Financial Technology and Innovation

♦ **ISSN:** 2836-5372 **Vol:** 05, No. 02, 2025 ♦ PP. 10–19
**Article DOI:** https://doi.org/10.54216/FinTech-I.050202


FULL LENGTH ARTICLE

# Platform Recovery Alignment in Financial Digital Transformation: A Friction-Aware Benchmarking Framework for Digital Finance Services

**N. Metawa**[1,*] **Iman Akour**[2] **Rania Itani**[3]

[1] College of Business Administration, University of Sharjah, UAE
[2] Department of Information Systems, University of Sharjah, UAE
[3] Murdoch University Dubai, UAE

**Emails:** nmetawa@sharjah.ac.ae · iakour@sharjah.ac.ae · rania.itani@murdoch.edu.au

Received: March 26, 2025 Revised: May 20, 2025 Accepted: August 28, 2025 ⋆ Corresponding author

## ABSTRACT

A key measure of financial digital transformation is the adoption, cost efficiency, channel migration, and transaction growth. They are all indicators of platform reach, but they provide little insight into what happens when a mishandled platform fails to provide a customer with their service, causes a customer to question their transactions, fails to inform a customer how much they are being charged, or has vulnerabilities that cause a security incident or disables the service for a customer due to restrictions. In this paper, a framework for platform recovery is designed that connects the type of digital-service failure to the depth of the institutional response to the failure, taking friction into account. Complaint narratives are symbolized by interpretable latent themes and intertwined with structured intake information in a chronologically validated response model. The predicted response depth provides a case-mix benchmark and recovery is compared with this benchmark at the company-product level, where results from the small samples are limited by empirical-Bayes shrinkage. The out-of-time evidence demonstrates that adding narrative themes modestly yet consistently to structured platform information boosts monetary-relief precision–recall area from 0.373 to 0.381 and any-relief precision–recall area from 0.519 to 0.525. The depth of recovery is different from material to material within the friction layers. The monetary-relief complaints rate is the highest, while the complaints rate for trust and security is the highest for the explanation closure. In units that are adequately observed, risk-adjusted alignment distinguishes recovery leaders from constrained, trust-intensive, and transaction-intensive types of units. The framework will go beyond just counting the number of complaints and relieve, and ask the more defensible question: Was the response from a financial platform deeper or shallower than what the service friction and case mix would reasonably suggest? It is not designed for automated adjudication or public ranking that would be part of a government system.

**Keywords:** Financial digital transformation ▪ Digital finance platforms ▪ Service recovery ▪ Complaint analytics ▪ Platform governance ▪ Text mining ▪ RegTech

## 1. INTRODUCTION

Financial institutions have moved a growing share of customer journeys from branches and call centers to mobile applications, web portals, payment interfaces, digital wallets, and platform-mediated ecosystems. Digital transformation in this setting is not simply the installation of technology. It alters value creation, organizational routines, customer touchpoints, data flows, and the boundaries between institutions and external providers [1, 2, 3]. Banking transformation therefore unfolds through a portfolio of interconnected platforms rather than a single channel project. The relevant managerial question is no longer whether a bank offers a mobile application, but whether the platform can support reliable transactions, understandable pricing, secure identity, accountable data use, and effective recovery when service breaks down.

The literature offers mature accounts of transformation strategy, organizational readiness, platform capability, and customer adoption [4, 5, 6, 7]. Financial research likewise explains how FinTech changes intermediation, lending, data use, and competition [8, 9, 10, 11]. Yet performance assessment remains weighted toward front-end outcomes: adoption, active users, digital sales, processing time, cost-to-income ratios, and channel migration. Such measures can improve while customer recovery deteriorates. A platform may process more transactions and still impose severe friction when a transfer fails, a payment is disputed, identity verification blocks legitimate access, or an account is closed without an intelligible pathway to resolution.

This gap matters because recovery is part of the platform value proposition. Digital services reduce the marginal cost of routine interactions, but they can also make exceptions harder to resolve. Automated workflows tend to perform well for standard cases and poorly at boundaries that require discretion, context, or cross-system coordination. Complaint records expose those boundaries. They reveal where users experience the platform as a sequence of hand-offs, inaccessible controls, inconsistent status messages, or unresolved liability rather than as a seamless service. Complaint outcomes additionally indicate whether the institution responded with an explanation, a non-monetary correction, or monetary relief.

Raw outcome rates, however, are not fair performance measures. Platform units face different complaint portfolios. A prepaid-card program may receive a concentration of access and balance disputes; a credit-card platform may face chargebacks and fee complaints; a wallet or transfer service may encounter fraud claims and transaction reversals. Comparing their monetary-relief rates without accounting for case mix confounds institutional response with the nature of the underlying friction. This paper addresses that problem by estimating an expected recovery depth from intake information and comparing it with the recovery actually delivered.

The resulting concept, *platform recovery alignment*, is a risk-adjusted diagnostic measure. Positive alignment indicates that observed response depth exceeds the level predicted from product, issue, channel, timing, narrative availability, and latent complaint themes. Negative alignment indicates a shallower response than expected. The measure is deliberately complaint-conditional. It does not estimate failure incidence because the database contains complaints rather than platform exposure, transaction volume, or total customers. Nor does it determine whether a complaint is legally meritorious. Its purpose is narrower and operational: to reveal where recovery behavior appears unusually deep or shallow after observable case mix has been considered.

The paper makes four contributions. First, it extends financial digital-transformation research from capability and adoption toward post-failure recovery. Second, it develops a friction taxonomy that connects platform breakdowns to resolution depth. Third, it combines interpretable narrative themes with structured platform information under a strict chronological design. Fourth, it introduces a shrinkage-adjusted company–product benchmark and uses it to identify distinct recovery archetypes rather than producing a simplistic league table.

## 2. CONCEPTUAL POSITION AND RELATED RESEARCH

### 2.1 Digital transformation as a service-system redesign

Digital transformation changes business models, internal coordination, and customer experience simultaneously [1, 2]. Reviews emphasize that successful transformation depends on strategy, structure, technology, leadership, and dynamic capabilities rather than isolated digitization [3, 4, 5]. The digital transformation canvas proposed by [6] similarly frames transformation as an orchestrated process that connects strategic intent, enabling resources, organizational change, and value outcomes.

Within banking, institutional journeys are heterogeneous. [12] describe stages and metrics for financial-institution transformation, while [7] identify organizational and contextual factors affecting banking transformation. Open-finance strategy further requires institutions to convert data access into customer value rather than treating data as an infrastructure asset alone [13]. These perspectives imply that platform performance should be judged through the complete service system, including exception handling and remediation.

Customer experience research supports this extension. Digital banking quality depends on usability, reliability, security, responsiveness, personalization, and trust [14, 15]. Adoption studies show that households do not select FinTech services solely because the services exist; perceived usefulness, convenience, risk, and institutional alternatives shape migration [16]. Recovery is therefore not an ancillary back-office function. It influences whether a user continues to trust the platform after a high-friction event.

### 2.2 Financial platforms, data, and accountability

Financial platforms produce information advantages and new forms of intermediation. Digital footprints can improve credit prediction [17]; FinTech lenders alter screening and borrower composition [18]; and platform-based lending can expand through specialist finance companies [19]. These benefits coexist with concerns about discrimination, unequal model effects, and opaque decision processes [20, 21]. As a result, accountability cannot be separated from platform innovation.

Complaint disclosure is one accountability mechanism. Public complaint information may affect consumer protection and institutional behavior [22], while complaint activity is associated with local trust conditions [23]. From an operational perspective, complaints are a distributed sensor network: each record documents a perceived failure at the boundary between a customer and a financial service. The signal is imperfect because reporting is selective and allegations are not adjudicated, but the aggregate record can reveal recurrent frictions that conventional platform dashboards overlook.

RegTech research has begun to turn this information into decision support. [24] develops explainable and fairness-aware automated monitoring of consumer complaints. [25] applies a residual convolutional architecture to complaint classification, and [26] combines deep language representations with multicriteria decision making for complaint prioritization. These studies establish the value of complaint analytics, yet they focus primarily on classification or priority. They do not benchmark the recovery delivered by a financial platform against the response expected from its observable complaint mix.

### 2.3 From complaint volume to recovery alignment

Three distinctions are required for a defensible platform benchmark. First, *failure incidence* differs from *recovery behavior*. Complaint counts cannot establish which platform fails more often without exposure denominators. Second, *response depth* differs from *complaint merit*. Monetary relief may reflect the institution's policy, evidence, legal obligations, or commercial judgment; explanation may be appropriate for a non-meritorious allegation. Third, *observed response* differs from *case-mix-adjusted response*. A platform receiving more severe or more remediable cases should not be compared with another platform through an unadjusted relief rate.

The framework therefore treats recovery depth as an ordinal operational outcome: explanation is coded 0, non-monetary relief 1, and monetary relief 2. A probabilistic model estimates the expected value of this outcome from intake characteristics. Company–product units are then assessed by the difference between observed and expected recovery depth. This residual approach is familiar in risk adjustment, but its application to digital-finance platform recovery is new.

The study addresses three research questions:

1. Which friction layers and latent narrative themes characterize complaints about digitally delivered financial products?
2. Does text–tabular fusion improve out-of-time prediction of response depth relative to structured intake information alone?
3. Can case-mix-adjusted recovery residuals distinguish meaningful company–product platform archetypes without relying on raw relief rates?

## 3. EVIDENCE ARCHITECTURE

### 3.1 Scope and analytical cohort

The empirical source is the Consumer Financial Protection Bureau Consumer Complaint Database [27]. The frozen extract contains complaints received from 1 January through 31 December 2024 and was assembled before the study cutoff of 31 January 2025. No observation, reference, or empirical claim dated after that boundary enters the manuscript.

The analysis focuses on four products that are strongly mediated through digital platforms: checking or savings accounts; credit cards; money transfer, virtual currency, or money services; and prepaid cards. Records are retained when the company response is a terminal category of closed with explanation, closed with non-monetary relief, or closed with monetary relief. The final analytical cohort contains 21,069 complaints across 316 companies. A narrative is available for 53.6% of observations. Narrative absence is preserved rather than treated as a reason for deletion because structured intake information remains available and disclosure choice may itself be informative.

The design is chronological. Complaints received from January through August form the training period (13,894 records), September and October form the validation and calibration period (3,647 records), and November and December form an untouched test period (3,528 records). The holdout includes 911 relief outcomes, of which 554 are monetary. This sequencing prevents vocabulary, categories, topic structure, and calibration parameters from learning from future complaints.

### 3.2 Response-depth outcome

Let $Y_i \in \{0,1,2\}$ denote the response to complaint $i$, where 0 represents explanation, 1 non-monetary relief, and 2 monetary relief. The expected recovery depth is

$$\widehat{R}_i = \Pr(Y_i = 1 \mid \mathbf{x}_i) + 2\Pr(Y_i = 2 \mid \mathbf{x}_i). \tag{1}$$

This expectation respects the ordered operational depth of the three responses without claiming that monetary relief is always normatively superior. It is a measure of remediation intensity, not legal correctness.

### 3.3 Friction layers

Issue and sub-issue fields are mapped to seven service-friction layers using transparent keyword rules established before test evaluation. *Access and identity* covers login, activation, verification, and legitimate-user access. *Transaction execution* includes transfers, payments, deposits, checks, withdrawals, and incorrect transaction amounts. *Trust and security* includes fraud, scams, unauthorized activity, identity theft, and lost-card events. *Pricing and transparency* covers fees, interest, overdrafts, charges, advertising, and disclosure. *Account lifecycle and support* includes opening, management, closure, investigation, and service interactions. *Data and reporting* covers credit reporting and score disputes. Remaining issues are grouped as other platform friction.

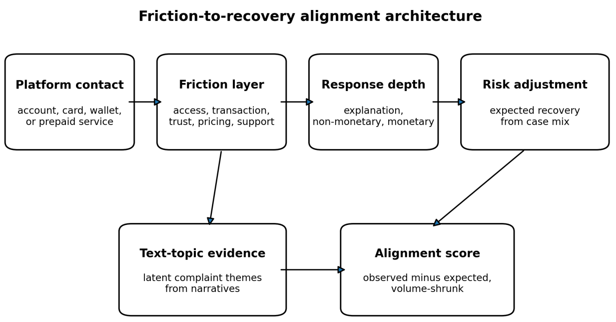
The taxonomy serves two purposes. It translates administrative complaint categories into platform-design domains, and it supports interpretable comparisons of recovery behavior. The rules are included in the analysis code, allowing exact replication or amendment for another jurisdiction.

### 3.4 Narrative themes

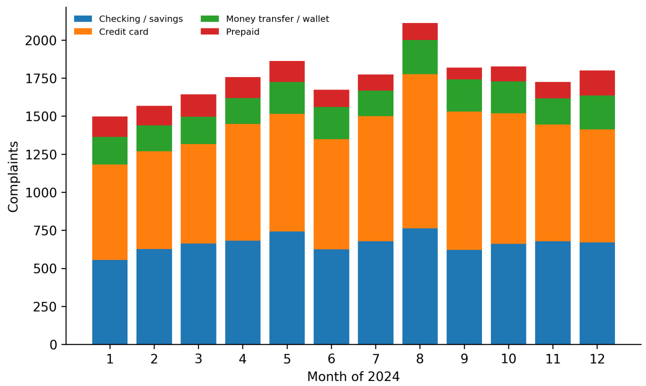
Narratives are cleaned by lowercasing, removing masked identifiers, numbers, punctuation, common stop words, institution-name tokens, and product-name tokens. The training corpus is represented through term frequency–inverse document frequency with unigrams and bigrams, a minimum document frequency of ten, and a 4,500-feature limit.

**Table 1.** Positioning of the study within related research

| Study | Primary focus | Evidence or method | Remaining gap addressed here |
|---|---|---|---|
| [1] | Multidisciplinary digital transformation | Integrative conceptual review | Does not operationalize customer recovery as a transformation outcome. |
| [2] | Strategy and organizational change | Systematic literature review | Limited treatment of post-failure platform performance. |
| [12] | Banking transformation stages and metrics | Financial-institution transformation framework | Metrics do not risk-adjust recovery for complaint case mix. |
| [7] | Drivers of banking transformation | Empirical study of transformation factors | Focuses on transformation determinants rather than service breakdowns. |
| [13] | Data strategy and customer value | FinTech and bank cases under open finance | Does not test whether data-enabled platforms deliver aligned remediation. |
| [14] | Digital-banking customer experience | Systematic review | Identifies service-quality dimensions but not outcome-based recovery benchmarking. |
| [24] | Explainable complaint monitoring | RegTech design and predictive evaluation | Classifies complaints without company–product recovery residuals. |
| [26] | Complaint prioritization | Deep learning and multicriteria decision making | Optimizes priority rather than case-mix-adjusted institutional response. |




(a) Friction-aware modeling and benchmarking architecture.




(b) Monthly complaint mix across the four platform products.

**Figure 1.** Research architecture and temporal evidence. Training, calibration, and final evaluation follow the order in which complaints arrived.

Non-negative matrix factorization extracts eight themes from training records only. Theme weights for validation and test complaints are obtained by projecting them into the fixed training basis.

The themes are labeled from their highest-weight terms: (1) fraud disputes and unauthorized transactions; (2) formal escalation and resolution requests; (3) funds, deposits, and account closure; (4) fees, charges, and balances; (5) identity theft and fraudulent account opening; (6) severe harm and legal escalation; (7) credit reporting and late-payment remarks; and (8) gift-card access and payment declines. The labels summarize word distributions rather than assigning a definitive cause to each complaint.

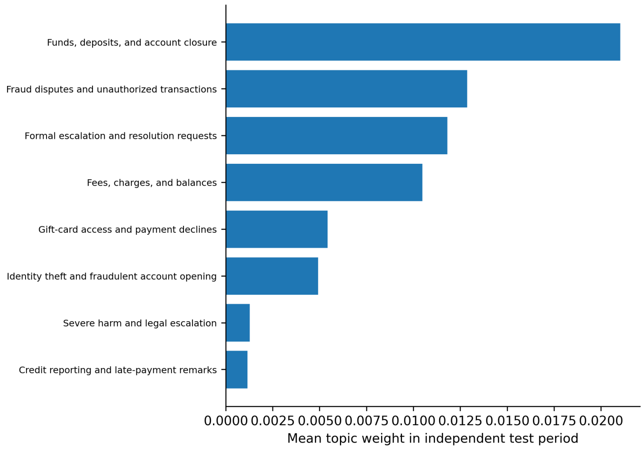
## 4. FRICTION-AWARE EMPIRICAL FRAMEWORK

### 4.1 Competing specifications

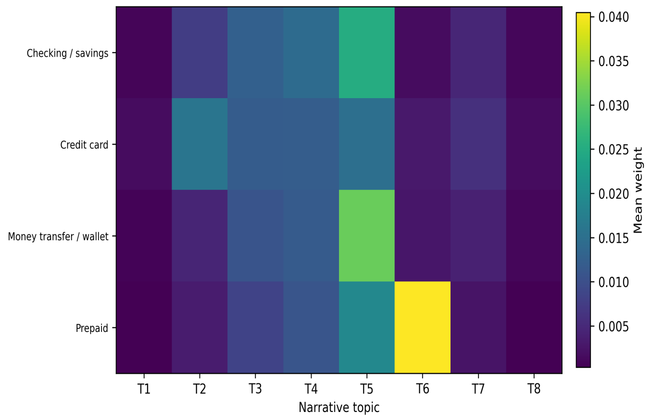
Four models are compared. The *prior* model assigns the training-period class distribution to every complaint. The *structured* model uses product, sub-product, issue, sub-issue, company, consumer tags, submission channel, friction layer, narrative availability, narrative length, receipt day, and weekend status. Infrequent company categories and one-hot levels are pooled to limit overfitting. The *topic profile* model uses only the eight narrative-theme weights. The *platform fusion* model combines structured and topic features.

Each predictive model is a regularized multinomial linear classifier optimized with log loss. Sparse linear estimation is suitable for the high-dimensional categorical design, permits stable chronological retraining, and avoids placing architectural complexity ahead of interpretability. The monetary class receives a larger training weight because it is the least frequent response. Models are calibrated on September–October probabilities through multinomial logistic recalibration. Probability calibration is essential because Equation 1 and the platform benchmark depend on the magnitude of predicted probabilities, not only their rank [28].

Performance is evaluated on November–December complaints through accuracy, balanced accuracy, macro and weighted F1, multiclass log loss, monetary-relief precision–recall area, and any-relief precision–recall area. Precision–recall measures are emphasized because the monetary outcome remains the minority class. Interpretability and calibration are treated as design requirements rather than post-hoc




**(a)** Average theme weight by chronological split.




**(b)** Theme intensity by product in the independent test period.

**Figure 2.** Latent narrative structure. Institution and product names are excluded from the topic vocabulary to keep the themes centered on service friction.

additions [29, 30].

### 4.2 Platform recovery alignment

For company–product unit $g$, observed and expected recovery are

$$\overline{R}_g^{\text{obs}} = \frac{1}{n_g} \sum_{i \in g} Y_i, \tag{2}$$

$$\overline{R}_g^{\text{exp}} = \frac{1}{n_g} \sum_{i \in g} \widehat{R}_i. \tag{3}$$

The raw alignment residual is

$$A_g^{\text{raw}} = \overline{R}_g^{\text{obs}} - \overline{R}_g^{\text{exp}}. \tag{4}$$

Small groups can produce extreme residuals by chance. The reported score therefore applies empirical-Bayes-style shrinkage toward zero:

$$A_g = \frac{n_g}{n_g + \lambda} A_g^{\text{raw}}, \tag{5}$$

where $\lambda = 30$. Descriptive platform comparisons are restricted to units with at least 20 test complaints. Equation 5 does not create a causal estimate. It is an out-of-time diagnostic of whether observed response depth differs from the model-implied case mix.

### 4.3 Recovery archetypes

Adequately observed units are clustered using standardized alignment, relief and monetary rates, shares of trust, access, and transaction friction, and narrative availability. Four-cluster $k$-means with repeated initialization yields interpretable profiles. Clustering is used to organize platform conditions rather than to label institutions as good or bad. The analysis retains company–product granularity because the same institution may operate materially different recovery processes across products.

**Table 2.** Product outcomes in the independent test period

| Product | $n$ | Relief | Monetary | Mean depth | Narrative |
|---|---|---|---|---|---|
| Checking/savings | 1,350 | .216 | .150 | .366 | .587 |
| Credit card | 1,509 | .335 | .170 | .505 | .514 |
| Transfer/wallet | 394 | .124 | .102 | .226 | .652 |
| Prepaid card | 275 | .233 | .204 | .436 | .542 |

## 5. RESULTS

### 5.1 Platform complaint composition

Table 2 shows the independent test composition. Credit cards form the largest group and have the highest overall relief rate, 33.5%. Prepaid-card complaints have the highest monetary-relief rate, 20.4%, while money-transfer, virtual-currency, and money-service complaints have the lowest mean recovery depth. Narrative availability is highest for transfer and wallet complaints, consistent with customers needing free text to describe complex transaction sequences or disputed transfers.

The latent themes remain broadly stable across time, which supports chronological transfer, but product profiles differ. Fraud-dispute language is prominent across transaction-oriented products; identity-theft language is more concentrated where unauthorized account opening is alleged; gift-card access language is localized to prepaid products. This heterogeneity confirms that a single raw complaint count obscures materially different failure mechanisms.

### 5.2 Predictive evidence

Table 3 and Figure 3 compare the four specifications. The fusion model achieves the best macro F1 (0.506), accuracy (0.762), and log loss (0.609). Its monetary PR-AUC is 0.381, compared with 0.373 for structured features and 0.157 for the prevalence baseline. Any-relief PR-AUC rises from 0.519 to 0.525 when themes are added. The incremental gain is modest, but it is consistent across probability-based measures.

The topic-only model performs poorly as a hard classifier and does not exceed the prior on macro F1. This result is substantively important. Narrative themes describe the experienced friction, but the response is also shaped by product rules, issue categories, institution, channel, and other structured context.

Free text should therefore complement rather than replace platform process data. The fusion result supports a restrained interpretation of narrative analytics: it improves discrimination at the margin while structured operational fields carry most of the response signal.

The confusion matrix shows the operational limitation behind the moderate balanced accuracy. Explanation dominates predicted labels, and non-monetary relief is difficult to distinguish from both adjacent categories. The problem is not simply class imbalance: non-monetary recovery covers heterogeneous actions such as corrections, fee reversals without cash payment, or account changes. For platform governance, calibrated expected depth is therefore more useful than treating the predicted class as an automated decision.

### 5.3 Friction and recovery depth

Recovery differs sharply by friction layer (Table 4). Pricing and transparency complaints have the highest mean response depth (0.580) and monetary-relief rate (26.9%). Transaction-execution complaints also show relatively high monetary relief (20.2%), consistent with cases in which a disputed amount, fee, transfer, or payment can be quantified. Trust and security complaints, by contrast, are closed with explanation in 87.7% of cases and have the lowest mean depth among adequately populated layers. A fraud allegation can be severe while still being difficult to resolve through immediate compensation because liability, authentication, merchant evidence, and customer conduct may remain contested.

These contrasts justify case-mix adjustment. A platform concentrated in pricing disputes would mechanically exhibit more monetary relief than one concentrated in security allegations, even if both applied comparable recovery standards. Raw monetary rates would attribute the difference to platform quality; the proposed benchmark assigns part of it to the observable complaint portfolio.

### 5.4 Platform recovery alignment

Forty-four company–product units meet the minimum-volume criterion. Figure 4 reveals substantial dispersion around the observed-equals-expected line. The shrinkage adjustment narrows residuals for smaller units but does not eliminate meaningful separation. Some units deliver response depth materially above their predicted case-mix benchmark, while others fall below it.

The ranking should not be read as a public quality table. First, complaint propensity may differ by customer base and platform design. Second, the data do not contain transaction exposure or verified loss. Third, company responses are recorded categories, not independent judgments of appropriateness. Fourth, the model adjusts only for observed variables. The value of alignment lies in internal investigation: a materially negative product residual can prompt a review of escalation rules, refund authority, evidence collection, hand-offs, or communication design. A positive residual can identify practices worth examining, but not necessarily copying without understanding their costs and context.

Company–product granularity is consequential. The same institution can exhibit different alignment across credit-card, checking, prepaid, or transfer platforms. This pattern supports the view that digital transformation is implemented through product operating models rather than through a single institution-wide maturity level. Governance should therefore trace the specific platform journey that produces the residual.

### 5.5 Recovery archetypes

The cluster analysis organizes the 44 units into four profiles (Figure 5). Nine *recovery leaders* combine positive mean alignment (0.092) with relatively high monetary and overall relief rates. Twenty-seven *recovery-constrained* units have negative mean alignment ($-0.065$), lower relief, and a complaint mix weighted toward transaction and access friction. Five *trust-friction-intensive* units receive a large share of fraud and security complaints and exhibit near-neutral but slightly negative alignment. Three *transaction-friction-intensive* units have high overall relief but almost no monetary response in the test window; their negative alignment indicates that the form of recovery is shallower than the model expects, even where non-monetary corrections are frequent.

The archetypes show why a single transformation score is inadequate. A trust-intensive platform may need stronger investigation transparency and provisional-credit logic; a transaction-intensive platform may require reconciliation and reversal redesign; a constrained platform may need broader authority for frontline recovery; and a leader may need controls to ensure that generous remediation is sustainable and consistently governed. Each profile points to a different transformation intervention.

## 6. PLATFORM DESIGN AND SUPERVISORY IMPLICATIONS

### 6.1 Recovery as a transformation metric

Digital-transformation programs should add recovery alignment to their performance architecture. Traditional indicators remain useful, but they should be complemented by measures of how exceptions are handled. A practical dashboard can track, by product and friction layer, complaint volume, narrative-theme movement, expected recovery depth, observed recovery depth, alignment residuals, and the stability of those residuals over time. Such a dashboard would turn complaints from a lagging reputational count into a diagnostic view of platform operations.

The metric should be embedded in a review process rather than displayed as an unqualified score. Product owners can inspect negative residuals alongside process traces: number of hand-offs, authentication failures, evidence requests, reversal time, provisional-credit decisions, escalation authority, and communication events. This investigation connects the statistical signal to controllable platform design.

### 6.2 Friction-specific transformation priorities

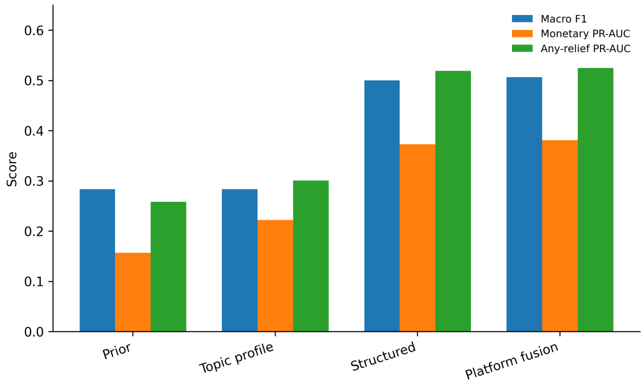
The results indicate that different friction layers demand different design responses.

*Transaction execution* calls for event-level traceability. Customers and staff need a consistent record of authorization, routing, settlement, reversal, and status. Platform integration should reduce contradictory states across the application, ledger, payment processor, and support system.

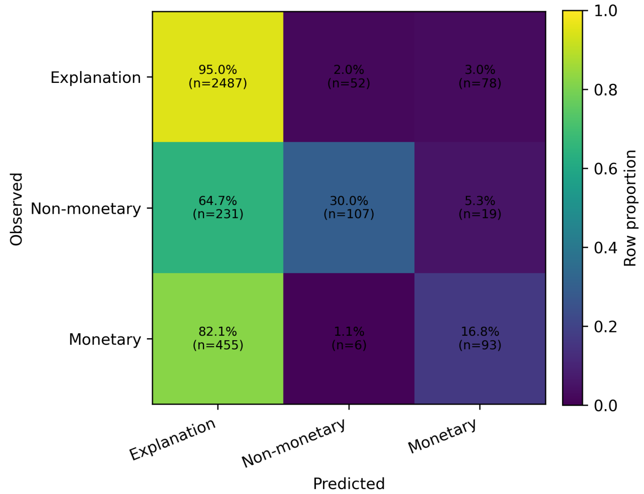
*Pricing and transparency* requires explainable fee and interest logic before and after the transaction. Because these

**Table 3.** Out-of-time model performance on November–December 2024 complaints

| Model | Accuracy | Balanced acc. | Macro F1 | Weighted F1 | Log loss | Monetary PR-AUC | Any-relief PR-AUC |
|---|---|---|---|---|---|---|---|
| Platform fusion | .762 | .473 | .506 | .718 | .609 | .381 | .525 |
| Structured | .758 | .469 | .501 | .714 | .610 | .373 | .519 |
| Topic profile | .742 | .333 | .284 | .632 | .741 | .222 | .301 |
| Prior | .742 | .333 | .284 | .632 | .748 | .157 | .258 |




**(a)** Comparison of general and rare-outcome measures.




**(b)** Confusion matrix for calibrated platform fusion.

**Figure 3.** Out-of-time predictive results. Narrative themes add probability-ranking value, although explanation remains the easiest response to classify.

**Table 4.** Response depth by platform-friction layer

| Friction layer | $n$ | Expl. | Non-mon. | Mon. | Depth |
|---|---|---|---|---|---|
| Transaction execution | 1,217 | .737 | .061 | .202 | .465 |
| Lifecycle/support | 755 | .797 | .086 | .117 | .319 |
| Other friction | 563 | .655 | .298 | .046 | .391 |
| Pricing/transparency | 402 | .689 | .042 | .269 | .580 |
| Access/identity | 402 | .761 | .077 | .162 | .400 |
| Trust/security | 187 | .877 | .011 | .112 | .235 |

**Table 5.** Platform recovery archetypes

| Archetype | Units | Mean $n$ | Alignment | Monetary |
|---|---|---|---|---|
| Recovery leaders | 9 | 84.0 | .092 | .350 |
| Recovery constrained | 27 | 75.8 | $-.065$ | .087 |
| Trust-friction intensive | 5 | 38.4 | $-.018$ | .103 |
| Transaction-friction intensive | 3 | 68.7 | $-.039$ | .000 |

complaints show high monetary recovery, they may reveal preventable design ambiguity or rules that can be corrected upstream.

*Access and identity* requires recovery pathways that preserve security without trapping legitimate customers. Strong authentication is not sufficient if an exception cannot move from automated verification to accountable human review.

*Trust and security* requires evidence-sensitive communication. The low observed recovery depth does not establish under-remediation, but it indicates that explanation dominates. Institutions should test whether explanations describe the investigation, liability rule, and appeal route clearly enough to restore trust.

*Lifecycle and support* points to cross-channel continuity. Closure, restriction, and account-management cases often span multiple systems. Transformation should preserve case context across automated and human interactions rather than forcing customers to reconstruct the history at each hand-off.

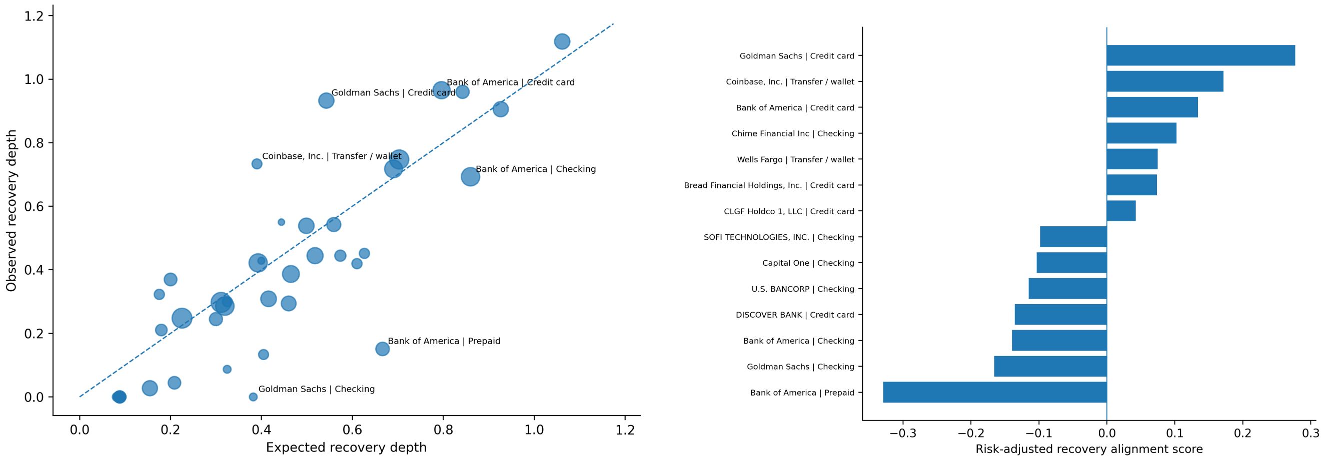
### 6.3 Use in regulatory technology

Supervisors can use the framework to identify outliers for examination, but safeguards are essential. Alignment should be calculated within comparable products and periods; uncertainty and volume thresholds should accompany every score; models should be recalibrated as issue taxonomies or response practices change; and narrative processing should exclude personally identifying information. Most importantly, the score should never determine enforcement or consumer entitlement by itself. Complaint allegations and institutional responses require human interpretation.

The framework is also compatible with explainable RegTech principles [24]. It uses transparent friction rules, low-dimensional themes, calibrated probabilities, and explicit residual formulas. These elements permit a reviewer to understand why expected recovery is high or low. More complex language models could improve prediction, but additional performance would need to justify reduced transparency, greater computational burden, and more difficult temporal governance.

## 7. BOUNDARY CONDITIONS AND ROBUSTNESS CONSIDERATIONS

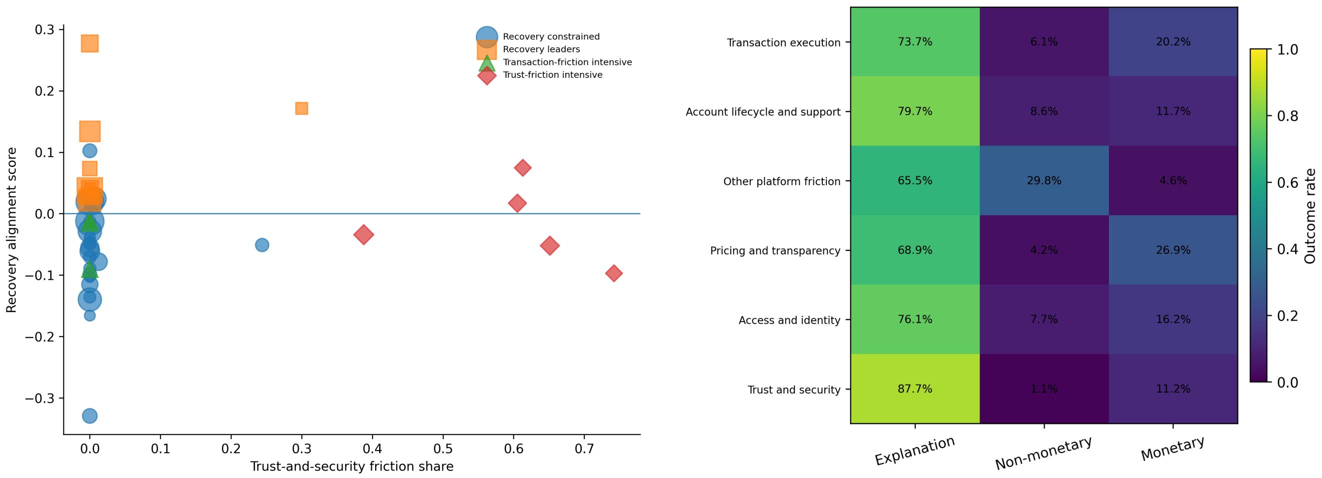
The study has six principal limitations. First, the complaint database is not a random sample of customers. Awareness, trust, severity, language, and willingness to complain influ-




**(a)** Observed versus expected recovery depth by platform unit.

**(b)** Shrinkage-adjusted alignment for adequately observed units.

**Figure 4.** Risk-adjusted recovery alignment. Distance above or below the diagonal reflects deeper or shallower recovery than predicted from observable case mix. Scores are diagnostic and complaint-conditional.




**(a)** Platform units by expected recovery and alignment archetype.

**(b)** Outcome composition across friction layers.

**Figure 5.** Recovery archetypes and friction-specific response patterns. The profiles describe operating conditions rather than institutional merit.

ence selection [23]. Second, exposure denominators are unavailable. The analysis cannot compare failure rates per customer or transaction. Third, response categories are supplied by companies and do not independently verify the validity of either the complaint or the remedy. Fourth, the friction taxonomy is rule-based. Although transparent, it may compress heterogeneous issues or miss context that a manually coded validation study would recover.

Fifth, the model adjusts for observed intake characteristics only. Unmeasured severity, evidence quality, contractual terms, loss amount, customer history, or legal obligations may explain part of the residual. Alignment is therefore associative, not causal. Sixth, the empirical setting covers complaints from one U.S. state during one year. California offers a large and diverse market, but institutional and regulatory conditions may differ elsewhere.

Several design choices reduce, but do not remove, these risks. Chronological separation protects against future leakage. Narrative themes are learned only from training complaints, and brand tokens are removed to prevent themes from becoming proxies for institutions. Probability calibration is performed before the test period. Small-unit residuals are shrunk toward zero, and archetype analysis uses only units with at least 20 test complaints. The complete extract, code, intermediate tables, and figure-generating scripts are supplied for independent audit.

Future work should add platform exposure, confirmed financial loss, resolution time, repeat contact, appeal outcomes, and customer-level satisfaction. Multi-jurisdiction replication could test whether friction-to-recovery relationships are stable under different liability regimes. A longitudinal study could also evaluate whether negative alignment predicts later operational incidents or whether transformation interventions improve alignment without simply increasing compensation.

## 8. CONCLUSION

The financial digital transformation should be defined by the efficiency of a platform in performing standard transactions, but also by how it reacts in case of failure. This paper introduces a friction aware benchmark that associates complaint type and narrative context with the depth of expected response, as well as compares the expectation to the actual response (recovery) at the company-product level.

Evidence backs three conclusions. First, the signal of the platform is structured, while the signal of the narrative themes is interpretable and the relief ranking is calibrated, and the latter signal adds a modest and regular improvement to the former signal. Secondly, there is a friction-specific difference between how pricing complaints are treated and how security complaints are treated. Third, the interpretation of the platform performance is altered by case-mix adjustment. Raw relief rates combine service friction and institutional response, while recovery alignment extracts one of the service residual, which can be used to guide investigation.

The framework is deliberately conservative in nature. It is not a basis for complaint merit, does not imply failure rates of platforms, does not automate enforcement. Its role is to give the financial institutions and the supervisors an observable answer to the more precise transformation question: given the problems that users had, was the platform more or less effective than its observable case mix would suggest? This prompts recovery governance to become a central issue in financial digital transformation.

## DATA AVAILABILITY

The analytical extract, data dictionary, code, intermediate outputs, and figures are included in the reproducibility package. The source data are available from the Consumer Financial Protection Bureau.

## ETHICAL CONSIDERATIONS

The analysis uses publicly released complaint records. Complaint narratives are treated as allegations rather than adjudicated facts. The framework is designed for aggregate diagnosis and human-supervised review, not automated entitlement, enforcement, or adverse action.

## REFERENCES

[1] P. C. Verhoef, T. Broekhuizen, Y. Bart, A. Bhattacharya, J. Q. Dong, N. Fabian, and M. Haenlein, "Digital transformation: A multidisciplinary reflection and research agenda," *Journal of Business Research*, vol. 122, pp. 889–901, 2021.

[2] A. Hanelt, R. Bohnsack, D. Marz, and C. Antunes Marante, "A systematic review of the literature on digital transformation: Insights and implications for strategy and organizational change," *Journal of Management Studies*, vol. 58, no. 5, pp. 1159–1197, 2021.

[3] S. Nadkarni and R. Pruegl, "Digital transformation: A review, synthesis and opportunities for future research," *Management Review Quarterly*, vol. 71, no. 2, pp. 233–341, 2021.

[4] S. Kraus, S. Durst, J. J. Ferreira, P. Veiga, N. Kailer, and A. Weinmann, "Digital transformation in business and management research: An overview of the current status quo," *International Journal of Information Management*, vol. 63, p. 102466, 2022.

[5] D. Plekhanov, H. Franke, and T. H. Netland, "Digital transformation: A review and research agenda," *European Management Journal*, vol. 41, no. 6, pp. 821–844, 2023.

[6] G. Elia, G. Solazzo, A. Lerro, F. Pigni, and C. L. Tucci, "The digital transformation canvas: A conceptual framework for leading the digital transformation process," *Business Horizons*, vol. 67, no. 4, pp. 381–398, 2024.

[7] J. A. Porfirio, J. A. Felicio, and T. Carrilho, "Factors affecting digital transformation in banking," *Journal of Business Research*, vol. 171, p. 114393, 2024.

[8] A. V. Thakor, "Fintech and banking: What do we know?" *Journal of Financial Intermediation*, vol. 41, p. 100833, 2020.

[9] A. W. A. Boot, P. Hoffmann, L. Laeven, and L. Ratnovski, "Fintech: What's old, what's new?" *Journal of Financial Stability*, vol. 53, p. 100836, 2021.

[10] D. Broby, "Financial technology and the future of banking," *Financial Innovation*, vol. 7, p. 47, 2021.

[11] V. Murinde, E. Rizopoulos, and M. Zachariadis, "The impact of the fintech revolution on the future of banking: Opportunities and risks," *International Review of Financial Analysis*, vol. 81, p. 102103, 2022.

[12] A. Papathomas and G. Konteos, "Financial institutions digital transformation: The stages of the journey and business metrics to follow," *Journal of Financial Services Marketing*, vol. 29, no. 2, pp. 590–606, 2024.

[13] L. Grassi, N. Figini, and L. Fedeli, "How does a data strategy enable customer value? the case of fintechs and traditional banks under the open finance framework," *Financial Innovation*, vol. 8, p. 75, 2022.

[14] S. Chauhan, A. Akhtar, and A. Gupta, "Customer experience in digital banking: A review and future research directions," *International Journal of Quality and Service Sciences*, vol. 14, no. 2, pp. 311–348, 2022.

[15] G. Zouari and M. Abdelhedi, "Customer satisfaction in the digital era: Evidence from islamic banking," *Journal of Innovation and Entrepreneurship*, vol. 10, p. 9, 2021.

[16] M. Juenger and M. Mietzner, "Banking goes digital: The adoption of fintech services by german households," *Finance Research Letters*, vol. 34, p. 101260, 2020.

[17] T. Berg, V. Burg, A. Gombovic, and M. Puri, "On the rise of fintechs: Credit scoring using digital footprints," *The Review of Financial Studies*, vol. 33, no. 7, pp. 2845–2897, 2020.

[18] M. Di Maggio and V. Yao, "Fintech borrowers: Lax screening or cream-skimming?" *The Review of Financial Studies*, vol. 34, no. 10, pp. 4565–4618, 2021.

[19] M. Gopal and P. Schnabl, "The rise of finance companies and fintech lenders in small business lending," *The Review of Financial Studies*, vol. 35, no. 11, pp. 4859–4901, 2022.

[20] R. Bartlett, A. Morse, R. Stanton, and N. Wallace, "Consumer-lending discrimination in the fintech era," *Journal of Financial Economics*, vol. 143, no. 1, pp. 30–56, 2022.

[21] A. Fuster, P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther, "Predictably unequal? the effects of machine learning on credit markets," *The Journal of Finance*, vol. 77, no. 1, pp. 5–47, 2022.

[22] Y. Dou and Y. Roh, "Public disclosure and consumer financial protection," *Journal of Financial and Quantitative Analysis*, vol. 59, no. 5, pp. 2164–2198, 2024.

[23] R. M. Hayes, F. Jiang, and Y. Pan, "Voice of the customers: Local trust culture and consumer complaints to the cfpb," *Journal of Accounting Research*, vol. 59, no. 3, pp. 1077–1121, 2021.

[24] M. Siering, "Explainability and fairness of regtech for regulatory enforcement: Automated monitoring of consumer complaints," *Decision Support Systems*, vol. 158, p. 113782, 2022.

[25] D. O. Oyewola, T. O. Omotehinwa, and E. G. Dada, "Consumer complaints of consumer financial protection bureau via two-stage residual one-dimensional convolutional neural network," *Data and Information Management*, vol. 7, no. 4, p. 100046, 2023.

[26] C. Vairetti, I. Aranguiz, S. Maldonado, J. P. Karmy, and A. Leal, "Analytics-driven complaint prioritisation via deep learning and multicriteria decision-making," *European Journal of Operational Research*, vol. 312, no. 3, pp. 1108–1118, 2024.

[27] Consumer Financial Protection Bureau, "Consumer complaint database: 2024 extract," Data set, 2024. [Online]. Available: https://www.consumerfinance.gov/data-research/consumer-complaints/

[28] T. M. Silva Filho, H. Song, M. Perello-Nieto, R. Santos-Rodriguez, M. Kull, and P. A. Flach, "Classifier calibration: A survey on how to assess and improve predicted class probabilities," *Machine Learning*, vol. 112, no. 9, pp. 3211–3260, 2023.

[29] A. Barredo Arrieta, N. Diaz-Rodriguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence: Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information Fusion*, vol. 58, pp. 82–115, 2020.

[30] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova, and C. Zhong, "Interpretable machine learning: Fundamental principles and 10 grand challenges," *Statistics Surveys*, vol. 16, pp. 1–85, 2022.