-attribute fairness tests. Bias drift, whereby models develop new discriminatory patterns post-deployment as social conditions shift, disproportionately affects older adults (55+), immigrants, and seasonal businesses. Mitigation tools adopted by 2026 include SHAP and LIME for explainability, counterfactual fairness testing, and continuous bias-monitoring dashboards.

4.2. Regulatory framework comparison

Table 3: Regulatory Framework: India vs. United States

Feature	India	United States
Model	Centralised (RBI-led)	Fragmented multi-agency (CFPB, FTC, OCC, state regulators)
Primary statutes	RBI Digital Lending Guidelines 2022; DPDP Act 2023	ECOA, FCRA, GLBA, CFPB Personal Financial Data Rights Rule
Credit reporting	High-frequency updates (5×/month); mandatory borrower pre-notification before default reporting	Annual report entitlement; FCRA adverse-action notice requirement
Explainability	Mandatory model explainability for digital lenders; grievance redress within 30 days	Adverse-action reason codes required; CFPB scrutiny of AI explanations growing
Data governance	Strict consent requirements; data localisation (DPDP Act)	GLBA financial privacy; no federal data-localisation mandate
Third-party liability	Fintech intermediary liability under RBI guidelines	Emerging CFPB guidance; no comprehensive fintech liability statute
Dispute resolution	Compensation penalties for delayed resolution	FCRA dispute mechanism; CFPB complaint portal

5. Discussion

The findings confirm the paradox of inclusion at a structural level. Alternative data demonstrably improves credit access for thin-file SMEs—a genuine financial inclusion gain validated across both countries. Yet the same AI systems encode historical discrimination when trained on legacy lending data, deploy proxy variables that indirectly encode protected characteristics, and generate intersectional disadvantages that single-attribute bias tests fail to surface. The paradox is not a technical anomaly; it is a predictable consequence of applying optimisation logic to historically inequitable datasets (5, 13, 14).

The regulatory comparison reveals a substantive philosophical divergence. India's RBI model is proactive, prescriptive, and borrower-centric: it mandates explainability, sets update frequency, assigns dispute compensation, and localises data. The US model is reactive and distributed: anti-discrimination statutes were

written for human lenders, and their application to black-box ML models remains contested, with the CFPB increasingly relying on supervisory guidance rather than statute. Neither model is currently sufficient: India's prescriptive rules risk regulatory lag as models evolves; the US's fragmentation creates compliance arbitrage opportunities for fintech lenders operating across state lines (15, 16).

Three governance interventions emerge as critical. First, explainable AI mandates—requiring SHAP-based or equivalent model-level explanations for adverse decisions—must be standardised across jurisdictions, not left to supervisory discretion. Second, continuous bias monitoring, treating fairness metrics as operational KPIs equivalent to model accuracy, must be embedded in model lifecycle governance. Third, cross-jurisdictional regulatory coordination between the RBI and US agencies (CFPB, OCC) on shared standards for alternative data use, proxy-variable restrictions, and third-party audit requirements would reduce the regulatory arbitrage that currently allows discriminatory models to persist in less-scrutinised lending channels (4, 15, 16).

Conclusion and future directions

This paper demonstrates that the "paradox of inclusion" is structurally embedded in current AI-based credit scoring practice across both India and the United States. Alternative data expands access for underserved SMEs—but without robust governance, the same systems deepen inequalities along gender, race, geography, and age lines. The comparative analysis shows that India's centralised, borrower-centric regulatory model provides stronger borrower protections than the US's fragmented approach, yet both fall short of the proactive, algorithmic-lifecycle oversight required to address bias drift and intersectional discrimination.

Three future research directions are high-priority. (i) Empirical bias auditing: field studies with real lending data across both jurisdictions would move the evidence base from documented pattern-observation to causal measurement, directly informing regulatory thresholds. (ii) Federated fairness frameworks: privacy-preserving federated learning approaches could enable cross-institution bias monitoring without exposing proprietary model architectures or customer data, addressing the primary obstacle to regulatory oversight. (iii) Regulatory convergence modeling: simulation studies examining the credit-access and bias-reduction effects of harmonising RBI and US fair-lending standards would provide quantitative foundations for international policy coordination—currently absent from both academic and regulatory discourse.

Acknowledgements

The authors thank the Department of Computer Science at Pillai College of Arts, Commerce and Science for research facilities and support, and Instructor Abhijeet Salve for guidance throughout this project.

References

1. Agarwal, S., & Zhang, J. (2020). FinTech, lending and payment innovation: A review. *Asia-Pacific Journal of Financial Studies, 49*(3), 353–367. https://doi.org/10.1111/ajfs.12294
2. Barocas, S., & Hardt, M. (2019). *Fairness and machine learning: Limitations and opportunities*. https://fairmlbook.org
3. Berg, T., Burg, V., Gombočić, A., & Puri, M. (2020). On the rise of fintech credit reporting. *Review of Financial Studies, 33*(7), 2874–2923. https://doi.org/10.1093/rfs/hhz099
4. Consumer Financial Protection Bureau. (2022). *Supervisory highlights: Fair lending*. U.S. Government.
5. Cowgill, B., Dell'Acqua, F., Deng, S., Hylton, K., Kotcherlakota, N., & Perkowski, J. (2020). Bias and productivity in humans and algorithms. *ACM Conference on Economics and Computation*.

6. Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring. *Applied Soft Computing, 91*, 106263. https://doi.org/10.1016/j.asoc.2020.106263
7. Cornelli, G., Frost, J., Gambacorta, L., & Jagtiani, J. (n.d.). *The impact of fintech lending on credit access for U.S. small businesses*. Bank for International Settlements.
8. Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? *Journal of Finance, 77*(1), 5–47. https://doi.org/10.1111/jofi.13090
9. International Finance Corporation. (2017). *MSME finance gap*. World Bank Group.
10. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1–35. https://doi.org/10.1145/3457607
11. Organisation for Economic Co-operation and Development. (2022). *Financing SMEs and entrepreneurs 2022*. OECD Publishing. https://doi.org/10.1787/e9073a0f-en
12. Reserve Bank of India. (2022). *Guidelines on digital lending* (RBI/2022-23/111).
13. Rona-Tas, A. (2017). The off-label use of consumer credit ratings. *Historical Social Research, 42*(1), 52–76.
14. Sahay, R., *et al.* (2020). *The promise of fintech: Financial inclusion in the post COVID-19 era* (IMF Departmental Paper No. 2020/009).
15. Siddiqi, N. (2017). *Intelligent credit scoring* (2nd ed.). Wiley.
16. Tan, Y., Floros, C., & Anchor, J. (2017). The profitability of Chinese banks. *Review of Accounting and Finance, 16*(1), 86–105.
17. World Bank. (2020). *Financial inclusion overview*. https://www.worldbank.org/en/topic/financialinclusion/overview
18. Xu, M., *et al.* (2005). Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment, 97*(3), 322–336.
19. Zest AI. (2023). *How AI-powered credit underwriting is changing lending* (White paper).
20. Elizabeth Kumar, I., Hines, K. E., & Dickerson, J. P. (n.d.). *Equalizing credit opportunity in algorithms: Aligning algorithmic fairness research with U.S. fair lending regulation*.