From Early Warning to Workout: Operationalizing AI-Assisted Restructuring Recommendations in Small-Business Lending Across Community Banks, Credit Unions, and CDFIs

Mariam Mkrtchyan [1], Gelashvili Lilit [2]

[1] University of Malta, MALTA
[2] University of Ljubljana, SLOVENIA

ABSTRACT

Small-business loan distress is usually recognized late, after missed payments or covenant breaches have already narrowed the options for a workable restructuring. A recent multi-institution validation study of an AI-assisted early-intervention and loan-restructuring recommendation engine across US community banks, credit unions and Community Development Financial Institution (CDFI) portfolios [1] demonstrated that such a system can be evaluated consistently across lenders that differ in charter, size, data maturity and mission. This article adopts that engine as its reference design and addresses the question that validation leaves open: how a lender should put the engine into everyday credit operations. We propose a five-layer implementation framework covering data readiness, risk-signal generation, restructuring-option ranking, human-in-the-loop credit decisioning and post-modification monitoring, set out a transparent expected-value formulation for ranking restructuring options, show how the framework should be adapted for community banks, credit unions and CDFIs, and map each layer to current US supervisory and consumer-protection expectations. We then implement the engine's two stages and run the back-test stage of our evaluation roadmap on public repayment data for 24,000 credit-card borrowers. The early-warning stage reached an AUC of 0.78 and, at a watch-list capacity of 12 percent, flagged 36 percent of eventual defaulters with 70 percent precision; its reason codes were stable across bootstrap refits (93 percent agreement). Lead time was the weak point: scored two months earlier, the engine flagged only 3 of the 316 defaulters whose first arrears appeared later, so most of its signal recognised delinquency that had already begun. The option-ranking stage favoured term extension for 51 percent of flagged borrowers, but 37 percent of recommendations changed when the assumed option effects were varied by up to 50 percent, so cure-rate evidence must be collected before the ranking is trusted. The results support the staged local evaluation proposed here; they come from consumer card data, because no public small-business dataset records repayment histories and restructuring outcomes, and they do not test institution-type effects.

Keywords: Small-Business Lending; Early-Warning Systems; Loan Restructuring; Community Banks; Credit Unions; CDFIs; Model Risk Management; Explainable AI; Human-in-the-Loop; Financial Inclusion

Introduction

Early, well-structured intervention is the least costly point at which a lender can preserve a viable small-business relationship. Once a borrower has missed several payments, cash reserves are often exhausted, collateral values may have weakened and the lender's choices narrow to forbearance, charge-off or liquidation. Earlier recognition, by contrast, leaves room for modest and reversible concessions such as a short payment deferral or a re-amortization that matches the borrower's seasonal cash flow.

Small firms are difficult to assess with conventional tools. They are informationally opaque, rarely produce audited statements and depend heavily on relationship lending, in which loan officers accumulate soft

information through repeated contact [2][3]. Information asymmetry of this kind can lead lenders to ration credit rather than price it [4], and smaller, more local institutions have historically held a comparative advantage in using soft information [5]. Community banks, credit unions and CDFIs therefore occupy a central position in small-business finance, yet they often lack the analytical infrastructure of larger lenders.

The multi-institution validation study by Karim and Elias [1] addresses this gap directly. It evaluates an AI-assisted engine that both identifies borrowers at risk of financial distress and recommends restructuring options, and it does so across three types of institution with very different portfolios, data systems and missions. Its central contribution is evidence that one recommendation engine can be validated across community banks, credit unions and CDFIs together rather than inside a single lender. That design matters because a model that holds across heterogeneous lenders is a far stronger candidate for broad adoption than a model tuned to one portfolio.

Validation, however, is not deployment. An institution that reads [1] must still decide how the engine's outputs enter its loan-review cycle, who may override them, how borrowers are informed of decisions, how modifications are documented and disclosed, and how examiners will evaluate the model. These operational questions determine whether a validated model actually changes outcomes for borrowers. This article treats the engine in [1] as a reference design and builds an implementation framework around it. Its contributions are fourfold:

- A five-layer operating model that adapts the validated engine to routine credit workflows, with defined owners and auditable outputs for each layer.
- A transparent formulation for ranking restructuring options and a human-in-the-loop decision protocol that preserves the relationship-lending advantage of community lenders.
- Guidance for adapting the framework to the distinct constraints of community banks, credit unions and CDFIs, mirroring the cross-institution design of the reference study.
- A mapping of every layer to current US supervisory and consumer-protection expectations, and a four-stage local evaluation roadmap with explicit decision gates.
- A back-test of the engine's two stages on public repayment data, which shows what the first stage of the evaluation roadmap reveals in practice and where the engine's evidence is weakest.

Background and Related Work

Distress Prediction for Small Firms

Distress prediction has a long history in accounting and finance. Early univariate work showed that individual financial ratios deteriorate well before failure [6], and multivariate discriminant analysis produced the widely used Z-score [7]. Logistic models later offered probabilistic estimates of bankruptcy that were easier to interpret [8]. These models were developed mainly for larger, publicly reporting firms. When adapted to small and medium-sized enterprises, models built specifically for that segment were shown to outperform generic corporate scores [9], underscoring that small-business distress has its own signature.

Machine Learning in Credit Risk

Machine-learning methods can combine accounting ratios with transaction-level behavior. Models trained on account and transaction data have been shown to forecast consumer credit delinquency months in advance [10], and large benchmarking exercises indicate that ensemble methods frequently outperform traditional logistic scorecards [11]. Gradient-boosted trees in particular have become a standard tool for tabular data because of their accuracy and scalability [12]. The cost of this flexibility is reduced transparency, which post-

hoc explanation methods such as LIME and SHAP aim to address by attributing each prediction to its input features [13][14].

Table 1 summarizes how these families of approaches compare on the dimensions that matter most for an early-intervention engine in community lending.

Table 1. Comparison of distress-prediction approaches for small-business early warning.

Approach	**Typical inputs**	**Strengths**	**Limitations for small-business use**
Ratio-based discriminant or logistic models [7][8]	Annual financial-statement ratios	Transparent; long track record; easy to explain to examiners	Depend on timely, reliable statements that many small firms do not produce
SME-specific scorecards [9]	Ratios calibrated on SME samples	Better fit to the small-firm segment	Still statement-based; slow to react to sudden cash-flow shocks
Machine-learning ensembles [10][11][12]	Ratios plus payment, deposit and transaction behavior	Higher accuracy; can detect stress months earlier	Lower transparency; need governance and explanation tooling
Validated early-warning and restructuring engine [1]	Multi-source borrower data across institution types	Couples detection with recommended remedies; validated across lender types	Requires local confirmation and operational integration

Restructuring as a Decision Problem

The troubled debt has long shown that private restructuring can preserve more value than formal default when it is undertaken early and on realistic terms [15]. Evidence from large-scale modification programs also shows that the design and targeting of modifications materially affect re-default and cure rates [16]. US banking regulators encourage prudent workouts rather than reflexive charge-offs. The 2020 interagency statement on loan modifications [17] and the 2023 interagency policy statement on accommodations and workouts [18] both treat restructuring as a legitimate credit tool when it is well documented and grounded in the borrower's repayment capacity. On the accounting side, ASU 2022-02 replaced troubled-debt-restructuring recognition for entities applying the current expected credit loss standard with enhanced disclosures of modifications made to borrowers experiencing financial difficulty [19]. Choosing among term extension, payment deferral, rate reduction, re-amortization or collateral changes is therefore a structured decision with regulatory, accounting and relationship consequences, not merely a prediction.

The Reference Engine and Its Validation

The study in [1] is notable because it unites two functions that are usually treated separately, distress detection and remedy recommendation, and validates them across community banks, credit unions and CDFIs at once. These lenders differ in charter, supervisor, borrower mix and mission. CDFIs in particular serve borrowers whom conventional scorecards often exclude. A model that performs consistently across this range offers a credible basis for adoption by institutions that could not build such a system alone. The present framework adopts three of its elements: the two-stage logic of a distress signal followed by ranked remedies, assessment across institution types, and the positioning of the engine as decision support rather than automation. Each is

extended below with defined owners, institution-specific configuration and a formal human-in-the-loop protocol.

The Intervention Window

The practical value of an early-warning engine lies in the time it buys. Figure 1 illustrates the concept schematically. A borrower's financial health typically erodes gradually before it collapses, while the range of feasible restructuring options narrows steadily as reserves are consumed and collateral weakens. Conventional detection relies on hard events such as a missed payment, which often occurs only after much of that range has disappeared. An engine that flags emerging stress from payment, deposit and transaction behavior creates an additional intervention window in which lighter and less costly remedies remain available.

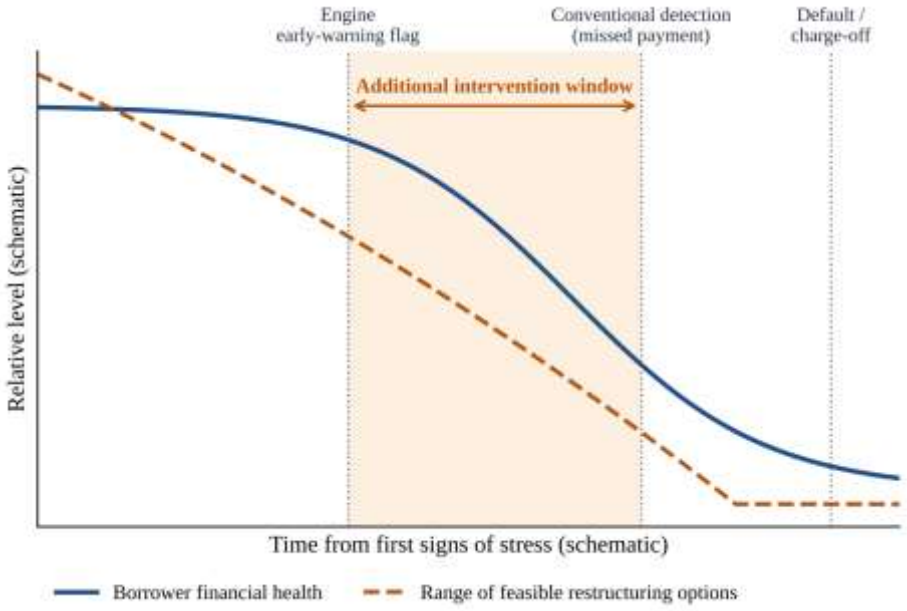

Figure 1. Schematic of the additional intervention window created by early warning

The length of this window, measured as the lead time between an engine flag and the first conventional delinquency signal, is one of the most important quantities an implementing institution should measure locally. A model can be accurate in a statistical sense yet operationally useless if it flags borrowers only days before they would have been identified anyway. The evaluation roadmap proposed later therefore treats lead time as a primary outcome alongside conventional accuracy.

Implementation Framework

The framework converts the validated engine in [1] into a repeatable monthly credit process organized in five layers, shown in Figure 2. Each layer has a clear owner and a defined output, so that examiners, internal audit and model validators can trace any restructuring back to the data and the human judgment behind it. A feedback loop from post-modification monitoring to the risk-signal and option-ranking layers allows the engine to be recalibrated as outcomes accumulate.

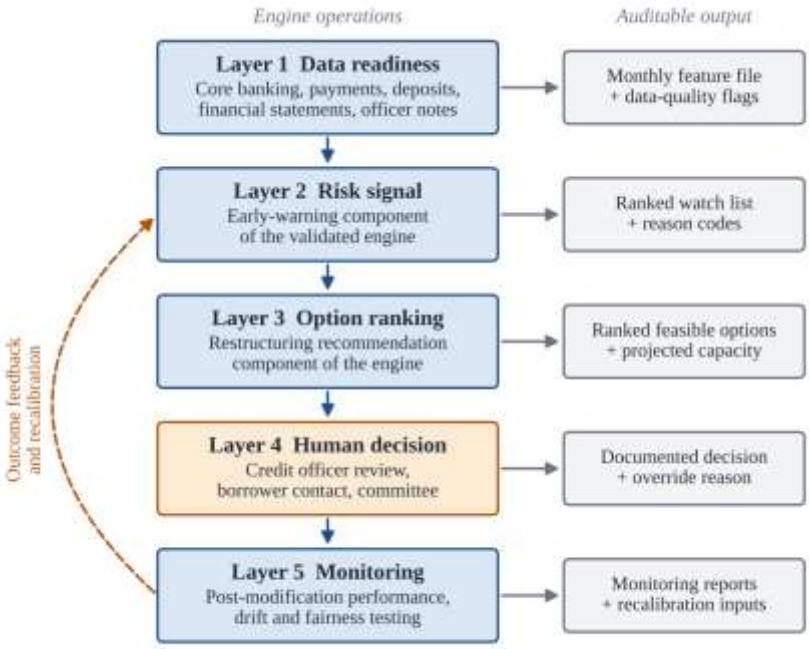

Figure 2. Five-layer framework for operationalizing the validated early-intervention and restructuring engine

Table 2 summarizes the purpose, main output and accountable owner for each layer. Separating the risk signal from option ranking reflects the two-stage logic of the reference engine and allows each stage to be monitored independently; a drop in option-acceptance rates, for example, can then be traced to the ranking layer without implying any fault in detection.

Table 2. Layers of the implementation framework, outputs and owners.

Layer	**Purpose**	**Main output**	**Owner**
1. Data readiness	Standardize core-banking, payment, deposit-flow and financial-statement data to the engine's input schema	Monthly borrower feature file with data-quality flags	Data and IT
2. Risk signal	Run the early-warning component to score distress risk	Ranked watch list with reason codes	Credit risk
3. Option ranking	For flagged borrowers, rank feasible restructuring options	Top options per borrower with projected repayment capacity	Credit risk
4. Human decision	Officer reviews, contacts the borrower, then accepts, modifies or overrides	Documented decision and override reason	Lending and credit committee

Layer	Purpose	Main output	Owner
5. Monitoring	Track modified loans and model behavior; feed outcomes back	Post-modification performance, drift and fairness reports	Model risk and internal audit

Layer 1: Data Readiness

Data readiness is the layer most likely to delay implementation, particularly at smaller institutions. The engine's inputs must be assembled monthly from systems that were not designed to share data, and each source carries characteristic quality problems. Typical domains include loan servicing, deposit and cash-flow behavior, financial statements, collateral, officer notes and local economic indicators; their characteristic problems include manual postings, stale statements, infrequent reappraisal and unstructured notes, all of which the readiness layer should flag before scoring. Machine-learning systems accumulate hidden maintenance costs through data dependencies and undeclared consumers [20], so each feed should be documented with an owner, refresh schedule and validation rule.

Layers 2 and 3: Risk Signal and Option Ranking

Layer 2 runs the early-warning component of the engine and produces a ranked watch list. Each flagged borrower should carry reason codes derived from feature-attribution methods [14], so that officers understand why the borrower was flagged and so that any adverse action can be explained accurately. Layer 3 then evaluates feasible restructuring options for each flagged borrower. One transparent way to express this ranking, consistent with the recommendation role of the reference engine, is an expected-value score for each candidate option o:

$$V(o) = P(cure \mid o) \times PV(cure, o) + [1 - P(cure \mid o)] \times PV(default, o) - C(o)$$

Here P(cure | o) is the estimated probability that the borrower returns to performing status under option o, PV(cure, o) and PV(default, o) are the present values of expected cash flows to the lender under each outcome, and C(o) captures administrative and relationship costs. Options are ranked by V(o) and filtered by policy constraints, such as maximum term extensions. Expressing the ranking in this form keeps the trade-off visible to credit committees and examiners. Table 3 sets out a catalogue of common options with the conditions under which each tends to suit a borrower.

Table 3. Catalogue of restructuring options considered by the option-ranking layer.

Option	Typically suited to	Key considerations
Short payment deferral	Temporary, identifiable cash-flow shock with a credible recovery date	Interest accrual and capitalization; risk of masking deeper problems
Interest-only period	Seasonal or cyclical revenue dips	Principal reduction is delayed; set a clear end date
Term extension or re-amortization	Sustained but moderate decline in debt-service capacity	Lower payment versus longer exposure; collateral coverage over the new term
Rate reduction	Viable business with payment burden as the main constraint	Direct income cost to lender; fair-lending consistency

Option	Typically suited to	Key considerations
Collateral or guarantee restructuring	Weak coverage but viable operations	Legal costs; borrower capacity to pledge additional support
Referral to technical assistance	Operational or management weaknesses (common in CDFI portfolios)	Combine with a financial modification; track outcomes separately

Human-in-the-Loop Decisioning

The engine should never modify a loan on its own. Figure 3 shows the decision protocol for Layer 4. A flagged borrower is first reviewed by a credit officer, who examines the reason codes and ranked options before contacting the borrower to verify current cash flow. The officer, or the credit committee for larger exposures, then accepts the top-ranked option, modifies its terms, or overrides the recommendation entirely. Whatever the decision, the reason is recorded in a structured field.

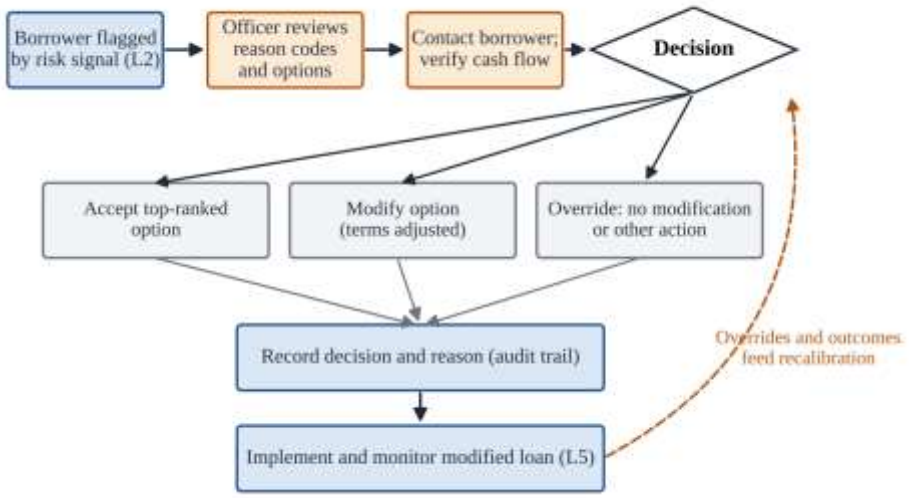

Figure 3. Human-in-the-loop decision protocol for engine recommendations

Override records serve three purposes. They form an audit trail that demonstrates effective challenge of the model, as model risk guidance expects [21]. They capture soft information that the engine cannot observe, preserving the relationship-lending advantage of community institutions [3][5]. And they provide training signal: systematic overrides in a particular segment or for a particular option indicate where the engine should be recalibrated. Override rates that are very low may suggest automation bias, in which staff accept recommendations without scrutiny, while very high rates suggest the engine is not trusted or not suited to the portfolio. Both extremes warrant review.

Adapting the Framework to Institution Type

Because the reference engine was validated across three types of institution [1], implementation should respect the same distinctions. Community banks, credit unions and CDFIs differ in their regulators, typical loan sizes, data maturity and mission, and each difference changes how the framework should be configured. Table 4 summarizes the main adaptations.

Table 4. Adapting the framework to community banks, credit unions and CDFIs.

Dimension	Community banks	Credit unions	CDFIs
Typical small-business exposure	Commercial and industrial and owner-occupied real estate loans	Member business loans, often smaller balances	Microloans and mission-driven loans to underserved borrowers
Data maturity	Moderate; statements often available	Variable; deposit data strong, statements sparse	Variable; strong relationship data, thin credit files
Integration point	Existing commercial loan review cycle	Member business lending review	Portfolio management and technical-assistance teams
Option emphasis	Term extension, re-amortization, collateral changes	Deferral and re-amortization	Deferral combined with technical assistance
Primary objective weighting	Loss minimization and relationship retention	Member service and loss minimization	Borrower survival and mission outcomes alongside loss

For CDFIs, the objective function in the option-ranking layer may legitimately place weight on borrower survival and community outcomes in addition to expected recovery. That weighting should be explicit and approved by the board rather than embedded silently in model parameters, so that the institution can explain to funders and examiners why a lower-recovery option was preferred.

Governance, Fairness and Regulatory Alignment

Adopting the engine is a model-risk decision first and a technology decision second. Table 5 maps the principal US supervisory and consumer-protection expectations to the layer of the framework where each is met.

Table 5. Mapping of supervisory and consumer-protection expectations to framework layers.

Expectation	Source	Where it is met
Independent validation, ongoing monitoring and effective challenge	Model risk management guidance [21]	Local evaluation stages; Layer 4 override records; Layer 5 reports
Workouts grounded in repayment capacity and well documented	Interagency workout and modification statements [18][17]	Layers 3 and 4
Disclosure of modifications to borrowers in financial difficulty	ASU 2022-02 [19]	Layer 4 documentation feeding financial reporting
Specific and accurate reasons for adverse credit decisions, including with complex models	CFPB Circular 2022-03 [22]	Layer 2 reason codes and feature attributions

Expectation	Source	Where it is met
Collection and fair-lending use of small-business application data	Section 1071 rule [23]	Fairness testing in Layer 5
Due diligence and oversight of vendor-supplied models	Third-party risk guidance [24]	Vendor review before Layer 1 go-live
Valid, reliable, fair, explainable and accountable AI	NIST AI RMF 1.0 [25]	All layers

Fairness deserves particular attention because community lenders, and CDFIs especially, serve many minority-owned and low-income-area businesses. Machine-learning credit models can widen disparities between groups even when they improve overall accuracy [26], and algorithmic lending has been shown to reduce, but not eliminate, pricing differences between groups [27]. Implementing institutions should therefore test whether flag rates and recommended concessions differ across groups at similar levels of estimated risk, using criteria such as equality of opportunity [28]. A recommendation engine can widen gaps quietly if one group is consistently routed to term extensions and another to rate relief without a credit-based reason.

Models also degrade as economic conditions shift. Concept drift, in which the relationship between inputs and outcomes changes over time, is a well-documented challenge for deployed predictive systems [29]. Layer 5 should therefore monitor input distributions, score distributions and calibration at least quarterly, with predefined thresholds that trigger review or recalibration.

Staged Local Evaluation

An institution should confirm the findings of [1] on its own portfolio before relying on them. Figure 4 presents an indicative four-stage roadmap, with decision gates set in advance by the model risk function. Each gate has explicit pass criteria, and failure at a gate returns the project to recalibration rather than forward to wider use.

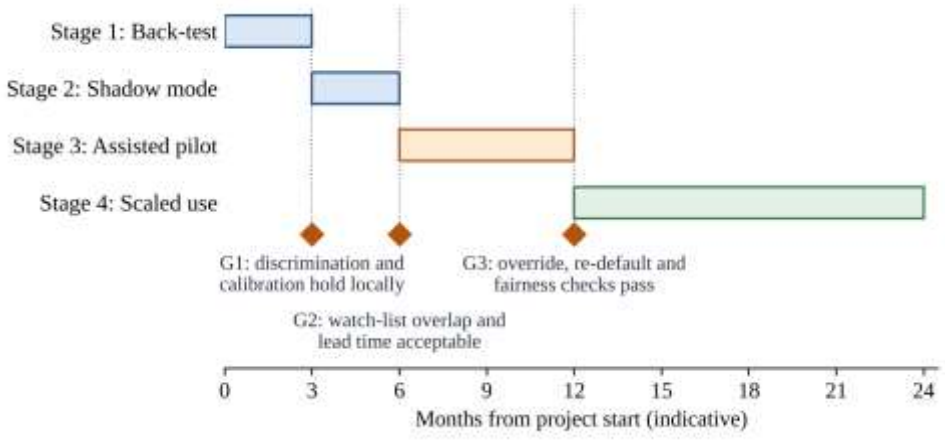

Figure 4. Indicative staged evaluation roadmap with decision gates

In the back-test stage, two to three years of historical loans are scored and the engine's flags and recommended options are compared with what actually occurred. In shadow mode, the engine runs live for one or two review cycles without its outputs being shown to officers, and its watch list is compared with the institution's

own. The assisted pilot exposes recommendations to a subset of officers or branches and records acceptance, overrides and time to intervention. Only after these stages does the engine move to portfolio-wide use with quarterly monitoring. Table 6 lists suggested outcome measures. Each should be reported by institution type and borrower segment, mirroring the design of [1].

Table 6. Suggested outcome measures for local evaluation and ongoing monitoring.

Measure	Definition	Stage used
Discrimination (AUC)	Ability of the risk score to rank eventual distressed borrowers above others	All stages
Calibration (Brier score) [30]	Agreement between predicted probabilities and observed distress rates	All stages
Lead time	Months between engine flag and first conventional delinquency signal	Back-test, shadow, pilot
Option acceptance and override rate	Share of recommendations accepted, modified or overridden, with reasons	Pilot, scaled use
Re-default rate	Share of modified loans returning to delinquency within 12 months	Pilot, scaled use
Disparity in flags and concessions	Differences across borrower groups at similar risk levels	All stages

Back-Test of the Reference Engine on Public Data

The back-test stage of the roadmap can be run before any institution's own data are available, provided a public dataset carries the signals the engine needs. No public small-business dataset records monthly repayment histories together with restructuring outcomes. We therefore used the most widely studied public repayment panel, the Taiwan credit-card default data[31] (24,000 borrowers in the openly distributed copy; default rate 22.1 percent). It records six months of repayment status, statement balances and payments, and whether the borrower defaulted the following month. These are consumer accounts, not small businesses, so the back-test checks the engine's mechanics and the evaluation measures in Table 6, not its small-business performance.

Implementation

Layer 1 checked completeness and coding: 1.1 percent of records carried undocumented education codes and about 10 percent had three or more months without a statement balance, the thin files the data-readiness layer is meant to surface. Layer 2 was a LightGBM classifier[32] trained on 56 percent of borrowers, with 14 percent for validation and 30 percent (7,200 borrowers) held out for testing. Its inputs were the credit limit, six months of repayment status, balances and payments, and derived utilisation, payment-ratio, balance-trend and arrears-count features. Sex, age, marital status and education were excluded from the model. The watch list was the top 12 percent of scores, a capacity set in advance on the validation data. Each flagged borrower received three reason codes from TreeSHAP attributions[14], mapped to plain-language categories. Layer 3 computed V(o) for five options from Table 3, with P(cure | o) obtained by applying an assumed odds uplift for each option to the model's cure probability. The assumed uplifts and concession costs are stated, not estimated: no public data record what happens to borrowers after a modification. Observable signals adjusted the uplifts in line with Table 3, for example deferral for a single recent missed payment with continuing payments, and term extension for sustained arrears with high utilisation.

Risk Signal and Lead Time

Table 7. Back-test of the risk-signal layer on held-out borrowers: scoring at the latest month versus two months earlier.

Measure	Scored at latest month	Scored two months earlier
Discrimination (AUC)	0.779	0.729
PR-AUC	0.562	0.464
Brier score[30]	0.133	0.149
Defaulters flagged (recall at 12% capacity)	0.363	0.303
Precision of watch list	0.697	0.551

The risk signal discriminated well and was well calibrated (expected calibration error 0.013). Its explanations, however, show what it was mostly detecting: for 84 percent of flagged borrowers the principal reason was recent repayment status, and for 15 percent it was the number of months in arrears. The lead-time measure in Table 6 tests this directly. Of the 1,592 defaulters in the test set, 874 had no arrears two months before the outcome month. The engine scored at that earlier point flagged 482 defaulters, but only 4 of them were still clean at the time. Of the 316 defaulters whose first arrears appeared in the final two months, it flagged 3. On repayment data alone, the intervention window that Figure 1 depicts was therefore very narrow. Earlier warning would need the cash-flow, transaction and account-behaviour data that small-business lenders hold and this public panel lacks, which is why the data-readiness layer comes first.

Option Ranking

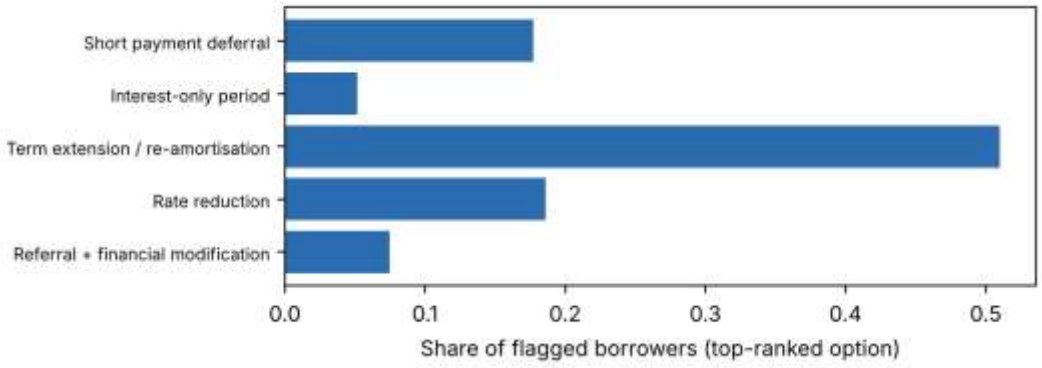

Figure 5. Top-ranked restructuring option for flagged test borrowers under the stated option-effect assumptions [authors' analysis].

Under the stated assumptions, term extension ranked first for 51 percent of flagged borrowers, rate reduction for 19 percent and a short deferral for 18 percent (Figure 5). Ranking by V(o) rather than offering every borrower a term extension raised the expected value to the lender for 49 percent of flagged borrowers, by NT$ 371 per borrower on average. The rankings were fragile, though. When each assumed uplift and concession cost was drawn independently from 50 to 150 percent of its stated value (200 draws), an average of 37 percent of borrowers received a different top option (10th to 90th percentile: 19 to 56 percent). The ranking layer is only as good as its estimates of P(cure | o). Those estimates must come from the institution's own modification history, collected during the assisted pilot and recorded as the re-default measure in Table

6, before option recommendations are presented to borrowers as more than one input to the officer's judgement.

Conclusion

The multi-institution validation reported in [1] establishes that an AI-assisted early-intervention and loan-restructuring engine can perform across community banks, credit unions and CDFIs. This article builds on that foundation by specifying how such an engine can be operated responsibly: a five-layer workflow with auditable outputs, a transparent option-ranking formulation, a human-in-the-loop decision protocol, institution-specific adaptation, explicit regulatory mapping and a staged local evaluation with decision gates. A back-test on public repayment data showed that the engine's two stages can be built and audited as specified, but also that its early warning mostly recognised delinquency already under way and that its option rankings depend heavily on cure-rate assumptions that only an institution's own modification history can supply. Together, these elements offer lenders a practical path from a validated model to earlier, fairer and better-documented workouts for financially distressed small-business borrowers, and they provide a common structure through which future implementation studies can extend the cross-institution evidence that the reference study began.

References

[1] Karim, N. B., & Elias, D. (2022). Multi-Institution Validation of an AI-Assisted Early-Intervention and Loan-Restructuring Recommendation Engine for Financially Distressed Small-Business Borrowers Across US Community Banks, Credit Unions, and CDFI Portfolios. *The Computertech*, 107-120.

[2] Berger, A. N., & Udell, G. F. (1998). The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle. *Journal of Banking & Finance*, *22*(6–8), 613–673.

[3] Petersen, M. A., & Rajan, R. G. (1994). The benefits of lending relationships: Evidence from small business data. *The Journal of Finance*, *49*(1), 3–37.

[4] Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American Economic Review*, *71*(3), 393–410.

[5] Berger, A. N., Miller, N. H., Petersen, M. A., Rajan, R. G., & Stein, J. C. (2005). Does function follow organizational form? Evidence from the lending practices of large and small banks. *Journal of Financial Economics*, *76*(2), 237–269.

[6] Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, *4*, 71–111.

[7] Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589–609.

[8] Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, *18*(1), 109–131.

[9] Altman, E. I., & Sabato, G. (2007). Modelling credit risk for SMEs: Evidence from the U.S. market. *Abacus*, *43*(3), 332–357.

[10] 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.

[11] 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.

[12] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM.

[13] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). ACM.

[14] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774).

[15] Gilson, S. C., John, K., & Lang, L. H. P. (1990). Troubled debt restructurings: An empirical study of private reorganization of firms in default. *Journal of Financial Economics*, *27*(2), 315–353.

[16] Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., Piskorski, T., & Seru, A. (2017). Policy intervention in debt renegotiation: Evidence from the Home Affordable Modification Program. *Journal of Political Economy*, *125*(3), 654–712.

[17] Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, National Credit Union Administration, Office of the Comptroller of the Currency, Consumer Financial Protection Bureau, & Conference of State Bank Supervisors. (2020). *Interagency statement on loan modifications and reporting for financial institutions working with customers affected by the coronavirus (revised)*.

[18] Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, National Credit Union Administration, & Office of the Comptroller of the Currency. (2023). *Policy statement on prudent commercial real estate loan accommodations and workouts*.

[19] Financial Accounting Standards Board. (2022). *Accounting Standards Update No. 2022-02, Financial instruments—Credit losses (Topic 326): Troubled debt restructurings and vintage disclosures*.

[20] Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. In *Advances in Neural Information Processing Systems* (Vol. 28, pp. 2503–2511).

[21] Board of Governors of the Federal Reserve System, & Office of the Comptroller of the Currency. (2011). *Supervisory guidance on model risk management* (SR Letter 11-7; OCC Bulletin 2011-12).

[22] Consumer Financial Protection Bureau. (2022). *Consumer Financial Protection Circular 2022-03: Adverse action notification requirements in connection with credit decisions based on complex algorithms*.

[23] Consumer Financial Protection Bureau. (2023). Small business lending under the Equal Credit Opportunity Act (Regulation B). *Federal Register*, *88*, 35150.

[24] Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, & Office of the Comptroller of the Currency. (2023). Interagency guidance on third-party relationships: Risk management. *Federal Register, 88*, 37920.

[25] National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce.

[26] 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.

[27] Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2022). Consumer-lending discrimination in the FinTech era. *Journal of Financial Economics, 143*(1), 30–56.

[28] Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems* (Vol. 29, pp. 3315–3323).

[29] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys, 46*(4), Article 44.

[30] Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review, 78*(1), 1–3.

[31] Yeh, I.-C., & Lien, C.-H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 36(2), 2473-2480.

[32] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30* (pp. 3146-3154).