SAFWAN ALMODHAYAN



OPINION AND ANALYSIS

Editorial date: September 22, 2026

The Grandmother Flagged by a Fraud System

When Caution Locked the Account

Amelie put her affairs in order before she needed anyone to order them for her. She gave her two grandchildren power of attorney so they could act on her behalf if she lost mental capacity. Then one day the door to her account would not open: her card did not work and online access was locked. She had not lost capacity and was still using the account regularly. But the bank registered the power of attorney and changed the access rights without asking the extra questions that could have revealed the simple fact missing from the procedure: the owner of the money was present, understood what she was doing and wanted to use it. She revoked the power of attorney and supplied proof. The legal record was corrected quickly; the account did not reopen with it. Restoring the card and online access took weeks. With limited mobility, she began taking a taxi to the branch to withdraw £50 in cash at a time. The complaint ended with the UK Financial Ombudsman Service requiring the bank to pay her travel costs and £350 for the distress and delay.1

This is a real case of a grandmother harmed by a system intended to protect authority over an account. Strictly speaking, it is not a case in which an artificial-intelligence model declared her a fraudster, and it should not be described as one. It does, however, open a larger question: how does a procedure designed to protect money become a barrier that keeps its owner from reaching it? Another Ombudsman case concerns Vera, who moved into a care home after

developing dementia. Her daughter Joanne paid the bills under power of attorney. When Joanne attempted a large transfer, the bank's fraud-detection system stopped it as unusual; staff later accepted that it was genuine. The stop also revealed that Joanne was using her mother's login details, contrary to the bank's rules. The Ombudsman considered it legitimate to end the sharing of credentials, but found that the bank had failed to help the attorney use the available mobile alternative. Practical assistance and £200 in compensation followed. This time a fraud-detection system was explicitly part of the story, but the problem was not only whether its suspicion was accurate. It was what happened after the suspicion had been resolved.2

On 28 June 2025, an HSBC customer anonymised by the Ombudsman as M tried to transfer £400 to a friend through the bank's app. The anti-fraud system suspended the payment and sent a message asking him to call. The first voice-verification attempt failed. On the second call, verification succeeded and an employee told him the payment would be released. Meanwhile, the first employee had already begun blocking the account; the second did not know what was happening along the parallel path. The next day M could not enter the app and was told to attend a branch. On 30 June, after sending the money to his friend another way, he left work for half a day and went in person. The Ombudsman did not say that the initial suspicion was itself wrong. It regarded the verification and initial restriction as reasonable; the compensable failure was telling M that the payment would be sent when that was no longer possible. It awarded him £200.3

These bureaucratic details are truer than an easy story about an evil bank and a stupid algorithm. Money was not withheld because a machine legally "accused" its owner of fraud, and the Ombudsman did not find every step unjust. The harm lived in the space between partially correct systems: an engine that noticed an anomaly, a voice check, one employee who changed the account's status, another who could not see the change, and an app that did not explain what had happened. Every component might have passed its individual test with excellent marks while the journey remained terrible. That is the paradox. A fraud system can succeed in raising a flag while the institution fails to lower it once suspicion is resolved. If success is measured at the instant of alert, where are the hours recorded during which an account holder remains standing beneath it?

Big Numbers and a Small Base Rate

Banks are right to use detection systems. Fraud is not a minor edge case that can be left to a pleasant employee with time to examine every payment. A UK Finance report published in May 2025 said British institutions prevented £1.45 billion in unauthorised fraud during 2024, according to industry data, while reported unauthorised-fraud losses reached £722 million across 3.13 million confirmed cases. Customers and institutions also lost £450.7 million to authorised push-payment fraud, in which account holders were deceived into sending the

money themselves. These are grave figures. They are not a comparative test proving that every interrupted transaction deserved to be stopped, nor do they tell us how many legitimate transfers were delayed or how many people had to prove their practical innocence of an automated suspicion.5

The size of the numbers invites an institution to use one measure: “How many crimes did we catch?” This is where a quiet mathematical distortion begins. Imagine one million transactions, of which one thousand are actually fraudulent—one tenth of one per cent. In a numerical illustration describing no particular bank, suppose a model catches 90 per cent of fraud but raises a false alarm on one per cent of legitimate payments. It detects 900 frauds, a triumph on a presentation slide. It also stops 9,990 legitimate transactions. Review staff now face 10,890 alerts, only about 8.26 per cent of them correct. For every true fraud alert, more than eleven legitimate payments are flagged. The conclusion is not that the model is worthless. It is that “catching 90 per cent of evil” tells us less than we imagine about the daily life through which customers must pass.7

This is the base-rate effect in a rare phenomenon. It is easy to read recall and mistake it for the probability that an alert is correct. The latter is precision, which also depends on the prevalence of fraud among transactions. A bank can improve recall by lowering the alert threshold, catching more attempts while sending thousands more genuine customers into review queues. Raise the threshold and some fraudsters pass. In statistics this is the familiar trade-off between false positives and false negatives. Inside a bank it does not remain on paper. It becomes an unpaid bill, a worker absent from a shift, a nurse waiting for wages, and a man trying to persuade his phone that his voice is still his own.

We must also stop the arithmetic itself from lying. The rates in the million-payment illustration are entirely hypothetical. They are neither a global fraud rate nor a measure of HSBC’s performance. “False positive” also requires care: a transaction may prove legitimate after investigation even though a temporary stop was reasonable in light of what was known at the time. A proper assessment asks not only whether risk appeared in the end, but whether the degree of suspicion, the scale of the intervention, the waiting time and the alternatives offered were proportionate to the risk knowable at the moment. Hindsight should not condemn every earlier precaution. Nor should the need for precaution excuse every delay that follows.

Not All Fraud Is the Same Fraud

Part of the confusion comes from the word “fraud.” One payment may be initiated by an attacker who stole card details or took control of an authentication method. Another may be ordered by the genuine account holder after believing a fraudulent story. In the first, a different device, location or usage pattern may be valuable evidence. In the second, the phone can be right, the fingerprint right, and the owner fully conscious; the defect lies in the story

that induced the instruction. A December 2025 joint report from the European Banking Authority and the European Central Bank said strong customer authentication remained effective against the kinds of fraud it was designed to counter, while criminals increasingly manipulated account holders into authenticating fraudulent payments themselves. It reported €4.2 billion in payment fraud in the European Economic Area during 2024. That is a value, not a percentage of the number of payments, and should not be mixed with the hypothetical illustration above.6

If a bank measures transaction risk by asking “does the payer have their original phone?”, it may improve its answer to an old question while fraud moves to another: why are they paying, to whom, and on the strength of what story? UK Finance has documented fraudsters coaxing people into surrendering one-time passcodes and the continued role of communications outside the bank in authorised-payment fraud. Device verification is not useless. It means that an adaptive adversary will try to make a human perform all the correct steps on the adversary’s behalf. In machine-learning terms, concept drift occurs: the relationship between patterns learned by a model and actual fraud changes. Innocent customers also change when they travel, retire, become ill or move country. What is statistically strange may then be ordinary in life.568

What an Algorithm Cannot Learn from Prevention

Research in this field has long moved beyond comparing algorithms by overall accuracy alone. A 2015 IEEE study by Andrea Dal Pozzolo and colleagues described the reality of fraud systems facing torrents of transactions while investigators can examine only a small set of the highest-risk alerts. Early information comes from investigators who contact customers; confirmed labels arrive days later when customers report payments they did not authorise. The samples are not equivalent. Investigators see cases the existing system already selected as risky, while the rest of the stream is not chosen in the same way. The team therefore handled investigator feedback and delayed confirmed labels separately before combining them, and found improved alert precision within its study environment.7

The team’s later IEEE study, published in 2018, analysed more than 75 million real card transactions over three years and examined the effect of class imbalance and changing patterns on system evaluation and updating. We cannot take that dataset and declare that every modern bank has the same rates, or that a model successful for one market, card population and period will perform identically elsewhere. The methodological point is nonetheless vital: a clean training sample assembled after events is not a faithful mirror of a decision made in seconds with incomplete information. Historical test performance cannot settle how a system behaves after fraudsters learn, customers change their habits and the truth about transactions arrives months after the first signal.8

There is a deeper problem I call blindness to the prevented transaction. If a system refuses a transfer before it is sent, it cannot observe what would have happened had it allowed it. Would the money have reached a fraudster and vanished, or a landlord waiting for rent? Prevention appears successful in the first case and as service failure in the second. But the record may never contain the counterfactual truth. Without independent verification, some systems may label a blocked payment “prevented fraud” because the same suspicion rule blocked it. If that label later trains another model, the model can learn to believe its predecessors’ judgments. This is not an accusation against a particular institution. It is a causal-inference problem: a decision changed what could be observed, so the data produced by that decision cannot alone prove it was right.

The problem grows when different risk scores produce different actions. High-scoring transactions receive intense human review, low scores pass unchallenged, and middling cases may go to extra authentication. Sample selection is then itself an outcome of the model, producing selection bias. Train the next algorithm only on cases that reached investigators and the whole world may look fraudulent. Treat every transaction not reported within a week as legitimate and the system may miss victims who discover the loss late or do not know how to object. The answer is not to make humans review everything. It is to create bounded measurement mechanisms and clearly assessable validation samples, while recording uncertainty, the time truth emerged and its source. An investigator is not a decorative button labelled “human review.” The investigator is an instrument for producing better knowledge than the score alone provides.

From Probability to Decision

This is why asking only for higher predictive accuracy is inadequate. An algorithm may output a probability that first needs calibration: a score of 0.8 should, as far as suitable samples allow, correspond to an actual probability rather than merely rank above 0.6. A second decision follows that calibration: what action should a given probability trigger? A high-value payment to a newly seen recipient may reasonably invite additional verification even when the estimated fraud probability is modest. If the payment is a medicine bill whose failure creates immediate harm, calling it unusual is not enough. The bank must consider the cost of interruption and whether a less damaging check is available.

In the simplest decision-theory form, expected cost of allowing a payment is the fraud probability multiplied by the resulting harm. Expected cost of prevention is the probability that the payment is legitimate multiplied by the harm of stopping it, with intervention costs and residual risk added. Reality is more complicated: banks reimburse some losses and not others; stopping a payment may save someone from deception they do not yet perceive; and law sometimes requires action that cannot be reduced to a profit equation. That complexity is

an argument for counting more, not for removing customer cost from the equation. Work on instance-dependent cost-sensitive learning has accordingly examined thresholds that vary with the cost of an error on each transfer, rather than assuming all mistakes are equal.9

The central principle is that reversibility of a protective decision is a security property, not a pleasant addition to user experience. A bank may hold a transfer for an hour while contacting its owner through a trusted channel; that precaution can be resolved. When the hold expands into a frozen account, then blocked digital access, then a phone message that never reaches an employee, the institution has moved from financial-risk management into confiscating time and the ability to act. A July 2024 supervisory report by the US Consumer Financial Protection Bureau documented institutions freezing accounts for suspected activity without notifying some customers or telling them how to restore access. Some dropped or blocked most incoming calls from numbers associated with frozen accounts. Consumers spent weeks or months without access to their money; examiners found some practices unfair.4

The organisation able to block a number in a phone system may have believed it was stopping a fraudster from using another support channel. It also blocked the channel the genuine customer needed to rebut suspicion. The rescue device became a door that opens only for those who do not need it. If a person can learn that an account is frozen but cannot reach anyone who can explain how to unfreeze it, the system becomes a closed loop: access is denied because of suspicion, and objection is denied because access was denied. This needs no advanced AI to occur. Automation magnifies it when restrictions spread across several parts of an institution without one record of the decision's state and the person responsible for ending it.

There is a strong argument on the other side. A customer can acknowledge initiating a payment and insist that it is safe while a scammer is coercing or scripting the answer. The Ombudsman published the case of Evan, whose large payments passed through friends to help another party; the bank believed the circumstances suggested possible deception or coercion. It restricted some transfers while leaving other ways to access cash and meet commitments. The Ombudsman did not uphold the complaint, though it emphasised regular review of the restrictions. The case reminds us that an account holder's truthful statement that they pressed "send" does not settle the safety of the purpose that led them to press it.10

I therefore dislike the easy metaphor of the fraud employee as a meddling guard. That employee may be the last person standing between a victim and the loss of a lifetime's savings. The real problem begins when we merge three claims: "we have reason to ask a question," "we have reason to delay a decision," and "we have reason to remove every option from the customer until further notice." The first may rest on strong data. The second requires a specific procedure. The third demands a higher justification and stronger safeguards. An institution that respects its customers does not leap from a probability on a screen to an open-ended

practical penalty. For fraud based on psychological manipulation, verification must be a dialogue capable of detecting pressure, rather than a script whose answers a criminal can teach the victim.

Harm Is Not Distributed Equally

Returning to Amelie, older people are not “high risk” in only one sense. Some may face greater danger of financial exploitation, while the cost of a mistaken block may also be greater for them. Someone without a car, unfamiliar with complex passwords, or dealing with the bank through a legal attorney may need added protection in one direction and a more accessible service route in another. A rule that treats “low digital activity” as low trustworthiness can turn numerical equality into service bias. A person who rarely travels or transfers money may become intensely suspicious when they need to pay one medical expense. Decline the payment and then require attendance at a branch, and their exceptional need becomes both the evidence against them and the reason they cannot clear it.

Unequal harm does not require a model to contain a variable named “age” or “disability.” Differences can emerge from apparently neutral features: device type, speed of receiving a code, language, calling hours, ability to travel to a branch, and presence of an authorised representative. Equal false-positive rates between two groups do not prove equal treatment if one group regains an account in five minutes and another in five days. Some groups disappear from the data because they abandon the service after failure. Time to lift a restriction, cost of reaching an appeal channel, and success in resolving cases with particular needs must therefore become part of the security test. A defence that repels an attacker but cannot recognise the rightful owner when they arrive remains incomplete.

The Cost-of-Prevention Test

I propose a falsifiable institutional exercise called the cost-of-prevention test. It is neither an existing regulatory standard nor a completed study, but a research framework a bank or supervisor can evaluate and revise. It starts with a representative sample of decisions, not only high-risk alerts. For each transaction, record the risk score at decision time, the type of concern, the action taken, available alternatives, start and end time of the restriction, method of resolution, and final outcome at a known level of confidence. Then estimate, by a declared method, the losses prevention likely avoided alongside the harm it imposed on legitimate transactions: customer and staff hours, direct fees and penalties, and differences in ability to restore service. Not every form of suffering should be forced into one monetary figure, but it should not remain absent from the decision board.

The test needs four kinds of experiment rarely shown in a conventional chart. First, replay decisions on historical records using only what was known at the time, documenting unavailable information. Second, run controlled interventions in a safe environment or on carefully governed samples: compare an alert asking for light authentication with one holding the payment, while monitoring customer harm and the risk of allowing it. Third, simulate an adaptive adversary moving from stolen credentials to deception of the account holder, or changing transfer value and timing. Fourth, test people for whom the first digital channel fails, including attorneys, disabled users and those without a recent device. No one can claim the test succeeded before stating the samples, selection method, customer rights and side effects; experiments should not put real money at risk without suitable controls and approvals.

A concise dashboard would contain numbers seldom gathered in one report: confirmed fraud value prevented; true-alert rate in the highest-risk group sent for review; share of legitimate payments requiring intervention; median and tail time to restore payment capability; share of restrictions lifted on the first trusted contact; cases closed without a clear human decision; recovery-time gaps between user groups; and fraud passing despite controls. Transactions abandoned before verification should be measured, but cannot be called prevented fraud or legitimate activity without further evidence. The indicator is not the count of red flags on a screen. It is the combination of risk removed and rights kept available, under an auditable definition.

When the Bank’s Architecture Changes

This logic may change the bank’s architecture. Instead of a “decline” spilling from one engine into dozens of channels, each intervention would have a unified state record explaining why it was imposed, how long it remains valid, what evidence permits reassessment, who can lift it and what the customer may still do during review. When an account holder proves identity through a trusted channel, that result should move to every relevant path; a second employee should not promise to send a transfer a first employee has already returned to suspension. If fraud conditions change during the wait, the case should be consciously reassessed rather than automatically extended forever. Where rules and risk allow, the bank can distinguish stopping one payment from paralysing the whole account. This is a testable design, not a promise that the hotline will become more sympathetic.

I expect a technical industry narrower and more original than the broad sign reading “AI that stops fraud.” Real value can lie in systems estimating individual harm, policy engines applying graduated measures, infrastructure testing decisions under concept drift, verification systems reusing a trusted result across bank channels, and audit platforms distinguishing prevention of crime from prevention of legitimate activity. Anyone who lays a classifier over old data,

boasts of a high detection rate and ignores alert cost is selling half the science of decision as the whole decision. It is unfair to call a system intelligent if it can stop a transfer in fifty milliseconds, then needs three days to learn that its owner proved the right to make it.

This industry does not erase legal accountability or make every customer the final judge of transaction risk. Anti-money-laundering rules or official orders may impose restrictions different from a fraud-indicator delay, and investigations may prevent disclosure of some information. Rights of appeal and compensation vary with jurisdiction, account and service. “Lift every block after one call” is therefore not a general policy. A narrower, stronger condition is defensible: every restriction should have a responsible owner, specified basis, proportionate degree, lawful review route where rules permit, and documented effect that can be examined after the case ends. Security unable to explain where its authority stopped is liable to extend that authority in proportion to its inability to explain.

When Does the Authority to Stop End?

These stories show, for different reasons, that a banking event does not end when a model produces a probability, an employee presses “block,” or an app displays “for your protection.” It ends when fraud is actually prevented, legitimate action is allowed, or the customer receives an explanation and a proper route to review within the limits available. Many systems learn from a fraudulent transaction that passed, and some learn from a false alarm. An institution seeking durable trust should also learn from the hour lost between them. Money in an account is not a theoretical promise of ownership. Its practical value is that its owner can use it when needed, unless a legitimate and proportionate reason prevents them.

The grandmother in the title is therefore clearer than every classification matrix. She does not ask the bank to abandon caution or automatically believe everyone saying “I own the account.” She asks only that caution not become the end of the story. Behind each risk flag stands someone who may be a victim, complicit, or simply a rightful owner whose behaviour departed this time from yesterday’s data average; each possibility needs a different path. The hardest part of fraud engineering is not teaching a machine to say “stop.” It is building an institution that knows when the authority to stop has ended, and how to tell the right person, after a comprehensible and prompt verification: “Here. This is your money.”

Sources and references

[1] Financial Ombudsman Service, “A bank account is blocked after grandchildren use power of attorney” — Official case study with no visible publication date; accessed 9 November 2026.

[2] Financial Ombudsman Service, “Power of attorney didn’t entitle Joanne to use her mum’s login” — Official case study with no visible publication date; accessed 9 November 2026.

[3] Financial Ombudsman Service, Final Decision DRN-5921701 — Events of 28–30 June 2025; the open decision file does not establish a specific issue date; accessed 9 November 2026.

[4] US Consumer Financial Protection Bureau, Supervisory Highlights, Issue 34, §2.5.1 — July 2024; examination findings generally concern 2023.

[5] UK Finance, Annual Fraud Report 2025, press release — 28 May 2025; 2024 data.

[6] European Banking Authority and European Central Bank, Joint EBA–ECB Report on Payment Fraud — 15 December 2025; 2024 data.

[7] Dal Pozzolo, Boracchi, Caelen, Alippi and Bontempi, "Credit Card Fraud Detection and Concept-Drift Adaptation with Delayed Supervised Information" — IEEE IJCNN 2015; original paper copy on an author's site.

[8] Dal Pozzolo et al., "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy" — Published online 14 September 2017; *IEEE TNNLS* 29(8), August 2018.

[9] Höppner, Baesens, Verbeke and Verdonck, "Instance-dependent Cost-sensitive Learning for Detecting Transfer Fraud" — Early publication 2021; *European Journal of Operational Research* 297(1), 2022, pp. 291–300.

[10] Financial Ombudsman Service, "Evan's bank blocked payments on his current account" — Official case study with no visible publication date; accessed 9 November 2026.