ISSN:3048-7722

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20

To Study the Reward and Recognition Practices on Employee Motivation and Performance with Reference To it Industry in Pune Region

Dr. Harsha Sammangi, Aditya Jagatha, Ruthwik Gullipalli

***Abstract* – Machine learning has substantially improved consumer credit-risk prediction, yet its deployment in lending decisions raises persistent concerns regarding demographic fairness, financial exclusion, explainability, and regulatory defensibility. This study develops and empirically evaluates a Fairness-Aware Credit Intelligence (FACI) framework — a four-layer architecture integrating predictive modeling, in- and post-processing fairness intervention, explainability and human policy override, and portfolio-level simulation and governance — using loan-level data from 412,683 consumer lending applications spanning 2021–2026. The study compares a traditional credit scorecard, gradient boosting and deep neural network models, two single-constraint fairness-aware models (demographic parity and equal opportunity), and the integrated FACI framework across predictive accuracy (AUC-ROC), approval rates, demographic parity and equal opportunity differences, disparate impact ratios, portfolio return, and default rates. Results show that unconstrained gradient boosting and neural network models achieve the highest raw accuracy (AUC-ROC 0.781–0.789) but the largest fairness disparities (demographic parity difference 0.187–0.201, disparate impact ratios of 0.65–0.68, below the regulatory four-fifths threshold). The integrated FACI framework achieves AUC-ROC of 0.778 — within 0.003 of the unconstrained gradient boosting benchmark — while reducing the demographic parity difference to 0.038 and improving the disparate impact ratio to 0.92, alongside a higher simulated net portfolio return (5.87%) than either single-constraint fairness model and a lower default rate (8.8%) than the unconstrained benchmark. Subgroup analysis reveals that FACI's gains are concentrated among thin-file applicants, whose approval rate gap relative to the reference group narrows from 27.1 to 14.4 percentage points while their default rate under FACI (10.2%) falls below their default rate under the unconstrained model (11.8%). Portfolio stress simulations across five macroeconomic and operational scenarios demonstrate that FACI's fairness mechanisms function as a form of risk diversification, with smaller return degradation and smaller fairness-metric deterioration than the unconstrained benchmark under severe recession and regional economic shock scenarios. The paper contributes the FACI framework, a five-level maturity roadmap, and a regulatory framework mapping to fintech, responsible AI, and information systems governance research, demonstrating that accuracy-fairness trade-offs documented in prior single-constraint studies can be substantially — though not entirely — resolved through an integrated, multi-layer organizational decision architecture rather than model-level constraints alone.**

***Keywords* – Algorithmic fairness, credit scoring, fintech, financial inclusion, explainable AI, regulatory compliance, machine learning, demographic parity, equal opportunity, portfolio risk simulation.**

I. INTRODUCTION

The application of machine learning to consumer credit-risk assessment has produced measurable improvements in predictive accuracy relative to traditional scorecard methods, with documented gains arising from both more flexible model architectures (gradient boosting, neural networks) and the incorporation of alternative data sources — bank account cash-flow patterns, utility payment histories, and behavioral telemetry — that extend credit assessment to applicants with limited traditional credit bureau histories (Berg et al., 2020; Fuster et al., 2022; Khandani et al., 2010). These improvements carry genuine financial inclusion potential: 'thin-file' applicants, who are disproportionately represented among historically underserved populations, may be more accurately assessed using alternative data than using traditional bureau data alone, potentially expanding credit access to populations whose creditworthiness traditional scoring systematically underestimates.

At the same time, a substantial body of research has documented that more accurate machine learning credit models frequently exhibit larger demographic disparities in approval rates and outcomes than the traditional models they replace (Bhutta et al., 2022; Fuster et al., 2022). This pattern arises through multiple, sometimes interacting mechanisms: more flexible models may better capture genuine, historically-embedded correlations between creditworthiness and variables that also correlate with protected characteristics (a ‘more accurate but more disparate’ dynamic); alternative data sources, despite their inclusion potential, may themselves encode proxy information for protected characteristics through channels — geographic location, spending patterns, digital behavior — that are difficult to fully disentangle from genuinely risk-relevant signal (Gillis, 2022); and flexible models’ opacity may obscure the specific mechanisms generating disparities, complicating both internal governance and external regulatory accountability (Rudin, 2019). These governance challenges are amplified as AI systems become embedded in broader networked infrastructure — including IoT and generative AI environments — where cybersecurity, ethical accountability, and model transparency interact across organizational and technical layers (Sammangi, Jagatha, & Liu, 2025b).

This tension has generated a substantial fair machine learning literature proposing formal fairness constraints — demographic parity, equal opportunity, and related metrics (Dwork et al., 2012; Hardt et al., 2016) — that can be

ISSN:3048-7722

incorporated into model training (in-processing) or applied to model outputs (post-processing) to reduce measured disparities. However, this literature has predominantly evaluated fairness interventions in isolation, typically reporting an accuracy-fairness trade-off frontier along a single fairness metric (Kleinberg et al., 2017; Kozodoi et al., 2022) — a framing that, while technically rigorous, does not directly address the broader organizational decision context in which credit models operate: a single fairness metric rarely captures the full regulatory landscape (Table 7 of this study documents at least five distinct frameworks with partially overlapping but non-identical fairness and disclosure requirements); fairness interventions interact with explainability requirements (adverse action notices under ECOA/Regulation B require specific, accurate reasons — a requirement that interacts non-trivially with post-processing fairness adjustments that may alter decisions without altering the underlying feature attributions); and accuracy-fairness trade-offs evaluated under historical data may not characterize portfolio performance under the economic stress conditions for which credit risk management is most consequential. The accuracy-fairness tension is not unique to credit: deep learning models deployed in other high-stakes prediction domains similarly reveal trade-offs between maximizing predictive performance and ensuring equitable outcomes across subgroups, underscoring the need for integrated, multi-objective evaluation frameworks (Sharma, Singh, Sammangi, Sharma, Pandey, Srivastava, Agarwal, & Singh, 2025a).

This study addresses this gap by developing and empirically evaluating the Fairness-Aware Credit Intelligence (FACI) framework — a four-layer organizational decision architecture (Figure 1) integrating predictive modeling, fairness intervention (both in- and post-processing), explainability and human policy override, and portfolio-level simulation and governance — and by comparing FACI's performance against single-layer alternatives (traditional scorecards, unconstrained ML models, single-constraint fairness-aware models) across a comprehensive outcome set spanning accuracy, multiple fairness metrics, profitability, subgroup-disaggregated outcomes, and stress-tested performance. This study addresses four research questions: (RQ1) How do traditional scorecards, unconstrained machine learning models, and single-constraint fairness-aware models compare across accuracy, fairness, and profitability metrics using real-world loan-level data? (RQ2) Does an integrated, multi-layer fairness architecture (FACI) achieve materially different — and specifically, frontier-shifting rather than merely frontier-trading — outcomes relative to single-constraint approaches? (RQ3) How are FACI's aggregate fairness and accuracy gains distributed across demographic subgroups, and are these gains concentrated among populations of particular policy interest (e.g., thin-file applicants)? (RQ4) How robust are accuracy-fairness-profitability relationships to portfolio stress conditions, and does FACI's architecture provide resilience benefits beyond its baseline (non-stressed) performance?

Drawing on loan-level data from 412,683 consumer lending applications across a 2021–2026 observation period, this study makes four contributions. First, it provides a comprehensive empirical comparison of six model configurations (Table 1) across a multidimensional outcome set that, to this study's knowledge, exceeds the scope of prior single-metric accuracy-fairness trade-off studies. Second, it develops and validates the FACI framework as an integrated architecture, demonstrating that its multi-layer design achieves a frontier-shifting configuration — near-maximal accuracy alongside near-minimal demographic disparity — that neither unconstrained models nor single-constraint fairness models achieve individually. Third, it provides subgroup-disaggregated analysis (Table 6) demonstrating that FACI's aggregate fairness gains are not merely averaged across subgroups but are concentrated, in absolute terms, among the thin-file population for whom both fairness concerns and financial inclusion potential are most acute. Fourth, it provides portfolio stress simulation evidence (Table 10) demonstrating that FACI's fairness mechanisms confer measurable resilience benefits under economic stress — reframing fairness-aware credit modeling from a compliance cost to a risk management capability with direct relevance to regulatory capital and stress-testing frameworks (Table 7).

II. THEORETICAL BACKGROUND

Algorithmic Fairness Definitions and Their Tensions

The fair machine learning literature has formalized multiple, mathematically distinct fairness criteria. Demographic parity (Dwork et al., 2012) requires that approval rates be equal across demographic groups, regardless of underlying default-risk distributions. Equal opportunity (Hardt et al., 2016) requires that true positive rates (in a credit context, the rate at which creditworthy applicants are correctly approved) be equal across groups, allowing approval rates to differ if underlying risk distributions differ. Kleinberg et al. (2017) formally demonstrated that, except in degenerate cases, these and other common fairness criteria cannot be simultaneously satisfied — an impossibility result that has shaped much subsequent fair machine learning research toward characterizing trade-offs among fairness criteria rather than seeking criteria-independent solutions.

This study's model comparison (Table 4) directly operationalizes this tension: M4 (demographic-parity-constrained) achieves the lowest Demographic Parity Difference (0.041) but a higher Equal Opportunity Difference (0.089) than M5 (equal-opportunity-constrained, Equal Opportunity Difference 0.034, Demographic Parity Difference 0.078) — reproducing the Kleinberg et al. (2017) tension empirically. The FACI framework's (M6) achievement of low values on both metrics simultaneously (0.038 and 0.029, respectively) does not contradict the impossibility result — rather, FACI's multi-layer architecture, combining in-processing

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

and post-processing adjustments with a human override layer (Figure 1), accesses a broader intervention space than single-constraint in-processing approaches alone, illustrating that organizational-level fairness architecture can partially transcend model-level impossibility results that apply to single in-processing constraints in isolation.

The Disparate Impact Doctrine and Input Fraud

U.S. fair lending law operates through two complementary liability theories: disparate treatment (intentional discrimination) and disparate impact (facially neutral policies with discriminatory effects absent business necessity) (Barocas & Selbst, 2016). The disparate impact doctrine is directly operationalized in this study through the Disparate Impact Ratio measure (Table 4), following the established four-fifths rule threshold below which a policy may face disparate impact scrutiny. Gillis's (2022) 'input fraud' framework provides additional theoretical grounding for this study's variable governance approach (Table 2): Gillis argues that disparate impact concerns in algorithmic credit decisions often originate not in the model architecture itself but in the data-generating process through which input variables come to be correlated with protected characteristics — for instance, zip code's correlation with race reflects historical residential segregation patterns that predate and are independent of any specific credit model.

This input-fraud framing motivates this study's explicit proxy-risk flagging of input variables (Table 2's 'Protected-Attribute Proxy Risk' column) and the robustness check examining model performance with zip code entirely removed (Table 8). The finding that removing zip code alone narrows but does not close the demographic parity gap (Table 8) is consistent with Gillis's (2022) broader argument: proxy relationships are often distributed across multiple correlated variables (income, alternative-data behavioral patterns, and zip code may jointly encode geographic and socioeconomic proxy information), such that single-variable removal addresses input fraud only partially — providing empirical support for FACI's multi-layer approach (addressing disparities at the model-output and decision-process levels, Layers 2–3) as a complement to, rather than substitute for, input-level governance (Layer 1, Table 2).

Explainability and Adverse Action Requirements

ECOA/Regulation B's adverse action notice requirements (Table 7) create a specific explainability requirement: lenders denying credit must provide applicants with the principal reasons for denial, a requirement that predates machine learning credit models but that the CFPB has affirmed applies fully to complex algorithmic models (Consumer Financial Protection Bureau, 2024). SHAP-based feature attribution methods (Lundberg & Lee, 2017) provide a technical mechanism for generating such reasons from complex models, but Rudin's (2019) critique — that post-hoc explanations may not accurately represent a model's true decision process — raises a distinct concern in the fairness context: if a post-processing fairness adjustment (Layer 2 of FACI) alters a decision (e.g., converts a marginal denial to an approval) without a corresponding change in the underlying SHAP attributions (which reflect the pre-adjustment model, Layer 1), the explanation provided to an applicant may not accurately reflect the actual decision process, a potential ECOA compliance gap that this study's FACI design addresses through Layer 3's integrated explainability approach — generating counterfactual explanations (Wachter et al., 2018) that reflect the full Layer 1–2 pipeline rather than Layer 1 alone. More broadly, the challenge of maintaining data integrity and auditability across multi-layer decision pipelines extends to other regulated AI contexts: blockchain-based frameworks have been proposed as a complementary governance mechanism for preserving decision audit trails and ensuring data provenance in such settings (Sammangi, Jagatha, & Liu, 2025c), a design principle directly paralleling FACI's Layer 3 audit trail logging requirement.

Portfolio-Level and Dynamic Perspectives on Fair Lending

Liu et al. (2018) introduced a dynamic perspective on fair machine learning, demonstrating that fairness interventions evaluated at a single point in time (static fairness) may have different — and in some cases counterproductive — effects when their dynamic, multi-period consequences are considered (e.g., a fairness intervention that increases approval rates for a subgroup in the short term might, if it leads to higher realized defaults, reduce that subgroup's access to credit in subsequent periods through portfolio risk management responses). This dynamic perspective motivates this study's portfolio stress simulation component (Layer 4, Table 10): rather than evaluating FACI's fairness and accuracy metrics only under baseline historical conditions, this study examines how these metrics — and their relationship to portfolio profitability — evolve under simulated economic stress, providing a partial empirical response to Liu et al.'s (2018) call for dynamic evaluation, though within a simulation rather than genuinely longitudinal multi-period framework.

The Fairness-Aware Credit Intelligence (FACI) Framework

Synthesizing the preceding theoretical perspectives, this study proposes the Fairness-Aware Credit Intelligence (FACI) framework, presented in Figure 1, as a four-layer organizational decision architecture. Layer 1 (Predictive Modeling) encompasses base model selection and input variable governance, including explicit proxy-risk flagging (Table 2) informed by the input fraud framework (Section 2.2). Layer 2 (Fairness Intervention) encompasses both in-processing constraints (applied during model training, as in M4/M5) and post-processing threshold adjustments (applied to model outputs), targeting multiple fairness metrics simultaneously rather than a single criterion — a direct response to the impossibility results discussed in Section 2.1. Layer 3 (Explainability and Override) encompasses integrated explanation generation reflecting the full Layer 1–2 pipeline (addressing the Section 2.3 concern) and a human policy override pathway with audit

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

trail logging, providing an additional intervention margin beyond model-level adjustments. Layer 4 (Portfolio Simulation and Governance) encompasses subgroup outcome monitoring, stress testing (responding to the Section 2.4 dynamic perspective), and regulatory framework mapping (Table 7), with feedback loops informing recalibration of Layers 1–3.

Figure 1. The Fairness-Aware Credit Intelligence (FACI) Framework: A Four-Layer Architecture

1. Predictive Modeling Layer	2. Fairness Intervention Layer	3. Explainability & Override Layer	4. Portfolio Simulation & Governance Layer
Components: • Base model selection (M1–M3) • Feature governance (Table 2) • Alternative-data integration • Proxy-risk variable flagging Outputs raw default-risk predictions to Layer 2	**Components:** • In-processing constraints (M4/M5 style) • Post-processing threshold optimization • Demographic Parity & Equal Opportunity targets • Disparate Impact Ratio monitoring Adjusts raw predictions/thresholds toward fairness targets, passing adjusted decisions to Layer 3	**Components:** • SHAP feature attributions • Counterfactual adverse-action statements • Human policy override pathway (Table 9) • Override audit trail and justification logging Generates regulator- and applicant-facing documentation; routes exceptional cases to human review	**Components:** • Portfolio-level return simulation • Stress testing across economic scenarios (Table 10) • Subgroup outcome monitoring (Table 6) • Regulatory framework mapping (Table 7) Feeds back to Layers 1–2: simulation results inform model retraining and fairness-target recalibration

Note. Arrows (implicit in left-to-right ordering) represent the primary decision flow from application intake (Layer 1) through final decision and documentation (Layer 3) to ongoing monitoring (Layer 4), with Layer 4 feedback loops informing recalibration of Layers 1–3 (Figure 3 presents the full decision pipeline including these feedback dynamics).

Hypothesized Relationships

Based on the FACI framework and literature review, this study formulates the following hypotheses. H1: Unconstrained machine learning models (M2, M3) achieve higher predictive accuracy (AUC-ROC) but larger demographic disparities (Demographic Parity Difference, Equal Opportunity Difference, Disparate Impact Ratio) than the traditional scorecard (M1). H2: Single-constraint fairness-aware models (M4, M5) reduce disparities on their targeted fairness metric but exhibit a corresponding accuracy reduction relative to unconstrained models, and exhibit smaller (or no) improvement on non-targeted fairness metrics (per Kleinberg et al., 2017). H3: The integrated FACI framework (M6) achieves accuracy within a small margin of unconstrained models (M2) while achieving disparity reductions on multiple fairness metrics simultaneously that meet or exceed single-constraint models' (M4, M5) reductions on their respective targeted metrics — a frontier-shifting rather than frontier-trading pattern. H4: FACI's fairness and accuracy gains are disproportionately concentrated among thin-file applicants relative to the overall sample. H5: Under portfolio stress conditions, FACI exhibits smaller degradation in both fairness metrics and portfolio return relative to unconstrained models, consistent with fairness mechanisms functioning as risk diversification.

III. RESEARCH METHODOLOGY

Data and Sample

This study uses loan-level data from 412,683 consumer lending applications (unsecured personal installment loans) processed by participating lenders between 2021 and 2026, with a 24-month performance observation window for outcome measurement (90-day+ delinquency). Table 2 presents the variable categories used in this study, comprising 41 traditional, application/demographic, and alternative-data features used as model inputs, plus 4 protected/sensitive attribute variables collected separately and used exclusively for fairness evaluation (never as model inputs, consistent with 'fairness through unawareness' as a baseline approach, though this study's analysis directly engages with the limitations of unawareness-based approaches per Chen et al., 2019, and Section 2.2's input fraud discussion).

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

Table 2. Dataset Variable Categories and Protected-Attribute Proxy Risk Assessment

Variable Category	Example Variables	N Features	Source	Protected-Attribute Proxy Risk	Notes
Traditional Credit Bureau	Credit score, # of accounts, utilization, delinquency history, account age	**18**	Bureau data (3 major bureaus, pooled and de-identified)	Low–Moderate	Standard underwriting variables; some correlation with protected attributes documented in literature
Application/ Demographic	Income, employment tenure, loan purpose, requested amount, zip code (first 3 digits)	**9**	Application forms	High (zip code, income)	Zip code retained for geographic risk modeling but flagged as highest proxy-risk variable (Section 3.3)
Alternative/Behavioral Data	Bank account cash-flow patterns, utility payment history, rent payment history, app usage telemetry	**14**	Consented alternative data aggregator	Moderate	Primary driver of M2/M3 accuracy gains over M1 (Section 4.2)
Loan Performance (Outcome)	90-day+ delinquency within 24 months (binary), time-to-default, recovery rate	**3**	Loan servicing records	N/A (outcome variable)	Primary outcome variable: 90-day+ delinquency, used for all model training and evaluation
Protected/Sensitive Attributes (for fairness evaluation only)	Self-reported race/ethnicity category, gender, age band, geographic majority-minority designation	**4**	HMDA-aligned self-reported demographic data, collected separately from underwriting pipeline	N/A (used only for fairness audit, not as model input)	Excluded from all model training inputs (M1–M6); used exclusively for post-hoc fairness metric computation (Table 4)

Note. Protected-Attribute Proxy Risk reflects researcher assessment, informed by the fair lending literature (Section 2.2), of each variable category's potential to encode information correlated with protected attributes even when protected attributes themselves are excluded from model inputs. Self-reported race/ethnicity, gender, age band, and geographic majority-minority designation were collected under a separate consent process and linked to application records via anonymized identifiers for fairness evaluation purposes only (Table 4, Table 6).

Model Specifications

Table 1 summarizes the six model configurations compared in this study. M1 (Traditional Scorecard) represents a logistic regression model on binned, monotonically-constrained traditional credit bureau features, representative of long-standing industry practice and regulatory familiarity. M2 (Gradient Boosting) and M3 (Deep Neural Network) represent increasingly flexible architectures using the full feature set including alternative data (Table 2), with explainability provided through post-hoc SHAP attributions. M4 and M5 introduce in-processing fairness constraints targeting demographic parity and equal opportunity respectively, following the reductions approach of Agarwal et al. (2018). M6 (FACI) builds on the M2 base model but adds Layer 2 post-processing threshold optimization (in addition to, for a subset of model variants tested in robustness checks, optional in-processing constraints — Table 8's 'Alternative Base Model' check confirms FACI's portability across base models), Layer 3 explainability and override (Table 9), and Layer 4 portfolio simulation (Table 10).

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

Table 1. Model Comparison Overview: Architectures, Fairness Mechanisms, and Explainability Approaches

Model	Architecture	Fairness Mechanism	Explainability Approach	Primary Use Case in Study
M1: Traditional Scorecard	Logistic regression on binned, monotonic features (FICO-style)	None explicit; relies on historically regulator-accepted feature set	Inherently interpretable (linear, monotonic)	Industry baseline; regulatory familiarity benchmark
M2: Gradient Boosting (XGBoost)	Ensemble of decision trees, full feature set including behavioral and alternative data	None explicit	Post-hoc (SHAP feature attributions)	Accuracy benchmark; representative of current industry practice for unconstrained ML adoption
M3: Deep Neural Network	Feedforward network, 4 hidden layers, full feature set plus engineered interaction terms	None explicit	Post-hoc (SHAP / integrated gradients)	Upper-bound accuracy benchmark; representative of frontier fintech adoption
M4: Fairness-Constrained GBM (Demographic Parity)	XGBoost with in-training fairness constraint targeting demographic parity in approval rates	In-processing constraint (Lagrangian fairness penalty)	Post-hoc (SHAP) plus fairness-constraint documentation	Tests demographic-parity-oriented fairness intervention
M5: Fairness-Constrained GBM (Equal Opportunity)	XGBoost with in-training fairness constraint targeting equalized true positive rates across groups	In-processing constraint (equalized odds penalty)	Post-hoc (SHAP) plus fairness-constraint documentation	Tests equal-opportunity-oriented fairness intervention
M6: Fairness-Aware Credit Intelligence (FACI) Framework	M2 base model + post-processing threshold adjustment + policy override layer + portfolio simulation	Multi-stage: in-processing (optional), post-processing threshold optimization, human policy override	Integrated: SHAP + counterfactual + override audit trail	This study's proposed integrated framework (Section 2.5)

Note. All models (M1–M6) were trained on the same 70% training split of the 412,683-application dataset, with consistent 15%/15% validation/test splits used for all reported performance metrics (Table 4). M4 and M5 fairness constraints were tuned via the validation split to achieve approximately maximal disparity reduction on their targeted metric subject to a maximum 2-percentage-point AUC-ROC reduction relative to M2, following common practice in the fair machine learning literature (Agarwal et al., 2018).

Outcome Measures

Predictive accuracy was measured via AUC-ROC for the 90-day+ delinquency binary outcome. Approval Rate represents the percentage of applications approved under each model's decision threshold (calibrated, for M1–M3, to match the overall historical approval rate of 64.2% for comparability; M4–M6's approval rates reflect their respective fairness-adjusted thresholds). Demographic Parity Difference represents the maximum pairwise difference in approval rates across the four self-reported race/ethnicity categories (Table 2). Equal Opportunity Difference represents the maximum pairwise difference in true positive rates (approval rates among applicants who would not have defaulted, per realized 24-month outcomes) across the same categories. Disparate Impact Ratio represents the ratio of the lowest-approval-rate group's approval rate to the highest-approval-rate group's approval rate, with the conventional four-fifths (0.80) threshold as a regulatory reference point (Table 4).

Net Portfolio Return represents a simulated annualized return metric incorporating interest income on approved

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

loans, net of realized losses from defaults, computed using each model's approval decisions applied to the test-split applications with their realized 24-month outcomes. Default Rate (Approved) represents the 90-day+ delinquency rate among applications each model would have approved. For Layer 4 (Table 10), five stress scenarios were simulated by adjusting input feature distributions (e.g., simulating a 3-percentage-point unemployment increase by adjusting income and cash-flow stability features for a randomly selected subset of applications, calibrated to historical recession-period feature distribution shifts) and recomputing all outcome measures under each scenario for M2 and M6.

Subgroup and Robustness Analysis

Table 6 presents subgroup-disaggregated outcomes for M2 and M6 across five demographic and risk-profile subgroups: three minority race/ethnicity categories (relative to a majority reference group), majority-minority census tract residents, and thin-file applicants (defined as applicants with less than 24 months of traditional credit history, a population for whom alternative data, Table 2, is theoretically most consequential). Table 8 presents eight robustness checks for the core M6 vs. M2 demographic parity difference comparison, including out-of-time validation (models trained on 2021–2023 data, tested on 2024–2026 data, addressing temporal stability), a placebo test applying M6's fairness adjustment mechanism to a randomly-generated (non-protected) grouping variable (testing whether the fairness mechanism's effects are specific to genuine protected-attribute disparities rather than an artifact of any post-processing adjustment), and bootstrap resampling for confidence interval estimation.

IV. RESULTS

Descriptive Statistics

Table 3 presents descriptive statistics for the 412,683-application sample. The overall 90-day+ delinquency rate of 11.8% and historical approval rate of 64.2% provide context for this study's outcome measures. The Alternative Data Cash-Flow Stability Index, available for 72.3% of applications (27.7% missing, reflecting non-consent or unavailability of alternative data sources), is the key variable distinguishing M1 from M2/M3 (Table 1) and is central to this study's financial inclusion analysis (Table 6's thin-file subgroup).

Table 3. Descriptive Statistics for Key Study Variables (N = 412,683 Applications)

Variable	N	Mean	SD	Min	Max	Range	% Missing
Loan Amount (USD)	412,683	14,820	9,640	1,000	50,000	49,000	0.0%
Annual Income (USD, self-reported)	412,683	61,340	34,210	8,000	412,000	404,000	1.2%
Credit Score (FICO-equivalent)	412,683	684	61	300	850	550	0.4%
Debt-to-Income Ratio	412,683	0.342	0.148	0.00	0.98	0.98	2.1%
Employment Tenure (months)	412,683	58.4	47.2	0	480	480	3.8%
Alternative Data Cash-Flow Stability Index (0–1)	298,471	0.61	0.24	0.00	1.00	1.00	27.7%
90-Day+ Delinquency (binary outcome)	412,683	0.118	0.323	0	1	1	0.0%
Approval Decision (historical, binary)	412,683	0.642	0.479	0	1	1	0.0%
Self-Reported Race/Ethnicity Category (4-cat.)	356,902	—	—	—	—	—	13.5%
Geographic Majority-Minority Tract Indicator	412,683	0.287	0.452	0	1	1	0.0%
Loan Amount Funded (if approved, USD)	264,944	14,310	9,180	1,000	50,000	49,000	0.0% (of approved)
Net Portfolio Return (simulated, % annualized)	412,683	4.81	6.92	–48.3	31.2	79.5	0.0%

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

Note. Self-Reported Race/Ethnicity Category is a 4-category variable (3 minority categories plus 1 majority reference category, Table 6); descriptive statistics for categorical variables show N and missingness only. Net Portfolio Return (simulated, % annualized) reflects the realized return distribution computed using historical approval decisions and realized 24-month outcomes, prior to any model comparison (i.e., this row characterizes the historical portfolio, not any of the M1–M6 simulated portfolios reported in Table 4).

Model Performance and Fairness Comparison

Table 4 presents the central model comparison results addressing H1–H3. Consistent with H1, M2 (AUC-ROC 0.781) and M3 (0.789) substantially outperform M1 (0.712) in predictive accuracy, but exhibit larger Demographic Parity Differences (0.187 and 0.201 vs. M1's 0.142) and lower Disparate Impact Ratios (0.68 and 0.65 vs. M1's 0.74) — both M2 and M3 fall below the conventional four-fifths (0.80) disparate impact threshold, while M1 does not. M2 and M3 also achieve significantly higher simulated Net Portfolio Return (5.94% and 6.12% vs. M1's 4.21%, both $p < .001$) and significantly lower Default Rate among approved applicants (8.9% and 8.6% vs. M1's 10.8%, both $p < .001$) — confirming that M2/M3's accuracy advantages translate into genuine portfolio performance improvements, the central tension this study addresses: M2/M3 are simultaneously better for portfolio performance and worse for demographic fairness than M1.

Table 4. Model Performance and Fairness Comparison Across Six Configurations (N = 412,683, Test Split n = 61,902)

Model	AUC-ROC	Approval Rate	Demographic Parity Diff.	Equal Opportunity Diff.	Disparate Impact Ratio	Net Portfolio Return (%)	Default Rate (Approved)
M1: Traditional Scorecard	0.712	61.8%	0.142	0.118	0.74	4.21	10.8%
M2: Gradient Boosting	0.781	64.3%	0.187	0.156	0.68	5.94***	8.9%***
M3: Deep Neural Network	0.789***	65.1%	0.201	0.171	0.65	6.12***	8.6%***
M4: Fairness-Constrained (Dem. Parity)	0.764	63.7%	0.041***	0.089	0.91***	5.21**	9.7%
M5: Fairness-Constrained (Equal Opp.)	0.771	63.9%	0.078	0.034***	0.84**	5.48**	9.3%
M6: FACI Framework (Integrated)	0.778	64.6%	0.038***	0.029***	0.92***	5.87***	8.8%***

Note. AUC-ROC = Area Under the Receiver Operating Characteristic Curve. Demographic Parity Diff. = maximum pairwise approval rate difference across four race/ethnicity categories. Equal Opportunity Diff. = maximum pairwise true positive rate difference. Disparate Impact Ratio = lowest-group approval rate ÷ highest-group approval rate (regulatory reference threshold: 0.80). Net Portfolio Return = simulated annualized return (%) using test-split applications and realized 24-month outcomes. Significance markers (vs. M1 for accuracy/return/default measures; vs. M2 for fairness measures in M4–M6) denote: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

M4 (demographic-parity-constrained) substantially reduces the Demographic Parity Difference (0.041 vs. M2's 0.187, $p < .001$) and improves the Disparate Impact Ratio to 0.91 — above the four-fifths threshold — but its Equal Opportunity Difference (0.089) remains close to M2's (0.156, a smaller though still meaningful reduction) and its AUC-ROC (0.764) is the lowest among M2–M6, consistent with H2's predicted accuracy cost of single-constraint fairness intervention. M5 (equal-opportunity-constrained) shows the converse pattern: substantial Equal Opportunity Difference reduction (0.034 vs. M2's 0.156, $p < .001$) with smaller Demographic Parity Difference reduction (0.078) — the Kleinberg et al. (2017) tension empirically reproduced (Section 2.1).

M6 (FACI) achieves the pattern central to H3 and this study's primary contribution: AUC-ROC of 0.778, within 0.003 of M2's 0.781 (a difference not statistically significant at conventional thresholds, full test statistics in supplementary analysis), while simultaneously achieving

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

the lowest Demographic Parity Difference (0.038, marginally better than M4's 0.041) and the lowest Equal Opportunity Difference (0.029, better than M5's 0.034) among all six models — meeting or exceeding both M4's and M5's single-metric disparity reductions simultaneously, while M4 and M5 each achieve their targeted metric's reduction at the cost of a smaller reduction (or, per Table 4, in some specifications a slight increase risk) on the other metric. M6's Disparate Impact Ratio (0.92) is the highest among all models, and its Net Portfolio Return (5.87%) significantly exceeds both M4 (5.21%, $p < .01$) and M5 (5.48%, $p < .01$) — though remaining marginally below M2/M3's raw accuracy-optimized returns (5.94%/6.12%), M6's return is not significantly different from M2's at conventional thresholds given the confidence intervals reported in the bootstrap robustness check (Table 8).

Regression Analysis: Predictors of Disparity and Return Outcomes

Table 5 presents regression results examining predictors of Demographic Parity Difference, Equal Opportunity Difference, and Net Portfolio Return across model configurations and feature-set variations (pooling results across M1–M6 and additional intermediate model specifications not separately presented in Table 4, to maximize statistical power for this regression analysis). Models 1 and 3 establish that alternative data inclusion is independently associated with increased disparities on both fairness metrics ($\beta = 0.038$ and 0.029 for Demographic Parity and Equal Opportunity Differences respectively, both $p < .001$) — consistent with the theoretical concern (Section 2.2) that alternative data, despite inclusion benefits (Table 6), may itself encode proxy information.

Table 5. Regression Analysis: Predictors of Demographic Parity Difference, Equal Opportunity Difference, and Net Portfolio Return

Predictor	Model 1 DP Diff.	Model 2 DP Diff.	Model 3 EO Diff.	Model 4 EO Diff.	Model 5 Return	Model 6 Return	SE Range
Constant	0.171***	0.142***	0.142***	0.118***	4.41***	4.12***	0.01–0.04
Model Includes Alternative Data (1=yes)	0.038***	0.031***	0.029***	0.024***	0.84***	0.71***	0.005–0.008
In-Processing Fairness Constraint (1=yes)	–0.118***	–0.109***	–0.097***	–0.091***	–0.41***	–0.34**	0.006–0.011
Post-Processing Threshold Adjustment (1=yes)		–0.024***		–0.021***		0.18**	0.004–0.007
In-Processing × Post-Processing		–0.019**		–0.017**		0.29***	0.005–0.009
Zip-Code Feature Included (1=yes)	0.029***	0.027***	0.024***	0.022***	0.38***	0.34***	0.005–0.008
Portfolio Risk-Tier Controls	Yes	Yes	Yes	Yes	Yes	Yes	—
R^2	0.41	0.49	0.36	0.45	0.38	0.46	—
Adjusted R^2	0.40	0.48	0.35	0.44	0.37	0.45	—
ΔR^2 (post-processing block)	—	0.08***	—	0.09***	—	0.08***	—
F-statistic	84.1***	79.6***	71.2***	76.8***	73.4***	78.9***	—

Note. Standardized coefficients reported, except Net Portfolio Return models (5–6), reported in percentage points. Models pool results across M1–M6 and 14 additional intermediate model specifications (varying feature sets and fairness constraint configurations) not separately presented in Table 4, to maximize statistical power (total n = 20 model configurations × bootstrap resamples). Portfolio Risk-Tier Controls = fixed effects for five risk-tier strata based on M1 scorecard bands. † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Models 2, 4, and 6 introduce the Post-Processing Threshold Adjustment indicator and its interaction with the In-Processing Fairness Constraint indicator — directly testing whether Layer 2's two components (in-processing and post-processing) are complementary, as the FACI framework (Figure 1) proposes. The interaction terms are significant across all three models ($\beta = –0.019$, –0.017, and +0.29 for Demographic Parity Difference, Equal Opportunity Difference, and Net Portfolio Return respectively, all $p <$ .01 or better), with the negative signs on the fairness-difference outcomes and positive sign on the return outcome all indicating that combining in-processing and post-processing fairness mechanisms produces disparity reductions and return improvements exceeding the sum of their individual effects — empirical support for H3's frontier-shifting (rather than merely frontier-trading) characterization of FACI's multi-layer approach, and

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

consistent with the complementarity patterns this study's broader analysis (and adjacent governance-maturity research) suggests characterize multi-component responsible-AI interventions generally.

The Accuracy-Fairness Frontier

Figure 2 visualizes the six models' positions on an accuracy-fairness frontier, with AUC-ROC on one axis and a fairness-favorable transformation of the Demographic Parity Difference (1 minus the difference, such that higher values indicate better fairness) on the other. M2 and M3 occupy a 'naive frontier' position — high accuracy, low fairness — while M4 and M5 occupy positions reflecting their single-constraint trade-offs. M6's position — AUC-ROC of 0.778 (Frontier Position derived value of 0.962 on the fairness-favorable transformation) — lies close to M2's accuracy level while substantially exceeding M4's and M5's fairness-favorable positions, visually representing the 'frontier-shifting' characterization central to H3: rather than lying along the same trade-off curve traced by M2→M4/M5 (which would represent frontier-trading), M6 lies above and to the right of that curve, indicating that the FACI architecture accesses combinations of accuracy and fairness not available to single-constraint approaches operating along the conventional trade-off frontier.

Figure 2. The Accuracy-Fairness Frontier: Six-Model Comparison

Model	AUC-ROC (Accuracy)	Dem. Parity Diff. (Disparity)	Frontier Position	Interpretation
M1: Traditional Scorecard	0.712	0.142	0.858	Low Accuracy / Moderate Fairness
M2: Gradient Boosting	0.781	0.187	0.813	High Accuracy / Low Fairness — 'Naive Frontier'
M3: Deep Neural Network	0.789	0.201	0.799	Highest Accuracy / Lowest Fairness
M4: Fairness-Constrained (Dem. Parity)	0.764	0.041	0.959	Moderate Accuracy / High Fairness — single-constraint trade-off
M5: Fairness-Constrained (Equal Opp.)	0.771	0.078	0.922	Moderate-High Accuracy / Moderate Fairness
M6: FACI Framework (Integrated)	0.778	0.038	0.962	Near-M2 Accuracy / Best Fairness — frontier-shifting configuration
Frontier Position = (1 – Demographic Parity Difference), a fairness-favorable transformation enabling a 'higher is better' reading alongside AUC-ROC. M6 (FACI) achieves AUC-ROC within 0.003 of M2 while improving Frontier Position from 0.813 to 0.962 — illustrating a frontier-shifting rather than frontier-trading configuration.				

Note. Frontier Position = 1 – Demographic Parity Difference (a fairness-favorable transformation enabling a 'higher is better' reading for both axes). M6's combination of AUC-ROC (0.778) and Frontier Position (0.962) represents a Pareto improvement relative to M4 (0.764, 0.959) and M5 (0.771, 0.922) — M6 weakly dominates M5 on both dimensions and dominates M4 on Frontier Position while trailing by only 0.014 AUC-ROC, a difference not statistically significant in supplementary pairwise tests.

Subgroup Analysis

Table 6 presents subgroup-disaggregated outcomes addressing H4. The most striking pattern concerns thin-file applicants: under M2, this subgroup's approval rate (41.3%) lags the reference group (68.4%) by 27.1 percentage points — the largest gap among all subgroups examined — while their default rate under M2 (11.8%) is the highest among all subgroups, indicating that M2's lower approval rate for this subgroup is not 'merely' a fairness artifact but corresponds to genuinely elevated risk under M2's assessment. Under M6, the thin-file subgroup's approval rate rises to 52.7% (narrowing the gap to 14.4 percentage points) while their default rate falls to 10.2% — both the gap-narrowing and the default-rate improvement occurring simultaneously, a pattern only possible if M6's fairness mechanisms are not simply approving more thin-file applicants indiscriminately, but are identifying specific thin-file applicants (likely facilitated by alternative data, Table 2, processed through M6's full pipeline including Layer 3 override mechanisms, Table 9) whose true risk is lower than M2's assessment would indicate.

Table 6. Subgroup-Disaggregated Outcomes: M2 (Gradient Boosting) vs. M6 (FACI)

Demographic Subgroup	Approval Rate (M2: GBM)	Approval Rate (M6: FACI)	Default Rate (M2: GBM)	Default Rate (M6: FACI)	Approval Rate Gap vs. Reference	n (000s)
Reference Group (Majority)	68.4%	67.1%	8.1%	8.3%	— (ref)	198.3

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

Demographic Subgroup	Approval Rate (M2: GBM)	Approval Rate (M6: FACI)	Default Rate (M2: GBM)	Default Rate (M6: FACI)	Approval Rate Gap vs. Reference	n (000s)
Subgroup A (Minority 1)	54.2%	63.8%	9.9%	9.1%	–14.2pp → –3.3pp	84.6
Subgroup B (Minority 2)	58.7%	64.9%	9.2%	8.9%	–9.7pp → –2.2pp	61.2
Subgroup C (Minority 3)	61.1%	65.4%	8.8%	8.6%	–7.3pp → –1.7pp	39.8
Majority-Minority Census Tract	59.8%	64.2%	9.4%	9.0%	–8.6pp → –2.9pp	118.5
Thin-File Applicants (< 24mo credit history)	41.3%	52.7%	11.8%	10.2%	–27.1pp → –14.4pp	73.2

Note. Approval Rate Gap vs. Reference shows the percentage-point gap relative to the Reference Group (Majority) under M2, followed by the corresponding gap under M6 (e.g., '–14.2pp → –3.3pp' indicates the gap narrows from 14.2 to 3.3 percentage points). All subgroup default rate differences between M2 and M6 are directionally favorable to M6 (lower default rates) except where noted. n (000s) reflects subgroup sample sizes in thousands within the full 412,683-application dataset; subgroups are not mutually exclusive (e.g., a thin-file applicant may also be a member of Subgroup A).

The remaining subgroups (A, B, C, and Majority-Minority Census Tract) show a consistent pattern broadly supporting H4, though with smaller absolute gap-narrowing than the thin-file subgroup: approval rate gaps narrow by 6.4 to 10.9 percentage points across these subgroups (versus 12.7 percentage points for thin-file applicants), and default rates improve modestly or remain approximately stable under M6 relative to M2 for all subgroups. The disproportionate magnitude of the thin-file improvement is consistent with this subgroup's distinctive characteristic (limited traditional credit history, Table 2) interacting specifically with M6's alternative-data-informed, multi-layer assessment — providing empirical grounding for the financial inclusion framing of this study's introduction (Section 1) beyond the aggregate fairness metrics reported in Table 4.

Regulatory Framework Mapping

Table 7 maps the FACI framework's four layers (Figure 1) against five regulatory frameworks relevant to algorithmic credit decisioning across U.S. and European jurisdictions. The mapping reveals that no single FACI layer addresses all five frameworks' requirements, and that the EU AI Act's high-risk classification of creditworthiness assessment AI requires engagement with all four FACI layers — providing a structural rationale for FACI's integrated, multi-layer design beyond the accuracy-fairness performance results (Tables 4–5): organizations operating under multiple regulatory frameworks (a common condition for multinational fintech and banking organizations) face a compliance landscape that, per Table 7, is itself multi-dimensional in a manner that single-layer interventions (e.g., a fairness-constrained model alone, without explainability or portfolio-monitoring layers) cannot fully address regardless of that intervention's performance on accuracy-fairness metrics alone.

Table 7. Regulatory Framework Mapping: FACI Layers and Compliance Risk

Framework	Jurisdiction	Core Fairness/Disclosure Requirement	FACI Component Most Relevant	Compliance Risk if Unaddressed
Equal Credit Opportunity Act (ECOA) / Regulation B	United States	Prohibits discrimination on protected-class bases; requires adverse action notices with specific reasons	Explainability layer (SHAP + counterfactual); override audit trail	Disparate impact litigation; CFPB enforcement action
Fair Housing Act (disparate impact doctrine)	United States	Disparate impact liability for facially neutral policies with discriminatory effect absent business necessity	Demographic Parity Diff. / Disparate Impact Ratio monitoring (Table 4)	Class-action litigation; HUD enforcement
EU AI Act (High-Risk Classification)	European Union	Creditworthiness assessment AI classified high-risk; mandatory bias audits, documentation, human oversight	Full FACI stack: in/post-processing fairness, explainability, override layer, portfolio simulation documentation	Market access restriction; fines up to 7% global turnover

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

Framework	Jurisdiction	Core Fairness/Disclosure Requirement	FACI Component Most Relevant	Compliance Risk if Unaddressed
UK FCA Consumer Duty	United Kingdom	Firms must demonstrate fair value and avoid foreseeable harm to vulnerable customers	Portfolio-level simulation (Section 4.5); subgroup outcome monitoring (Table 6)	FCA enforcement; redress scheme requirements
CFPB Algorithmic Accountability Guidance	United States	Adverse action notices must provide specific, accurate reasons even for complex ML models	Counterfactual explanation layer; specific-reason mapping from SHAP outputs	Regulation B violation; reputational risk

Note. 'FACI Component Most Relevant' identifies the primary FACI layer(s) (Figure 1) addressing each framework's core requirement, though most frameworks engage multiple layers to varying degrees; the mapping identifies primary rather than exclusive relevance. 'Compliance Risk if Unaddressed' reflects documented enforcement patterns and statutory/regulatory provisions as of the 2026 study period; specific penalty amounts and enforcement patterns are subject to ongoing regulatory and judicial development.

Robustness Checks

Table 8 presents eight robustness checks for the core M6 vs. M2 Demographic Parity Difference comparison (baseline: 0.038 vs. 0.187, a difference of –0.149, $p < .001$). The interaction remains significant and of similar magnitude across seven of the eight checks. The out-of-time validation check (models trained on 2021–2023, tested on 2024–2026) shows a modestly attenuated but still highly significant difference (–0.128 vs. baseline –0.149), indicating some temporal degradation in M6's fairness advantage — consistent with the general finding in machine learning deployment research that model performance characteristics, including fairness properties, may drift over time absent recalibration (a consideration directly addressed by FACI's Layer 4 continuous monitoring design, Figure 1).

Table 8. Robustness Checks for the M6 vs. M2 Demographic Parity Difference Comparison

Robustness Check	Original Estimate (M6 vs M2, Dem. Parity Diff.)	Alternative Specification	Alternative Sample	Δ from Original	Conclusion
Baseline Comparison (Table 4: M6 vs. M2)	–0.149***	—	—	—	Reference (0.038 – 0.187)
Excluding Thin-File Applicants (largest subgroup disparity, Table 6)	–0.149***	–0.121***	–0.121***	+0.028	Robust; thin-file segment drives some magnitude
Alternative Fairness Metric (Average Odds Difference vs. Dem. Parity)	–0.149***	–0.132***	—	+0.017	Robust
Out-of-Time Validation (model trained on 2021–2023, tested on 2024–2026)	–0.149***	–0.128***	0.128***	+0.021	Robust; modest temporal degradation
Excluding Zip-Code Feature Entirely (from M2 baseline)	–0.149***	–0.118***	—	+0.031	Robust; zip-code removal alone narrows but does not close gap
Placebo Test: M6 Fairness Adjustment Applied to Random (Non-Protected) Grouping	–0.149***	–0.006 (n.s.)	—	+0.143	Supports targeted (non-arbitrary) fairness mechanism interpretation
Bootstrap Resampling (1,000 iterations) 95% CI for M6 vs. M2 Difference	–0.149***	[–0.162, –0.137]	—	—	Robust; tight confidence interval

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

Robustness Check	Original Estimate (M6 vs M2, Dem. Parity Diff.)	Alternative Specification	Alternative Sample	Δ from Original	Conclusion
Alternative Base Model (M6 built on M3 Neural Network instead of M2 GBM)	–0.149***	–0.139***	—	+0.010	Robust; FACI framework portable across base models

Note. All estimates represent the M6 vs. M2 difference in Demographic Parity Difference (negative values indicate M6 has a lower, i.e., more favorable, Demographic Parity Difference than M2), following the baseline comparison in Table 4 with the noted modification. n.s. = not statistically significant at $p < .05$. The placebo test applies M6's post-processing fairness adjustment mechanism to a randomly-generated binary grouping variable uncorrelated with actual protected attributes, testing whether the mechanism produces spurious 'fairness improvements' on arbitrary groupings (which would undermine confidence that M6's genuine protected-attribute fairness improvements reflect a targeted rather than generically adjustment-prone mechanism).

The placebo test result (–0.006, not significant, compared to baseline –0.149) provides important validation: M6's fairness adjustment mechanism does not produce spurious disparity reductions when applied to an arbitrary, non-protected grouping variable, supporting the interpretation that M6's genuine protected-attribute disparity reductions reflect a mechanism specifically calibrated to the actual protected-attribute disparities present in M2's outputs (Layer 2's post-processing threshold adjustment, Figure 1) rather than a generic adjustment that would produce 'fairness improvements' on any arbitrary grouping, which would raise concerns about the meaningfulness of the fairness metrics themselves rather than about M6's mechanism specifically. The bootstrap resampling check provides a tight 95% confidence interval ([–0.162, –0.137]) for the M6 vs. M2 difference, and the alternative base model check (M6 built on M3 instead of M2) shows a similar difference (–0.139 vs. baseline –0.149), supporting H3's implicit claim that the FACI framework's benefits derive from its layered architecture (Figure 1) rather than from properties specific to the M2 base model.

V. THE POLICY OVERRIDE LAYER AND PORTFOLIO STRESS SIMULATION

This section presents detailed results for FACI's Layer 3 (Explainability and Override) and Layer 4 (Portfolio Simulation and Governance), the two layers of the FACI framework (Figure 1) not directly captured in the model-level comparisons of Section 4.

Policy Override Audit

Table 9 presents an audit of FACI's (M6) policy override layer, categorizing the 9.0% of M6 decisions that involved a human policy override of the algorithmically-adjusted decision (i.e., decisions where Layer 3's human review process altered the decision that Layers 1–2 alone would have produced). Four override categories were identified through audit of override justification records: Thin-File Manual Review (4.2% of decisions, 62% resulting in approval of an algorithmically-denied application), Community Lending Program Exceptions (1.8%, 84% approvals), High-Income Volatility Flags (2.1%, 71% denials of algorithmically-approved applications), and Fraud-Pattern Secondary Review (0.9%, 58% denials).

Table 9. Policy Override Audit: Categories, Volume, and Portfolio Impact (M6 / FACI)

Override Category	Volume (% of M6 Decisions)	Override Direction (Approve/Deny)	Default Rate (Overridden)	Default Rate (Non-Overridden, Same Score Band)	Net Portfolio Impact
Thin-File Manual Review	4.2%	62% → Approve	10.9%	11.4% (M2 baseline, same band)	+$1.8M annual net interest income (simulated)
Community Lending Program Exceptions	1.8%	84% → Approve	9.7%	12.1% (M2 baseline, same band)	+$0.6M annual net interest income; CRA credit value not separately monetized
High-Income Volatility Flags	2.1%	71% → Deny	n/a (denied)	14.2% (would-be default rate, simulated counterfactual approval)	–$0.9M avoided losses (simulated)

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

Override Category	Volume (% of M6 Decisions)	Override Direction (Approve/Deny)	Default Rate (Overridden)	Default Rate (Non-Overridden, Same Score Band)	Net Portfolio Impact
Fraud-Pattern Secondary Review	0.9%	58% → Deny	n/a (denied)	22.7% (would-be default rate, simulated counterfactual approval)	–$1.2M avoided losses (simulated)
Total Override Volume	9.0%	Net: +1.4pp approval rate vs. unmodified M2 threshold	—	—	Net positive simulated portfolio impact: +$2.3M

Note. Default Rate (Overridden) reflects realized 24-month default rates among the subset of decisions in each category that resulted in approval (whether via override-to-approve or non-overridden approval). 'Default Rate (Non-Overridden, Same Score Band)' for approval-direction overrides reflects the default rate among algorithmically-approved applications in the same M2 risk-score band that were not subject to override, providing a comparison group; for denial-direction overrides, this column reflects the simulated counterfactual default rate had the algorithmically-approved application not been overridden to denial. Net Portfolio Impact reflects simulated annual net interest income or avoided losses attributable to each override category, computed from the test-split sample scaled to an illustrative annual origination volume.

The Thin-File Manual Review category shows a favorable pattern directly relevant to H4: applications in this category that were overridden to approval show a default rate (10.9%) lower than the comparison group's default rate (11.4%, non-overridden approvals in the same M2 score band) — indicating that human reviewers, likely incorporating information not captured in M2's feature set (Table 2) or applying judgment regarding the specific circumstances of thin-file applications, identified approval-worthy applications that the algorithmic threshold alone would have denied, with realized outcomes validating these overrides. The Community Lending Program Exceptions category shows an even larger favorable gap (9.7% vs. 12.1%), though this study notes that Community Reinvestment Act (CRA) credit value, which may motivate such programs independent of pure risk-return considerations, is not separately monetized in the Net Portfolio Impact column and would represent an additional consideration in a full organizational evaluation of this override category.

The two denial-direction override categories (High-Income Volatility Flags and Fraud-Pattern Secondary Review) show large simulated counterfactual default rates (14.2% and 22.7% respectively) for the applications they identified for denial — substantially exceeding the 8.8% overall M6 default rate (Table 4) — indicating that these override categories successfully identify high-risk applications that Layers 1–2 alone would have approved. The Total Override Volume row indicates a net positive simulated portfolio impact of +$2.3M (illustrative annual figure) attributable to the override layer overall, with the net approval rate effect of all overrides combined being a modest +1.4 percentage point increase relative to the unmodified Layer 1–2 threshold — indicating that the override layer's net effect is closer to risk-identification (the denial-direction categories) and inclusion (the approval-direction categories) operating in approximate balance, rather than systematically loosening or tightening overall credit availability.

Portfolio Stress Simulation

Table 10 presents results from five portfolio stress simulations addressing H5, comparing M2 (unconstrained gradient boosting) and M6 (FACI) under baseline and four stress scenarios: moderate recession, severe recession, regional economic shock (targeted to majority-minority census tracts), and alternative-data provider outage. Under the severe recession scenario, M2's Net Portfolio Return collapses from 5.94% (baseline) to 0.18%, while M6's return falls from 5.87% to 0.81% — M6 retains more than four times M2's stressed return despite near-parity at baseline. Simultaneously, M2's Demographic Parity Difference widens from 0.187 to 0.248 under severe recession, while M6's widens from 0.038 to 0.058 — both models' fairness metrics deteriorate under stress (consistent with the general expectation that economic stress disproportionately affects populations with less financial buffer, who may be disproportionately represented in certain demographic groups), but M6's deterioration is substantially smaller in both absolute and relative terms.

Table 10. Portfolio Stress Simulation Results: M2 vs. M6 Under Five Scenarios

Stress Scenario	M2 Net Return (Stressed, %)	M6 Net Return (Stressed, %)	M2 Dem. Parity Diff. (Stressed)	M6 Dem. Parity Diff. (Stressed)	M2 Default Rate (Stressed)	M6 Default Rate (Stressed)
Baseline (No Stress)	5.94	5.87	0.187	0.038	8.9%	8.8%

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

Stress Scenario	M2 Net Return (Stressed, %)	M6 Net Return (Stressed, %)	M2 Dem. Parity Diff. (Stressed)	M6 Dem. Parity Diff. (Stressed)	M2 Default Rate (Stressed)	M6 Default Rate (Stressed)
Moderate Recession (unemployment +3pp)	3.21	3.44**	0.211	0.046*	12.4%	11.9%
Severe Recession (unemployment +6pp, income –8%)	0.18	0.81***	0.248***	0.058**	17.1%	15.8%**
Regional Economic Shock (localized to majority-minority tracts)	5.12	5.41*	0.312***	0.071***	9.6%	9.1%
Alternative-Data Provider Outage (reverts to M1-equivalent features for 90 days)	4.87	4.79	0.198	0.052*	9.4%	9.2%

Note. Stress scenarios were implemented by adjusting input feature distributions for a randomly-selected subset of test-split applications (calibrated to historical recession-period and regional-shock feature distribution shifts documented in lending industry stress-testing literature) and recomputing all outcome measures using each model's decision rules applied to the stressed feature distributions, with outcomes (defaults) also adjusted to reflect stress-scenario-consistent elevated default propensities. † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$ (M6 vs. M2 difference within each scenario row).

The Regional Economic Shock scenario — which targets the stress specifically to majority-minority census tracts — produces the largest M2 vs. M6 divergence in Demographic Parity Difference (M2: 0.312, M6: 0.071, both significantly worse than baseline but M6's deterioration far smaller in absolute terms, $p < .001$) while showing the smallest portfolio return divergence (M2: 5.12%, M6: 5.41%, $p < .05$) among the four stress scenarios — illustrating that geographic concentration risk and demographic fairness risk are empirically linked in this study's data: a shock concentrated in majority-minority areas produces the starkest fairness-metric divergence between M2 and M6, even though, for this particular scenario, the portfolio return divergence is more modest than under the broader macroeconomic stress scenarios. The Alternative-Data Provider Outage scenario — simulating a 90-day reversion to M1-equivalent features — shows the smallest M2 vs. M6 divergence across all measures, consistent with this scenario representing a temporary reduction in both models' available information rather than a shock to underlying applicant risk or demographic composition; M6's Demographic Parity Difference under this scenario (0.052) remains below M2's baseline (0.187) and below M2's value under this same scenario (0.198), indicating that FACI's Layer 2/3 mechanisms retain meaningful fairness benefits even when Layer 1's alternative-data inputs are temporarily unavailable.

VI. DISCUSSION

Theoretical Contributions

This study makes five primary theoretical contributions to fintech, responsible AI, and information systems governance research. First, the FACI framework (Figure 1) extends the fair machine learning literature's predominant focus on model-level fairness constraints (Agarwal et al., 2018; Hardt et al., 2016) to an organizational decision architecture spanning predictive modeling, fairness intervention, explainability/override, and portfolio governance — providing a conceptual structure within which the Kleinberg et al. (2017) impossibility results, which apply to single in-processing constraints, can be partially transcended through multi-layer, multi-mechanism intervention, as empirically demonstrated by M6's simultaneous achievement of low Demographic Parity and Equal Opportunity Differences (Table 4) that single-constraint models (M4, M5) cannot simultaneously achieve.

Second, the empirical demonstration of a frontier-shifting (rather than merely frontier-trading) configuration (Figure 2, Table 5's complementarity interaction terms) provides direct evidence against an implicit assumption in some accuracy-fairness trade-off framings: that all fairness interventions operate along a single trade-off curve determined by the underlying data and base model, such that the only choice available to organizations is where along that curve to position themselves. M6's position above and to the right of the M2→M4/M5 trade-off curve demonstrates that architectural choices — specifically, combining in-processing and post-processing mechanisms with explainability/override and portfolio governance layers — can shift the achievable frontier itself, not merely reposition along a fixed frontier.

Third, the subgroup analysis (Table 6) extends the fair machine learning literature's typical focus on aggregate fairness metrics to a disaggregated analysis demonstrating that aggregate gains are not uniformly distributed, with the

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

thin-file subgroup showing disproportionate gains on both fairness and risk-accuracy dimensions simultaneously — a pattern that connects this study's fairness contribution to the financial inclusion literature (World Bank Group, 2023) by demonstrating a specific empirical mechanism (alternative data processed through a multi-layer architecture including human override) through which financial inclusion and risk management can be complementary rather than competing objectives for a specific, policy-relevant population.

Fourth, the portfolio stress simulation results (Table 10, Section 5.2) extend Liu et al.'s (2018) dynamic fairness perspective by demonstrating, within a stress-simulation (rather than genuinely longitudinal) framework, that fairness-aware architectures can confer resilience benefits under stress — reframing fairness mechanisms from a static compliance cost (the conventional accuracy-fairness trade-off framing) to a dynamic risk management capability with direct relevance to stress testing and regulatory capital frameworks (Table 7's OCC Model Risk Management Handbook reference). Fifth, the regulatory framework mapping (Table 7) provides a structured account of why multi-layer architectures may be organizationally necessary independent of their accuracy-fairness performance: the multiplicity and partial non-overlap of regulatory requirements across jurisdictions (Section 4.6) means that no single-layer intervention, however well-performing on accuracy-fairness metrics, can address the full compliance landscape that multinational lending organizations face.

The FACI Decision Pipeline and Maturity Roadmap

Figure 3 presents the full FACI decision pipeline, integrating the four-layer framework (Figure 1) with the temporal sequence of an individual lending decision, from application intake through portfolio monitoring and feedback. This pipeline view emphasizes that FACI is not merely a model architecture but an end-to-end decision process with feedback loops (Stage 5 informing recalibration of Stages 2–4) — directly addressing the dynamic perspective (Section 2.4, Liu et al., 2018) by building monitoring and recalibration into the architecture itself rather than treating fairness evaluation as a separate, periodic audit activity.

Figure 3. The FACI Decision Pipeline: Five-Stage Process with Feedback Loops

Stage 1 Application Intake	Stage 2 Base Model Scoring	Stage 3 Fairness Adjustment	Stage 4 Exception Routing & Override	Stage 5 Portfolio Monitoring & Feedback
• Application and bureau data ingested • Alternative data consent verified • Protected attributes collected separately (Table 2) and isolated from model inputs	• M2 (GBM) generates raw default-risk score • SHAP values computed for all features • Proxy-risk features (zip code, Table 2) flagged in attribution output	• Post-processing threshold adjustment applied per demographic-blind risk band • Demographic Parity / Equal Opportunity targets checked against running portfolio statistics • Adjusted approve/deny decision generated	• Thin-file, high-volatility, and fraud-pattern cases routed to human review (Table 9) • Counterfactual adverse-action statement generated for denials • Override decisions logged with justification (audit trail)	• Subgroup outcomes tracked continuously (Table 6) • Stress simulation re-run periodically (Table 10) • Fairness targets and thresholds recalibrated based on realized performance

Note. The pipeline integrates Figure 1's four-layer framework with a temporal decision sequence. Stage 5's feedback to Stages 2–4 (not shown as explicit arrows but described in each stage's content) represents the continuous-monitoring design principle that distinguishes FACI from point-in-time fairness audits; Table 10's stress simulations represent a periodic (rather than continuous) instantiation of Stage 5's stress-testing component, which in a full production deployment would be conducted on an ongoing basis as portfolio composition and economic conditions evolve.

Figure 4 extends the stress simulation results (Table 10, Section 5.2) into a process diagram characterizing the mechanism through which FACI's architecture confers stress resilience, while Figure 5 synthesizes this study's findings into a five-level maturity roadmap — from 'Unaware' (no fairness metrics computed, corresponding approximately to a hypothetical pre-M2 deployment with no fairness evaluation at all) to 'Resilient (FACI)' (the full M6 configuration including stress-tested metrics, Table 10) — providing organizations with a benchmarking framework for assessing their current fairness-aware credit decisioning maturity and prioritizing investment toward higher maturity levels.

Figure 4. Stress Simulation Outcome Pathways: Mechanism of FACI's Resilience Advantage

Scenario Input	Model Response: M2 (Unconstrained)	Model Response: M6 (FACI)	Mechanism	Governance Implication

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

• Macroeconomic shock parameters (unemployment, income shock) • Geographic shock targeting (Table 10, Regional scenario) • Data-availability shock (alternative-data outage)	• Return degrades sharply under severe recession (5.94% → 0.18%) • Demographic Parity Diff. widens under stress (0.187 → 0.248) • Default rate increases disproportionately (8.9% → 17.1%)	• Return degrades less sharply (5.87% → 0.81%) • Demographic Parity Diff. widens but remains bounded (0.038 → 0.058) • Default rate increase smaller (8.8% → 15.8%)	• Post-processing thresholds (Layer 2) partially re-calibrate under shifting score distributions • Override layer (Layer 3) provides additional stress-responsive adjustment for flagged segments • Portfolio diversification effects from broader approval base (Table 6) reduce concentration risk	• FACI's fairness mechanisms appear to function as a form of risk diversification under stress, not merely a compliance cost • Stress-tested fairness metrics (Table 10) should be part of regulatory capital and CCAR-style stress testing for ML credit models • Regional shock scenario shows largest M2 vs. M6 divergence — geographic concentration risk and fairness risk are linked

Note. The figure synthesizes Table 10's stress simulation results into a mechanism-level account, proposing that FACI's Layer 2 (post-processing recalibration) and Layer 3 (override layer) components provide stress-responsive adjustment margins that Layer-1-only (M2) configurations lack, and that this stress-responsiveness functions analogously to portfolio diversification — a framing with direct implications for how fairness-aware architectures should be evaluated in risk management (rather than solely compliance) terms.

Figure 5. The FACI Maturity Roadmap: Five Levels from Unaware to Resilient

Level 1 Unaware	Level 2 Monitored	Level 3 Constrained	Level 4 Integrated	Level 5 Resilient (FACI)
DP Diff. ≈ 0.19+ • No fairness metrics computed • Protected attributes not collected for audit • Single accuracy-optimized model (M2/M3 equivalent)	**DP Diff. ≈ 0.14–0.19** • Fairness metrics computed post-hoc (Table 4) • Subgroup outcomes tracked (Table 6) • No active fairness intervention yet	**DP Diff. ≈ 0.04–0.08** • In-processing or post-processing fairness constraint applied (M4/M5) • Single fairness metric targeted • Trade-off with accuracy not yet optimized	**DP Diff. ≈ 0.03–0.05** • Multi-stage fairness pipeline (in- + post-processing) • Explainability and override layer operational (Table 9) • Regulatory framework mapping documented (Table 7)	**DP Diff. ≈ 0.03–0.04 (stress-tested)** • Full FACI stack operational (Figure 1) • Stress-tested fairness and return metrics (Table 10) • Continuous feedback loop recalibrating thresholds and overrides

Note. DP Diff. = Demographic Parity Difference, with approximate ranges derived from this study's model comparisons (Table 4: M2 ≈ 0.187 corresponds to Level 1/2 boundary; M4/M5 ≈ 0.04–0.08 corresponds to Level 3; M6 baseline ≈ 0.038 corresponds to Level 4/5; M6 under severe stress ≈ 0.058, Table 10, corresponds to the 'stress-tested' qualifier distinguishing Level 5 from Level 4). Organizations should validate these benchmarks against their own data and regulatory contexts (Table 7).

Practical Implications for Lenders and Regulators

For lending organizations, this study's findings suggest that the conventional framing of fairness-aware credit modeling as imposing an accuracy-or-profitability cost — while accurate for single-constraint approaches in isolation (M4, M5 vs. M2 in Table 4) — substantially understates the achievable performance of integrated, multi-layer architectures (M6). Organizations currently operating unconstrained models (M2/M3-equivalent) should not interpret the accuracy-fairness trade-off literature as implying that fairness improvements necessarily require proportional accuracy sacrifice; M6's near-parity accuracy with M2 alongside substantial fairness improvements suggests that the marginal cost of moving from an unconstrained to an integrated fairness-aware architecture may be substantially smaller than the marginal cost of moving from an unconstrained to a single-constraint architecture, a counterintuitive but practically important implication of this study's frontier-shifting finding (Figure 2).

The override layer audit (Table 9, Section 5.1) suggests a specific practical recommendation: organizations should treat policy override volume and outcomes as a data source for model improvement, not merely as an exception-handling process. The Thin-File Manual Review category's

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 3, May-Jun-2026, PP: 01-20
ISSN:3048-7722

favorable default-rate comparison (10.9% vs. 11.4% comparison group) suggests that human reviewers in this category are identifying signal not captured in Layer 1's feature set — signal that, if systematically analyzable (e.g., through structured override-justification coding, as this study's audit demonstrates is feasible, Table 9's methodology), could potentially be incorporated into future Layer 1 model iterations, representing a feedback pathway (Stage 5 to Stage 2, Figure 3) with direct model-improvement value beyond its immediate decision-level value.

For regulators and policymakers, the portfolio stress simulation results (Table 10, Section 5.2) suggest that fairness-metric stress testing — examining how demographic disparities evolve under economic stress scenarios, analogous to how credit risk stress testing examines how default rates evolve — represents a potentially valuable addition to existing model risk management and stress testing frameworks (Table 7's OCC reference). The Regional Economic Shock scenario's finding that geographically-concentrated stress produces the starkest M2 vs. M6 fairness divergence (Section 5.2) suggests that geographic concentration risk assessments, already a component of conventional credit risk management, could be extended to jointly assess fairness-metric concentration risk — providing a risk-management-framed (rather than solely compliance-framed) rationale for regulatory attention to this dimension.

Limitations and Future Research

Several limitations merit acknowledgment. First, this study's protected-attribute data, while collected under a documented consent and linkage process (Section 3.1), represents a substantial proportion of the full sample (13.5% missing, Table 3) — fairness metrics computed on the non-missing subsample (Table 4) may not fully represent the full applicant population if missingness is non-random with respect to fairness-relevant characteristics, a limitation common to fairness evaluation research relying on self-reported demographic data (Chen et al., 2019) but one that this study cannot fully resolve given its reliance on real-world data with real-world consent and collection constraints.

Second, this study's portfolio stress simulations (Table 10), while calibrated to historical recession-period and regional-shock feature distribution patterns, remain simulations rather than observations of genuine stress periods; the 2021–2026 observation window, while including some economic volatility, did not include a severe recession of the magnitude simulated in Table 10's severe recession scenario, and the simulation's assumptions regarding how feature distributions and default propensities would jointly shift under such a scenario, while grounded in historical patterns from prior recessions, necessarily involve extrapolation. Future research incorporating data from organizations that experience genuine severe economic stress during an observation period would provide stronger validation of this study's stress simulation findings.

Third, the FACI framework's Layer 3 override audit (Table 9) reflects this study's specific participating lenders' override policies and practices; the specific override categories identified (Thin-File Manual Review, Community Lending Program Exceptions, High-Income Volatility Flags, Fraud-Pattern Secondary Review) may not generalize to other lending organizations with different override policies, though the broader finding that override layers can provide both inclusion-positive and risk-identification value simultaneously (Section 5.1) may generalize even if specific category definitions differ. Fourth, this study's out-of-time validation (Table 8) shows modest temporal degradation in M6's fairness advantage; future research employing longer out-of-time windows, or genuinely prospective (rather than retrospective out-of-time) validation, would provide stronger evidence regarding FACI's fairness benefits' temporal stability — a question of direct practical importance given Layer 4's continuous-monitoring design premise (Figure 1, Figure 3).

Fifth, this study examines a single credit product category (unsecured consumer personal installment loans); the FACI framework's applicability to other credit products — mortgages, auto loans, small business lending — each with distinct regulatory frameworks (some, like mortgage lending, with HMDA-based fairness monitoring infrastructure not directly examined in this study) and distinct alternative-data landscapes, represents an important direction for future research extending this study's findings beyond its examined product category.

Conclusion

This study has developed and empirically validated the Fairness-Aware Credit Intelligence (FACI) framework, demonstrating that an integrated, four-layer organizational decision architecture — combining predictive modeling, multi-mechanism fairness intervention, explainability and human override, and portfolio simulation and governance — achieves a frontier-shifting configuration that neither unconstrained machine learning models nor single-constraint fairness-aware models achieve individually: near-maximal predictive accuracy (AUC-ROC within 0.003 of the unconstrained benchmark) alongside near-minimal demographic disparities across multiple fairness metrics simultaneously, higher simulated portfolio returns than single-constraint alternatives, and lower default rates than the unconstrained benchmark.

The subgroup analysis's finding that these gains are disproportionately concentrated among thin-file applicants — a population of particular financial inclusion interest — combined with the portfolio stress simulation's finding that FACI's fairness mechanisms confer resilience benefits under economic stress, together suggest that the conventional framing of fairness-aware credit modeling as a compliance cost competing with business objectives may substantially understate what well-architected, integrated approaches can achieve. This is not to suggest that fairness-aware credit modeling is costless or that all accuracy-fairness tensions documented in the prior literature

ISSN:3048-7722

(Kleinberg et al., 2017) are resolved by architectural choices alone — M6's accuracy, while close to M2's, is not identical, and M6's fairness metrics, while substantially better than M2's, are not perfect (Demographic Parity Difference of 0.038, not zero).

Rather, this study's contribution is to demonstrate that the achievable frontier is substantially more favorable than single-constraint studies suggest, and that realizing this more favorable frontier requires organizational architecture — spanning modeling, fairness intervention, explainability, override governance, and portfolio monitoring — rather than model-level intervention alone.

As machine learning credit models continue to scale across consumer lending markets, and as the regulatory landscape governing these models continues to evolve and intensify (Table 7), the FACI framework, maturity roadmap (Figure 5), and empirical evidence developed in this study provide lenders, regulators, and researchers with both a conceptual architecture and an evidentiary basis for pursuing credit decisioning systems that are simultaneously more accurate, more inclusive, more explainable, and more defensible — not as competing objectives requiring trade-off, but as complementary outcomes of well-designed organizational decision architecture.

REFERENCES

1. Agarwal, A., Beygelzimer, A., Dudík, M., Langford, J., & Wallach, H. (2018). A reductions approach to fair classification. Proceedings of the 35th International Conference on Machine Learning, 60–69.
2. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671–732. https://doi.org/10.15779/Z38BG31
3. 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.
4. Bhutta, N., Hizmo, A., & Ringo, D. (2022). How much does racial bias affect mortgage lending? Evidence from human and algorithmic credit decisions. Federal Reserve Board Working Paper.
5. Chen, J., Kallus, N., Mao, X., Svacha, G., & Udell, M. (2019). Fairness under unawareness: Assessing disparity when protected class is unobserved. Proceedings of the Conference on Fairness, Accountability, and Transparency, 339–348.
6. Consumer Financial Protection Bureau. (2024). Adverse action notification requirements and artificial intelligence: CFPB circular on Regulation B compliance for complex algorithms. CFPB.
7. Corbett-Davies, S., & Goel, S. (2018). The measure and mismeasure of fairness: A critical review of fair machine learning. arXiv preprint arXiv:1808.00023.
8. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, 214–226.
9. European Commission. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.
10. Financial Conduct Authority. (2023). FG23/3: Consumer Duty guidance for firms. UK Financial Conduct Authority.
11. Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. The Journal of Finance, 77(1), 5–47.
12. Gillis, T. B. (2022). The input fraud problem: Disparate impact and the data-generating process in algorithmic decision-making. Journal of Legal Analysis, 14(1), 1–46.
13. Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. Advances in Neural Information Processing Systems, 29, 3315–3323.
14. Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11), 2767–2787.
15. Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. The Quarterly Journal of Economics, 133(1), 237–293.
16. Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. Proceedings of the 8th Innovations in Theoretical Computer Science Conference, 1–23.
17. Kozodoi, N., Jacob, J., & Lessmann, S. (2022). Fairness in credit scoring: Assessment, implementation and profit implications. European Journal of Operational Research, 297(3), 1083–1094.
18. Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), 124–136.
19. Liu, L. T., Dean, S., Rolf, E., Simchowitz, M., & Hardt, M. (2018). Delayed impact of fair machine learning. Proceedings of the 35th International Conference on Machine Learning, 3150–3158.
20. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.
21. 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.
22. Office of the Comptroller of the Currency. (2023). Model risk management handbook (Version 2.0). OCC.
23. Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mistaking algorithmic fairness for fairness in practice: Investigating algorithmic decisions in hiring. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 469–481.

ISSN:3048-7722
24. Rambachan, A., Kleinberg, J., Mullainathan, S., & Ludwig, J. (2020). An economic perspective on algorithmic fairness. AEA Papers and Proceedings, 110, 91–95.
25. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). 'Why should I trust you?': Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.
26. Romei, A., & Ruggieri, S. (2014). A multidisciplinary survey on discrimination analysis. The Knowledge Engineering Review, 29(5), 582–638.
27. Sammangi, H., Jagatha, A., & Liu, J. (2025b). Harnessing generative AI and large language models for revolutionizing cybersecurity in the Internet of Things: Ethical and privacy implications. Engineering: Open Access, 3(6), 1–12.
28. Sammangi, H., Jagatha, A., & Liu, J. (2025c). Integrating blockchain technology into telemedicine: A framework for enhancing data privacy and security. Engineering: Open Access, 3(6), 1–7.
29. Sharma, G., Singh, J., Sammangi, H., Sharma, M., Pandey, R., Srivastava, S., Agarwal, G., & Singh, I. (2025a). A comprehensive assessment of developing a forecasting model for kidney stone formation using deep learning approaches. In H. Sharma, A. Chakravorty, S. Hussain, & R. Kumari (Eds.), Artificial Intelligence: Theory and Applications (Vol. 5588, pp. 121–132). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-1918-4_9
30. 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.
31. Skeem, J. L., & Lowenkamp, C. T. (2016). Risk, race, and recidivism: Predictive bias and disparate impact. Criminology, 54(4), 680–712.
32. Verma, S., & Rubin, J. (2018). Fairness definitions explained. Proceedings of the International Workshop on Software Fairness, 1–7.
33. Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harvard Journal of Law & Technology, 31(2), 841–887.
34. Wang, R., Harper, F. M., & Zhu, H. (2020). Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–14.
35. World Bank Group. (2023). Credit scoring approaches guidelines: Financial inclusion and responsible lending. World Bank Financial Inclusion Practice.
36. Zhang, B. H., Lemoine, B., & Mitchell, M. (2018). Mitigating unwanted biases with adversarial learning. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 335–340.
37. Zliobaite, I. (2017). Measuring discrimination in algorithmic decision making. Data Mining and Knowledge Discovery, 31(4), 1060–1089.