

观点与分析

编辑日期: 2026年9月22日

被反欺诈系统怀疑的祖母

当谨慎关上账户的大门

埃米莉在需要别人替她安排生活之前，已经提前安排好了一切。她授权两个孙辈：若自己失去心智能力，他们可以代为处理事务。然而有一天，她的银行卡失效，网上银行也无法登录。她并未失去行为能力，仍在经常使用账户。银行登记授权书后更改了访问权限，却没有多问几个本可揭示简单事实的问题：钱的主人还在，理解发生了什么，也希望使用自己的钱。她撤销授权并提交证明；法律记录很快恢复，账户的大门却没有随之打开。银行卡和网上访问花了数周才恢复。由于行动不便，她只能乘出租车到网点，每次取50英镑现金。英国金融申诉专员最终要求银行报销交通费，并就困扰和延误赔偿350英镑。1

这是一位祖母被权限保护机制伤害的真实案例；严格说来，并不是人工智能模型宣布她是骗子，不能这样描述。它却适合开启一个更大的问题：保护金钱的程序怎样变成阻止主人使用金钱的障碍？同一机构还记录过维拉的案例。她因失智入住护理院，女儿乔安凭法律授权替她付账。一笔大额转账因异常被欺诈检测系统拦下，员工随后确认交易真实；但审查发现女儿在使用母亲的登录信息，违反银行条款。申诉专员认为停止共享登录信息合理，问题在于银行没有帮助合法代理人从自己手机上使用替代方案。银行最后提供实际协助并赔偿200英镑。这一次，欺诈检测系统确实进入故事；问题不只在怀疑是否准确，也在怀疑消除后发生了什么。2

2025年6月28日，一名汇丰客户——申诉专员隐去姓名，只称他为M——试图在手机应用中向朋友转账400英镑。反欺诈系统暂停交易并发短信要求致电。第一次通话的声纹核验没有通过；第二次通过后，一名员工告诉他转账会放行，而第一名员工已经开始封锁账户。第二名员工不知道另一条流程正在发生。次

日，客户无法登录应用，被要求亲自到店。6月30日，他请了半天假前往网点，此时早已通过其他方式把钱给了朋友。申诉专员并未认定产生怀疑本身错误，而是认为核验与最初封锁均属合理；应赔偿的错误，是工作人员承诺款项会发出，而这已不可能。赔偿为200英镑。3

这些官僚细节比“邪恶银行加愚蠢算法”的简单故事更真实。资金并不是因为机器在法律上“指控”主人欺诈而被冻结，申诉专员也没有裁定每一步都不公。伤害发生在一组各自只对了一部分的系统之间：异常检测引擎、声纹验证、改变账户状态的员工、看不到该改变的另一名员工，以及无法向主人解释经过的应用。每个部件单独测试都可能表现优异，整段旅程却依然糟糕。悖论正在这里：系统可以成功升起红旗，机构却在疑点澄清后不会把它降下来。若成功只在告警一刻计分，账户主人站在旗帜下等待的那些小时又记在哪里？

大数字与小基数

银行使用检测系统有充分理由。欺诈不是可以留给一位耐心员工逐笔处理的小麻烦。UK Finance在2025年5月发布的报告称，根据行业数据，英国机构在2024年阻止了14.5亿英镑未经授权的欺诈，而已报告的未经授权欺诈损失为7.22亿英镑，涉及313万起确认案件；客户和机构还因账户主人受骗后亲自授权的转账欺诈损失4.507亿英镑。这些数字严重，却不是一项比较试验，不能证明每一笔被拦交易都该被拦，也没有告诉我们多少合法转账被延误、多少人不得不向自动怀疑证明自己的清白。5

大数字很容易诱使机构只问：“我们抓住了多少犯罪？”安静的数学偏差从这里开始。假设有100万笔交易，其中1000笔确属欺诈，即数量上的0.1%。再假设——这只是计算示例，不描述任何银行——模型能抓住90%的欺诈，却对1%的正常交易误报警。它会发现900笔欺诈，放在演示文稿上相当耀眼；也会拦下9990笔合法交易。核查团队面对10890条告警，其中真正欺诈只约占8.26%。每一条正确告警背后，有超过十一条来自正常交易。结论不是模型没有价值，而是“抓住九成坏交易”对人们实际经历的日常生活说明得远比想象中少。7

这就是稀有事件中的基率效应（Base Rate）。人们容易把召回率（Recall）误读成“任一告警为真的概率”，后者其实是精确率（Precision），还取决于交易中的欺诈占比。银行可以降低阈值提高召回率，多抓几次尝试，同时把成千上万名真实客户送入复核队列；提高阈值则会放走一些欺诈者。统计上，这是误报（False Positives）与漏报（False Negatives）的熟悉取舍。在银行里，取舍不会留在纸面：它变成未付账单、缺勤工人、未收到工资的护士，以及一个努力说服手机“我的声音仍是我的声音”的人。

但计算本身也不能撒谎。上述100万笔交易的比例完全是假设，不是全球欺诈率，也不衡量汇丰的表现。“误报”还需精确定义：交易调查后可能合法，但在当时已知信息下，暂时拦截仍可能合理。正确评估不只问最终是否存在风险，还问当时可知的怀疑程度、干预强度、等待时长与替代渠道是否相称。事后知识不能把一切先前谨慎都判为错误；同样，谨慎的必要也不能替一切后续延误辩护。

并非所有欺诈都是同一种欺诈

混乱的一部分来自“欺诈”这个词。攻击者可能窃取银行卡资料或控制认证工具后交易；也可能由真正的账户主人在骗局诱导下亲自下令。前者中，设备、地点和使用方式异常可能很有价值；后者中，手机正确、指纹正确，钱的主人也完全清醒，错误藏在促使他发令的故事里。欧洲银行管理局与欧洲央行2025年12月的联合报告指出，强认证对其针对的欺诈类型仍然有效，骗子却越来越多地诱导账户主人自行认证欺诈交易。报告称欧洲经济区2024年已报告支付欺诈金额为42亿欧元。这是金额，不是支付笔数比例，不能与前面的假设例子混用。6

若银行把风险问题理解为“付款人是否拿着原来的手机”，它可能把旧问题答得更好，而欺诈已转向“他为什么付款、付给谁、依据什么故事”。UK Finance记录了骗子诱使人们交出一次性密码，也记录了授权转账欺诈仍依赖银行外部的沟通渠道。这并不使设备核验失去意义，只说明会学习的对手会让人替他执行所有正确步骤。用机器学习术语说，概念漂移（Concept Drift）发生了：模型学到的模式与实际欺诈之间的关系改变。无辜客户的行为也会因旅行、退休、疾病或迁居而改变。统计上的陌生，有时只是生活里的正常。568

算法无法从“已阻止”中学到什么

该领域早已超越只按总体准确率比较算法的阶段。Andrea Dal Pozzolo等人在2015年IEEE会议论文中描述了真实环境：欺诈系统面对交易洪流，调查团队只能检查风险最高的一小部分。早期信息来自联系客户的调查员，确认标签则可能在数日后、客户报告未授权交易时才到达。这两类样本并不相同：调查员看到的交易，本就由旧系统选择为高风险；其余交易没有经历同样的选择。团队因此分别处理调查反馈与延迟确认标签，再组合结果，并在该研究环境中提高了告警精确率。7

该团队2018年又在IEEE期刊发表研究，分析三年间超过7500万笔真实银行卡交易，说明类别不平衡与模式漂移怎样影响评估和更新。不能据此宣称每家现代银行拥有相同比例，也不能说一个在某市场、某卡种和某时期成功的模型，在另一国家同样有效。但方法论结论很重要：事后整理干净的训练样本，不是几秒内、信息不全时作出决策的忠实镜像。历史测试领先，也无法决定当欺诈者从系统中学习、客户改变习惯、真相比怀疑信号晚数月到达时，系统会怎样表现。8

还有一个更深的问题，我称之为“被阻止交易的盲区”。系统若在转账前拒绝，就看不到放行后的反事实：钱会流向骗子后消失，还是到达等待房租的房东？前者让阻止看似成功，后者体现服务损失，但日志也许根本没有反事实真相。若缺少独立核查，一些系统甚至可能依据原来的怀疑规则，把被拦交易登记为“已阻止的欺诈”。这些标签再用于训练新模型，模型就可能学会相信自己的旧判断。这不是对任何具体机构的指控，而是因果推断问题：决策改变了可观察结果，决策产生的数据不能独自证明决策正确。

风险不同，干预也不同，问题随之加深。高分交易接受密集人工审查，低分交易直接通过，中间交易可能被送去追加认证。样本选择本身变成模型的结果，形成选择偏差（Selection Bias）。若下一代算法只在调查员见到的案例上训练，它会觉得整个世界都充满欺诈；若一周内无人报告就一律视为正常，又会

漏掉发现较晚或不知道如何申诉的受害者。解决办法不是让人检查一切，而是设计有限、受控、可明确评估的测量机制与核验样本，记录确定程度、真相出现时间和来源。调查员不是一个装饰性的“人工复核”按钮，而是生产比单一分数更好知识的工具。

从概率走向决策

因此，要求模型“更准确”还不够。算法给出的概率估计首先需要校准（Calibration）：0.8不应只表示排在0.6前面，还应在适当样本上尽可能与实际概率相符。之后仍有一个校准无法替代的决定：在某个概率下该做什么？一笔大额交易发往新收款人时，即使欺诈概率不高，追加核验也可能合理；若交易是到期药费，失败会立刻造成伤害，仅仅说它“不寻常”就不够。还要衡量停止交易的代价，以及是否有更少伤害的核验方式。

最简单的决策分析，会把放行的预期代价——欺诈概率乘以损害——与阻止的预期代价——正常概率乘以阻止造成的损害——比较，再加入干预成本及干预后的残余风险。现实更复杂：有些损失由银行赔、有些不赔；暂停可以保护尚未意识到受骗的客户；法律有时要求采取无法简化为利润公式的措施。但复杂性要求扩大计算范围，而不是把客户成本从公式中删掉。成本敏感学习（Cost-Sensitive Learning）研究因此为不同转账采用不同错误代价与阈值，而非假设每笔交易的错误都等价。9

关键原则是：保护性决定的可逆性（Reversibility）本身是一项安全属性，不只是用户体验上的善意。银行为通过可信渠道联系客户，暂停转账一小时，属于可以迅速判断的谨慎；若暂停变成冻结整个账户，再变成封锁数字访问，最后只剩一条无法触达负责人的电话信息，机构就把交易从金融风险区搬进了剥夺时间与行动能力的区域。美国消费者金融保护局（CFPB）2024年7月的监管报告记录过一些机构因可疑活动冻结账户，却不通知客户或不说明如何恢复，甚至屏蔽与冻结账户关联号码的大多数来电。有些消费者数周或数月无法使用资金，检查人员认定其中部分做法不公平。4

封锁号码的机构也许以为自己在防止骗子借客服渠道绕行，却同时封住真正主人消除怀疑所需的渠道。救援工具变成了一扇只对不需要它的人打开的门。客户知道账户冻结，却找不到能说明解冻办法的人，系统便成为闭环：因怀疑而阻断访问，又因阻断而阻断申诉。这种故障无需高级人工智能也会发生；当自动系统在机构不同角落施加限制，却没有统一记录说明决定状态及其终止责任人时，伤害会被放大。

另一面也有不可忽视的强理由。客户可能承认交易出自自己并坚持合法，却正在骗子压力下照稿回答。英国金融申诉专员公布过“Evan”案例：他的多笔大额交易经朋友转给另一方，银行认为可能存在欺骗或胁迫。银行限制部分转账，但保留其他取现与付款方式；申诉专员不支持投诉，同时强调要定期复核限制。它提醒我们：账户主人承认按下发送键，并不能自动证明导致他按键的目的安全。10

因此，我不喜欢把反欺诈员工描述为多管闲事门卫的简单比喻。他可能是阻止一生积蓄消失的最后一个人。真正问题在于混淆三种主张：“有理由提问”“有理由延迟决定”“有理由无限期剥夺客户的一切选择”。第一项可能有强数据依据，第二项需要明确程序，第三项要求更高理由与更强保障。尊重客户的机构不会从屏幕上的概率直接跳到开放式实际惩罚。面对心理操纵型欺诈，核验还应是能够发现压力的对话，而不是骗子轻易教受害者回答的自动问答。

伤害并不平均分配

回到埃米莉，老年人并不只有一种意义上的“高风险”。一些人更易受金融剥削，同时，错误阻止对他们的代价也可能更高。没有车、不熟悉复杂密码或通过法律代理人与银行往来的人，一方面需要更多保护，另一方面也需要更容易的服务路径。若规则把“较少数字活动”当作不可信，表面的统计平等会变成服务偏差。一个平时不旅行也不转账的人，为治疗突然支付一次费用，反而最可疑；若拒绝后又要求本人到店，我们便把他的特殊需要既当作不利证据，又当作无法洗清怀疑的原因。

伤害不均也不表示模型一定拥有“年龄”或“残障”字段。差异可以来自看似中性的变量：设备、接收验证码的速度、语言、致电时间、能否到店、是否有认可代理。即使两组误报率相等，也不能证明体验公平——一组五分钟恢复账户，另一组要五天。有些群体甚至不再出现在数据里，因为失败后已离开服务。因此，解除限制所需时间、申诉渠道成本，以及特殊需求案例的解决率，都应成为安全测试的一部分。挡住攻击者后仍无法认出亲自前来的权利人，防御便不完整。

“阻止成本”测试

我因此提出一项可被证伪的机构实验，称为“阻止成本测试”。它不是现行监管标准，也不是已完成研究，而是银行或监管方可以评估、修改的研究框架。样本应来自有代表性的决策，而不只是高风险告警。每笔交易记录决策时的风险分数、怀疑类型、实际干预、可用替代方案、限制起止时间、解决渠道，以及带有明确确定度的最终结果。随后按公开方法估计阻止避免的损失，也计算合法交易因阻止遭受的损害、客户与员工工时、直接费用和罚金，以及恢复能力的差异。痛苦不必被强行压成一个金额，但也不能在决策面板上完全缺席。

测试需要四类通常图表不提供的实验。第一，用历史记录重放决策，只使用当时已知信息，并标明未知数据。第二，在安全环境或审慎样本中做干预比较：轻量认证告警与冻结交易告警，同时观察客户损害和放行风险。第三，模拟会适应的对手，让欺诈从窃取登录资料转向诱导账户主人，或改变金额和时间。第四，测试无法通过第一数字渠道的人，包括法律代理人、残障人士和没有现代设备的用户。在说明样本、选择方法、客户权利与副作用之前，不能宣称通过测试；任何让真实资金暴露于风险的试验，也需要恰当控制和批准。

一张简短仪表板应同时显示平时很少出现在同一份报告中的指标：被阻止的已确认欺诈价值；送审最高风险组的告警精确率；合法交易需要干预的比例；恢复支付能力的中位时长和最慢分位时长；首次可信联系后解除限制的比例；没有明确人工决定就关闭的案例比例；不同用户群体的恢复时间差；以及绕过控制的欺诈规模。客户未完成核验就离开的交易也值得测量，但没有额外证据时，既不能算“已阻止欺诈”，也不能算正常交易。真正指标不是屏幕上红旗的数量，而是在可审计定义下，消除了多少风险、保留了多少权利。

当银行架构随之改变

这套逻辑也可能改变银行架构。与其让一个引擎的“拒绝”散向几十条渠道，不如让每次干预拥有统一状态记录：为何施加、有效多久、重新评估需要什么证据、谁有权解除、审查期间客户仍能做什么。账户主人通过可信渠道证明身份后，结果应传至所有相关流程，避免第二名员工仍承诺放行被第一名员工挂起的转账。等待中若欺诈环境改变，应有意识地重评，而非自动无限延长。在风险与规则允许时，银行还可以把拦截单笔交易与瘫痪整个账户分开。这是可验证的设计，不只是一句“热线会更有同理心”的承诺。

这里可能出现一个比“人工智能反欺诈”口号更窄、也更真实的技术产业：估计个人损害的系统、采用分级措施的政策引擎、在模式漂移下测试决策的基础设施、跨银行渠道复用可信核验结果的机制，以及区分“阻止犯罪”和“阻止合法活动”的审计平台。若只把分类器放在旧数据上，再展示高检出率而不计算告警成本，卖的只是半套决策科学。系统若能在五万分之一秒内拦下转账，却要三天才知道主人已证明权利，就不该仅凭速度被称为“智能”。

这种产业不会消除法律问责边界，也不会让客户成为自己交易风险的最终裁判。反洗钱规则或主管机关命令可能施加与欺诈指标引发的支付延迟不同的限制；调查中有些信息也不能披露。申诉和赔偿权还因法域、账户与服务而异。因此，“一次电话后解除所有限制”不能成为普遍政策。能够成立的是更窄而更强的条件：每项限制都应有责任主体、明确依据、相称程度、规则允许时的合法复核路径，以及事后可检查的记录。无法说明权力何时终止的安全系统，容易让权力随解释能力的不足而不断扩张。

停止权何时终止

这些故事原因各异，却共同说明：银行事件不会在模型输出概率时结束，不会在员工按下封锁键时结束，也不会在应用显示“为了保护你”时结束。它在欺诈真正被阻止、合法行动重新获准，或客户在允许范围内收到解释和正确复核路径时才结束。很多系统会从放行的欺诈交易中学习，也有系统从误报中学习；希望建立长期信任的机构，还应从二者之间被浪费的那一小时中学习。账户里的钱不是对所有权的理论承诺；它的实际价值，是主人需要时能够使用，除非存在合法而相称的理由阻止。

这就是为什么标题中的祖母比所有分类矩阵都清楚。她没有要求银行放弃谨慎，也没有要求自动相信每个自称账户主人的人。她只要求谨慎不要成为故事的结尾。每个风险标记背后，都可能是受害者、共犯，或只是这一次偏离昨日数据平均值的权利人；每种可能需要不同路径。反欺诈工程最难的并不是教机器说“停”。更难的是建立一个知道停止权何时结束的机构，并在清晰、快速的核验之后，对正确的人说：“请吧，这是你的钱。”

来源与参考资料

[1] Financial Ombudsman Service, A bank account is blocked after grandchildren use power of attorney — 无可见发布日期的官方案例；查阅于2026年11月9日。

[2] Financial Ombudsman Service, Power of attorney didn’t entitle Joanne to use her mum’s login — 无可见发布日期的官方案例；查阅于2026年11月9日。

[3] Financial Ombudsman Service, Final Decision DRN-5921701 — 事件发生于2025年6月28日至30日；公开决定文件未确认具体发布日期；查阅于2026年11月9日。

[4] US Consumer Financial Protection Bureau, Supervisory Highlights, Issue 34, § 2.5.1 — 2024年7月；检查发现主要来自2023年。

[5] UK Finance, Annual Fraud Report 2025, press release — 2025年5月28日，数据为2024年。

[6] European Banking Authority & European Central Bank, Joint EBA – ECB Report on Payment Fraud — 2025年12月15日，数据为2024年。

[7] Dal Pozzolo, Boracchi, Caelen, Alippi & Bontempi, Credit Card Fraud Detection and Concept-Drift Adaptation with Delayed Supervised Information — IEEE IJCNN 2015原始论文。

[8] Dal Pozzolo et al., Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy — 2017年9月14日在线发表；IEEE TNNLS 29(8)，2018年8月。

[9] Höppner, Baesens, Verbeke & Verdonck, Instance-dependent Cost-sensitive Learning for Detecting Transfer Fraud — 2021年提前发表；European Journal of Operational Research 297(1)，2022，291 – 300页。

[10] Financial Ombudsman Service, Evan's bank blocked payments on his current account — 无可见发布日期的官方案例；查阅于2026年11月9日。