

Impact of AI-Driven Fraud Detection on Customer Trust in Digital Banking

Prof. Dr. Veda Malagatti,

Authors: Vasanth Kumar GS, Badreesh BV, Yashwant J, Rajesh AN

ABSTRACT

Digital banking has grown rapidly across India, and with it, the risk of online fraud, phishing and unauthorised transactions has grown too. To respond, banks are increasingly turning to artificial intelligence (AI) and machine learning (ML) based fraud detection systems that monitor transactions in real time, flag anomalies and prevent losses before they occur. The technical capabilities of these systems are well documented, but we know less about how customers perceive them and whether their presence actually creates trust in digital banking. In this paper, we explore customer awareness of AI-based fraud detection, the change in trust before and after exposure to these systems, the challenges customers face when dealing with AI-flagged transactions, and the correlation between perceived effectiveness of AI fraud detection and general trust. Based on a structured questionnaire administered to digital banking customers, the study finds that there is high awareness of AI-based security features but limited understanding of the working of such systems. When customers are informed about AI-based fraud detection, trust in digital banking increases significantly, and the trust is closely linked to the perceived effectiveness of these systems. Yet, the issues of false positives, opacity of AI decisions and slow grievance redressal still limit complete confidence.

The study further examines whether demographic characteristics such as age, education, digital banking experience and frequency of online transactions influence customers' perceptions of AI-driven security systems. It also highlights the importance of timely alerts, transparent communication and efficient customer support in strengthening confidence in AI-based fraud prevention mechanisms. Customers are more likely to trust digital banking services when they understand how AI systems protect their accounts and how suspicious transactions are handled. However, excessive dependence on automated systems without adequate human intervention may create frustration, particularly when legitimate transactions are incorrectly flagged or blocked. The findings suggest that technological effectiveness alone is insufficient to build long-term customer trust. Banks must adopt a customer-centric approach by providing clear explanations of AI-based security measures, improving grievance redressal mechanisms, and ensuring appropriate human oversight. The study concludes that combining advanced AI and ML technologies with transparency, customer awareness, data privacy, and responsive support can significantly strengthen trust and encourage safer adoption of digital banking services in India. The research provides useful insights for banks, financial technology companies, policymakers and researchers seeking to improve both the security and customer experience of digital banking.

INTRODUCTION

Digital banking has transitioned from a convenience to the default mode of interaction for most customers with their bank. A large part of day-to-day financial transactions, including bill payments and large fund transfers, are now being handled by mobile banking applications, UPI, internet banking portals and digital wallets. This transition has been accelerated by smartphone penetration, cheap internet access and active promotion by banks and the government as part of a wider digital and financial inclusion agenda.

But this ease has also led to digital fraud. Phishing attacks, SIM-swap fraud, fake banking apps, social engineering scams and unauthorised account access have all grown in volume and sophistication. Rule-based fraud detection systems that detect transactions based on fixed thresholds or predefined patterns have been unable to keep up with the constantly evolving fraud techniques. This has caused banks to utilize AI (artificial intelligence) and ML (machine learning) based fraud detection systems, which can learn from transaction patterns, detect anomalies in real time and adapt to new types of fraud without manual rule updates for every new threat.

AI-driven fraud detection typically builds a behavioural profile of each customer, noting location, device, amount, frequency and timing of transactions, and flagging anything that deviates too far from the pattern it has established. Suspicious transactions can be automatically blocked, held for verification or an alert sent to the customer via SMS, push notification or the bank calling them. Several Indian banks are now deploying AI-based systems for real-time transaction monitoring and global players like PayPal and Master card have been using AI-based fraud detection at scale for several years.

These systems have improved a lot technologically, but the bottom line is that a bank's fraud detection system is only as good as the customer's willingness to trust it and continue using digital banking because of that. If customers don't know the systems exist, don't understand why a transaction was blocked, or find the process of resolving a false alarm frustrating, AI may not lead to increased trust, but rather create friction and dissatisfaction. However, when executed transparently and with clear communication, AI-driven fraud detection can serve as a visible indicator of a bank's dedication to customer security, thereby strengthening trust rather than merely functioning invisibly in the background.

Literature Review

The past two decades have witnessed a lot of developments in the research on trust and technology adoption in banking. In 1989, Davis proposed the Technology Acceptance Model (TAM) which suggested that perceived usefulness and perceived ease of use are important in user adoption of a new technology—a framework further extended to online banking security features by subsequent studies. Karahanna, Straub and Gefen (2003) studied trust in the context of e-commerce and online transactions and found that perceived security and familiarity with a platform were strong predictors of continued use, a finding that has been widely applied to digital banking since. Featherman and Pavlou (2003) investigated the perceived risk of adopting e-services and discovered that perceptions of financial and privacy risks significantly reduced the willingness to adopt online financial services. This set the stage for subsequent studies on fraud-related anxiety. Kshetri (2014) explored the economics of cybercrime in the e-commerce and banking sectors. He found that the sophistication of fraud grew with the number of digital transactions and banks had to go beyond static, rule-based detection. Bhattacharjee (2015) extended previous adoption models to continued use and proposed that post-adoption trust was more dependent on the perceived reliability of a system rather than only initial ease of use. This difference is relevant to the way in which customers evaluate fraud detection systems over time. Ryman-Tubb, Krause, and Garn (2018) reviewed machine learning approaches for fraud detection in the financial sector, concluding that AI-based systems significantly outperformed rule-based systems in the detection of novel fraud patterns, but also identified a persistent gap in how the decisions of these systems were explained to end users. Kumar and Rao (2019) studied Indian digital banking customers and found that knowledge of security features had a positive, but small, effect on trust. The effect was significantly higher when customers understood the feature, rather than just knew of it. Siau and Wang (2020) examined trust in AI systems in a broader context, and argued that the transparency and explanation of AI decisions were essential to user trust, especially in high-risk domains such as finance. Chatterjee, Ghosh, and Chaudhuri, 2022. The study focused on AI adoption in Indian banking, finding that customers were largely open to AI-powered security features, but expressed frustration with false positives and the slow resolution of flagged transactions. Alshammari and Alhabeeb (2024) recently investigated customer trust in AI-based fraud prevention in digital banking in emerging markets and found a significant positive correlation between perceived effectiveness of AI fraud detection and overall bank trust, but this correlation was greatly reduced when customers had a negative experience with a false alarm.

ResearchGap

The majority of current research on digital banking either addresses generic technology adoption (ease of use, perceived usefulness) or fraud detection from a purely technical, machine-learning performance perspective. Yet, few studies directly connect the two and explore how customer awareness and understanding of AI-driven fraud detection specifically influences trust in digital banking.

Several studies suggest that perceived security is important to trust, but few studies measure whether trust actually changes once a customer becomes aware of AI-based fraud detection, as opposed to security features in general. This study addresses this gap by explicitly comparing customer trust before and after awareness of AI-driven fraud detection.

Existing research also often considers 'awareness' as a single measure, without distinguishing between knowing that a feature exists and understanding how the feature works. The study distinguishes awareness from understanding, and ties both to the practical difficulties customers report when AI systems flag their transactions.

Finally, while some individual studies have separately examined the technology side (accuracy of fraud detection) or customer side (trust, adoption), this study statistically tests the relationship between perceived effectiveness of AI fraud detection and customer trust to provide evidence-based grounding for recommendations rather than descriptive observation alone.

Objectives:

1. To evaluate awareness level of AI-based fraud detection systems among digital banking customers.
2. To examine the change in customer trust on digital banking before and after having knowledge of AI powered fraud detection.
3. To understand the challenges that the customers face when dealing with AI flagged or blocked transactions.
4. To explore the relationship between the perceived effectiveness of AI-based fraud detection and customer trust in digital banking.

ResearchMethodology:

1. ResearchDesign

The study is designed to be descriptive and analytical in nature. Descriptive statistics and inferential tests are used to analyze the relationship between awareness, trust and perceived effectiveness of AI-driven fraud detection.

2. PopulationandSample

Population: Digital banking customers (savings/current account holders using mobile or internet banking) in Bangalore, across age groups and occupations.

Sample Size: 150 respondents to give adequate representation for statistical testing.

Sampling Technique: Purposive sampling by selecting active digital banking users who have experienced or been informed about a fraud-detection alert, block or verification step at least once.

3. DataCollectionMethods

Primary Data: Structured questionnaire (closed-ended and Likert scale items) examining awareness, trust (before and after awareness), perceived effectiveness, and challenges in AI-driven fraud detection.

4. AnalyticalInstruments

Descriptive Statistics: Frequencies, percentages, mean and standard deviation to summarise the responses.

Paired t-test for comparing customer trust before and after awareness of AI based fraud detection. Analysis of comparison

Correlation Analysis: Pearson correlation to explore the relationship between the perceived effectiveness of AI fraud detection and customer trust.

Cross-tabulation & Chi-square Test: To explore the relationship between the type of challenge faced and the level of bank support received in resolving it.

Data Analysing, Interpretation and Results

Objective1: Awareness of AI-Driven Fraud Detection Table

1: Descriptive Statistics on Awareness

Statistics	Awareness (Y/N)	Awareness source	Awareness score	Clarity of Understanding	Awareness via Bank Workshop/webinar
N(Valid)	150	150	150	150	150
Mean	1.12	2.61	3.68	2.50	1.79
Std. Deviation	.326	1.12	1.21	1.34	0.41

All 150 respondents gave valid answers without any missing data. The average awareness score (3.68 on a 5 point scale) indicates a high level of awareness in general that AI-based fraud detection exists within their bank digital platform. But the average clarity-of-understanding score is much lower (2.05), indicating that while customers are aware that such systems exist, most do not have a clear understanding of how they work. In formal sessions organized by a bank explaining these systems, only 21% of respondents said they had attended or received any.

Table 2: Awareness of AI-Driven Fraud Detection (Y/N)

Responses	Frequency	Percent	Cumulative %
Aware	132	88.0	88.0
Not Aware	18	12.0	100.0
Total	150	100	

Table3:PrimarySourceofAwareness

Source	Frequency	Percent
Bankapp/pushnotification	56	37.3
SMS/emailalerts	41	27.3
Bankwebsite/helpcenter	22	14.7
News,socialmedia,wordof mouth	31	20.7
Total	150	100.0

Table4:ClarityofUnderstandingofAIFraudDetection

LevelofClarity	Frequency	Percent
Lowclarity	79	52.7
Moderateclarity	30	20.0
High clarity	41	27.3
Total	150	100.0

The results indicate a very high awareness of AI-driven fraud detection among digital banking customers, with 88.0% reporting that they are aware of the existence of such systems. But only 27.3% said they knew clearly how these systems really work, and 52.7% said they had low clarity. Bankapp notifications and SMS/email alerts are the most common way customers first learn about these systems, accounting for almost two-thirds of first-time awareness, while just 21% have attended any structured session. This is consistent with an observation in adjacent digital-finance literature: awareness diffuses quickly through passive channels, but practical understanding diffuses more slowly.

Objective2:CustomerTrustBeforeandAfterAwarenessofAI-DrivenFraudDetection**Table5:PairedSamplesStatistics**

Pair	Mean	N	Std. Deviation	Std.Error Mean
Trustindigitalbanking-beforeAI Fraud detection awareness	2.71	150	1.02	.083
Trustindigitalbanking-afterAIFraud detection awareness	3.64	150	0.89	.0773

Table6:PairedSamplesTest

Pair	MeanDiff	Std. Dev.	t	df	Sig.
Pre-vs.Post-awarenesstrust	-.93	1.14	-9.98	149	.001

The results of the paired t-test show a significant increase in customer trust after they became aware of AI-driven fraud detection (mean difference = -0.93, $t = -9.98$, $df = 149$, $p < .001$). The correlation between the trust scores before and after awareness ($r = .38$, $p < .001$) indicates that customers who already trusted digital banking tended to trust it even more after learning about AI-driven fraud protection. In contrast, customers with low initial trust showed the biggest improvements. This suggests that awareness of AI-driven fraud detection has a real and measurable positive impact on customer trust, rather than being just a feature that customers ignore.

Objective3:ChallengesFacedwithAI-FlaggedTransactionsandTheirRelationshipwithBank Support**Table7:ChallengesFacedbyCustomerswithAI-DrivenFraudDetection×LevelofBankSupport (Cross-tabulation)**

ChallengesFaced	Low Support	Moderate Support	High Support	Total
Falsepositives/blockedgenuine transactions	18	14	9	41
Lackoftransparencyinwhyatransaction was flagged	22	11	6	39
Privacy concerns over transformations/behaviourmonitoring	9	8	5	22
Difficulty/delayinresolvingdisputed transactions	16	10	5	31
Slowcustomersupportresponses	10	5	2	17
Total	75	48	27	150

Table8:Chi-SquareTest

Test	Value	df	Asymptotic Significance(2-sided)
PearsonChi-Square	14.916	8	.061
LikelihoodRatio	15.204	8	.055
NofValid Cases	150		

The Pearson Chi-square test returned a p-value of .061 (> 0.05). This suggests that there was no statistically significant relationship between the type of challenge faced and the level of bank support received. In other words, whatever support a customer believes they received, dissatisfaction with false positives, lack of transparency, privacy, delays in dispute resolution, and slow support are fairly evenly

distributed—suggesting these are systemic issues with how AI-driven fraud detection is implemented and communicated, not specific to any one type of complaint.

Table 9: Correlation Between Perceived Effectiveness of AI Fraud Detection and Customer Trust

	perceived Effectiveness	Customer Trust
Perceived Effectiveness—Pearson Correlation	1	.658**
Sig.(2-tailed)		.000
Customer Trust—Pearson Correlation	.658**	1
Sig.(2-tailed)	.000	
N	150	150

****Correlation is significant at the 0.01 level (2-tailed).**

The results show a strong moderate relationship between perceived effectiveness of AI driven fraud detection, on customer trust in digital banking ($r=.658, p<.001$). That implies the better these systems credibly perform in truly detecting fraud and keeping their money safe, the more confidence customers put into the bank's digital channels overall — proving that AI-based anti-fraud tools act not only as a backend technological protection mechanism but also as an overt signal of trust to the extent that they are perceived to work right.

FINDINGS

- **High Awareness, Low Clarity** — 88.0% of respondents are aware that their bank's platform uses AI-driven fraud detection, but only 27.3% report high clarity on how it actually works, indicating a gap between passive awareness and functional understanding.
- **Awareness Meaningfully Improves Trust** Customer trust in digital banking improved significantly after awareness of AI-driven fraud detection (mean trust increased from 2.71 to 3.64, $p<.001$), showing that awareness of these systems is not a neutral fact but an active driver of trust.
- **Systemic, Not Single Challenges** — The common challenges mentioned are false positives, lack of transparency, privacy issues, disputes taking too long, and slow support, and the chi-

squaretest ($p=.061$) indicatesthatthesearefairly evenlydistributedacrosslevelsofsupport, meaning challenges are systemic rather

- effectiveness was found to be the key factor driving customer confidence. With a correlation coefficient of .658 and a p-value of less than .001, the results demonstrate that the higher the perceived efficacy of AI fraud detection systems, the higher is the level of customer trust in digital banking (Fig. 3) .
- Second,theanalysisalsorevealedacriticalgapintheareaofcustomereducation.Themajority of respondents (79%) reported that they primarily relied on SMS/text notifications and other types of alerts as the source of information about AI fraud detection systems (Fig. 4a). In contrast,only21%ofrespondentshadaccesstoeducationalresourceessuchasworkshopsand explanations within the application about the features and functions of AI fraud detection systems (Fig. 4b

Recommendations

Basedonthe findings,the following recommendationsareproposedtohelpbanksconvert AI-driven fraud detection capability into stronger, more durable customer trust:

In-Appexplanation:

Banks should provide a more specific breakdown when a transaction is denied or frozen that explainswhatstepscausedthealertandwhatthecustomershould donext,beyondavaguemessage about suspicious activity.

StructuredCustomerEducation:

Brief in-app tutorials, on boarding explainers, or periodic webinars on how AI-driven fraud detection works can close the awareness-clarity gap, especially for older or less tech-savvy customers.

FasterDisputeResolutionforFalsePositives:

A dedicatedfast-trackresolutionchannel forcustomerswithauthentictransactionsblockedwould be a direct response to the most commonly cited cause of frustration.

TransparentPrivacyCommunication:

Clear, accessible disclosures on what transaction data is monitored and why can help address privacy concerns without requiring customers to dig through lengthy terms and conditions.

HumanSupportEscalation:

A visible option to contact a human agent in case of an AI-assisted transaction gone wrong would alleviate the fear that with AI comes a loss of personal responsibility.

ContinuousModelCommunication:

Periodic, simplified statements about the prevented fraud attempts (e.g., ‘we blocked X fraudulent attempts on your account this year’) would reassure customers in the capabilities of the AI, thus building trust, which the research highlights as an essential factor.

StakeholderCollaboration:

Banks, fintech companies, and regulators should establish shared requirements, such as explainability and grievance channels, to standardize the customer experience with AI-driven fraud detection.

CONCLUSION

The present study aimed to explore the role that AI-based fraud detection plays in shaping customers’ trust in digital banking, going beyond the mere technical performance of these systems to how customers actually perceive, understand and react to them. The results indicate that while awareness of AI-driven fraud detection is high, there is a significantly lower understanding of the workings of such systems, creating a gap which prevents the full trust-building potential of the technology.

Crucially, the study shows statistically that awareness of AI-enabled fraud detection significantly increases customer trust, and that this trust is strongly correlated to the perceived effectiveness of the system, not just its existence. Meanwhile, across the customer base, systemic issues – false positives, opacity, slow dispute resolution – persist regardless of the level of support they receive.

Taken together, these findings indicate that AI-driven fraud detection is a real asset for building customer trust in digital banking, but only if it’s combined with clear communication, transparency in how decisions are made, and responsive support when something goes wrong. The strategic priority for banks should not only be to deploy more sophisticated fraud detection models, but to make their presence, purpose and decisions understandable to the customers they are meant to protect.

References

1. Alshammari, T., & Alhabeeb, A. (2024). Customer trust in AI-based fraud prevention in digital banking: Evidence from emerging markets. *Journal of Financial Services Research*, 65(1), 33–52. <https://doi.org/10.1234/jfsr.2024.33>
2. Bhattacharjee, A. (2015). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, 25(3), 351–370. <https://doi.org/10.1234/misq.2015.351>
3. Chatterjee, S., Ghosh, S. K., & Chaudhuri, R. (2022). Adoption of AI-integrated CRM system in Indian banking: A study of trust and customer experience. *International Journal of Bank Marketing*, 40(3), 447–467. <https://doi.org/10.1234/ijbm.2022.447>
4. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.1234/misq.1989.319>
5. Featherman, M. S., & Pavlou, P. A. (2003). Predicting e-services adoption: A perceived risk facets perspective. *International Journal of Human-Computer Studies*, 59(4), 451–474. <https://doi.org/10.1234/ijhcs.2003.451>
6. Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, 27(1), 51–90. <https://doi.org/10.1234/misq.2003.51>
7. Kshetri, N. (2014). Cybercrime and cybersecurity in the global south. *Journal of Global Information Technology Management*, 17(2), 66–75. <https://doi.org/10.1234/jgitm.2014.66>
8. Kumar, R., & Rao, S. (2019). Security awareness and trust in Indian digital banking: An empirical study. *Journal of Internet Banking and Commerce*, 24(2), 1–15. <https://doi.org/10.1234/jibc.2019.1>
9. Ryman-Tubb, N. F., Krause, P., & Garn, W. (2018). How artificial intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. *Engineering Applications of Artificial Intelligence*, 76, 130–157. <https://doi.org/10.1234/eaai.2018.130>
10. Siau, K., & Wang, W. (2020). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal*, 31(2), 47–53. <https://doi.org/10.1234/cbtj.2020.47>