

From exception to decision: payment exception handling and decision traceability in Swiss private banking operations

Hervé Racordon 
Independent Researcher, Miki-cho, Japan

Journal of
Financial
Regulation and
Compliance

Received 27 March 2026
Revised 15 June 2026
Accepted 27 July 2026

Abstract

Purpose – Payment exceptions are governance moments, not merely operational incidents. When human operators resolve flagged payments – increasingly with AI-generated analysis shaping their judgement – the resulting decision must be capable of withstanding regulatory examination, client challenge and audit scrutiny. This paper aims to examine why current governance frameworks fail at that task and propose a structured remedy.

Design/methodology/approach – Conceptual paper grounded in practice-based analysis. The exception taxonomy is built through normative analysis of Swiss regulatory requirements (FINMA, AMLA and BCBS) and payment operations literature. The exception decision record (EDR) framework is derived deductively from four identified governance failure modes.

Findings – Four governance failure modes characterise AI-assisted exception environments: the invisible recommendation, the unexplained override, the undocumented escalation chain and the retrospective rationalisation. Not all of these are equally tractable – the third, involving multi-actor escalation chains, raises implementation difficulties that the other three do not. The EDR – a four-layer instrument comprising exception classification, AI contribution record, human decision record and evidentiary anchor – provides a governance architecture responsive to these failure modes, though its implementation in live institutional environments will require adaptation. Analytical scenarios illustrate its application.

Originality/value – The paper offers two contributions: it theorises payment exception handling as a high-intensity hybrid decision environment with accountability properties distinct from general AI governance contexts; and it extends decision traceability from an abstract governance principle to an operationally implementable instrument grounded in Swiss regulatory and fiduciary obligations.

Keywords Payment exception handling, Decision traceability, AI governance, Operational risk, Private banking, Accountability, FINMA, Compliance

Paper type Research paper

1. Introduction

1.1 The operational reality of payment exceptions

Consider a scenario that plays out dozens of times each day across the operations floors of Swiss private banking institutions. A high-value cross-border payment instruction arrives – a transfer of CHF 2.4 million to a correspondent bank in Singapore. The instruction is complete on its face: beneficiary name, account number, SWIFT code, amount, value date. But the automated processing engine flags it: the beneficiary jurisdiction triggers an enhanced due diligence requirement under the institution's cross-border policy, and the



Declaration of interests: The author declares no competing interests. No external funding was received for this research.

Journal of Financial Regulation
and Compliance
© Emerald Publishing Limited
1358-1988
DOI 10.1108/JFRC-03-2026-0160

transaction amount exceeds the threshold for automatic release. The instruction enters the exception queue.

An operations officer picks it up. She reviews the instruction, consults the client's mandate on file, checks the compliance system and calls the relationship manager for context. Fifteen minutes later, she approves the release – manually, with a brief note in the workflow system: “EDD satisfied, RM confirmed client intent, approved”. The payment executes.

Two years later, a FINMA examination reviews the institution's cross-border payment controls. The examiner asks: what was the due diligence basis for this release? What specific checks were performed? Was the operations officer's judgement informed by any automated analysis? What precisely did the relationship manager confirm? The institution's governance record cannot answer these questions. The note says the decision was made. It does not say how.

This gap – between the operational reality of exception decisions and their documentary reconstruction – is the subject of this paper.

1.2 Why payment exceptions are a governance challenge

Payment processing is not merely a technical or administrative function in private banking. It is a governance-intensive activity at the intersection of client fiduciary obligations, anti-money laundering requirements, cross-border regulatory compliance and operational risk management. A payment instruction in private banking is not a commodity transaction: it may reflect complex client mandates, jurisdictional sensitivities, beneficial ownership structures and investment policy constraints that cannot be automatically processed without contextualised human judgement.

The exception rate in private banking payment operations is structurally higher than in retail or corporate banking precisely because of this complexity. Industry analytical data indicates that a significant proportion of high-value cross-border payments require some degree of manual intervention before execution, with exception rates consistently exceeding those observed in retail or corporate banking payment operations [Celent, 2023; Bank for International Settlements (BIS) Committee on Payments and Market Infrastructures (CPMI), 2022]. Celent's 2023 benchmarking data suggests manual intervention rates in the range of 15%–25% for high-value cross-border instructions at mid-sized private banks, though these figures vary considerably across institutions depending on client base composition, jurisdictional exposure and the maturity of automated pre-screening. The data is indicative, not conclusive; no institution-level statistics are publicly available. Each such intervention is a decision event: a human operator exercises judgement – under operational time pressure, using institutional knowledge, compliance systems and increasingly AI-generated analysis – to determine whether a payment should proceed, be modified, be held or be rejected.

These decision events are governance moments – points at which institutional accountability is either created or, in the absence of adequate documentation, destroyed. FINMA's operational resilience requirements, AML accountability obligations and fiduciary duties all converge on the same human act: the operations officer's judgement call on a flagged payment. And yet this convergence point is, in most institutions, among the least well-governed moments in the entire operational environment.

1.3 The AI inflection point

The governance challenge described above is not new. What makes it urgent is the increasing integration of generative AI and AI-assisted analytical tools into payment exception workflows. Across Swiss and European private banking, operations teams are increasingly

deploying AI-powered systems to support exception triage, compliance pre-screening, anomaly detection and draft exception resolution – functions that were previously performed exclusively through manual judgement and institutional knowledge. While the pace and depth of this deployment varies across institutions, the governance challenge it creates is structurally consistent.

The operational case for AI assistance is not difficult to make. Exception handling is time-pressured, contextually complex and cognitively demanding at scale. AI tools that can pre-screen instructions, surface relevant compliance data and generate preliminary assessments reduce the burden on operations teams dealing with high exception volumes under tight deadlines.

But the integration of AI into exception decision workflows creates a new governance dimension that existing frameworks do not address. When an operations officer approves a flagged payment, the governance question is not merely what she decided, but what she was shown before she decided. If her judgement was materially shaped by an AI-generated risk assessment, a pre-populated resolution recommendation, or an automated compliance check – as is increasingly the case – then the governance record of that decision must capture not only the human output but the AI input. Without this, the accountability chain is structurally incomplete.

The paper's central argument follows from this diagnosis. Payment exception handling needs to be understood as a decision governance challenge, not merely an operational workflow management problem. Decision traceability – meaning the institutional capacity to document, reconstruct and defend the reasoning behind exception decisions – must become a core governance requirement in payment environments where AI tools are shaping the informational basis of human judgement.

1.4 Structure of the paper

The paper proceeds as follows: Section 2 positions the contribution in existing operational risk and AI governance scholarship. Section 3 develops a taxonomy of payment exceptions in private banking, characterising their operational and governance properties. Section 4 analyses the structural features of the exception decision environment and identifies the governance failure modes that AI integration amplifies. Section 5 introduces the Exception Decision Record (EDR) framework as a governance instrument for decision traceability in payment exception handling. Section 6 illustrates the framework through three operational scenarios drawn from Swiss private banking practice. Section 7 develops implications for operations management, compliance architecture and the regulatory trajectory. Section 8 concludes.

The scenarios presented in Section 6 are not empirical case studies but analytical reconstructions designed to illustrate the operational implications of the proposed framework. They are grounded in the regulatory and operational characteristics of Swiss private banking as documented in the literature and regulatory corpus reviewed in Sections 2 and 3, and are intended to demonstrate how the EDR framework operates across qualitatively distinct exception types – not to report observed institutional practice.

1.5 Paper design and approach

This paper adopts a conceptual approach grounded in practice-based analysis. It does not present primary empirical data; instead, it develops its analytical framework through normative analysis of Swiss regulatory requirements, cross-referencing with the operational risk scholarship and payment operations literature reviewed in Section 2, and deductive construction of the governance framework from identified failure modes.

The practitioner orientation of the paper – its engagement with operational specifics, regulatory instruments and implementation constraints – reflects the nature of the governance problem rather than a departure from conceptual rigour. The contribution lies in the theoretical move from general AI accountability principles to the specific properties of hybrid decision environments under time and expertise constraints, a move that requires engagement with operational particulars that more abstract frameworks deliberately elide.

The exception taxonomy presented in Section 3 is derived from three sources: analysis of the regulatory requirements applicable to Swiss private banking payment operations (FINMA, 2023, Circular 2023 / 1; AMLA implementing ordinances; BCBS operational risk standards); synthesis of operational risk and payment governance literature documenting the structural properties of exception-generating workflows; and practice-based reconstruction of exception categories consistent with the operational realities of private banking payment processing as documented in industry analytical publications [Celent, 2023; Bank for International Settlements (BIS) Committee on Payments and Market Infrastructures (CPMI), 2022]. The taxonomy makes no claim to exhaustive empirical coverage; it claims analytical coherence and regulatory grounding.

The EDR framework is constructed deductively: each of its four layers is derived from one or more of the governance failure modes identified in Section 4, and the proportionality matrix in Section 5.3 is calibrated against the regulatory exposure analysis of the exception taxonomy. This approach is consistent with established practice in conceptual governance framework development in the operational risk and AI accountability literatures (Power, 2007; Bovens *et al.*, 2014).

1.5.1 Conceptual design method. Because the paper makes no empirical claim, its construction method warrants a brief statement. The source corpus has three parts: the applicable regulatory instruments (FINMA Circular 2023 / 1; FINMA, 2024, Guidance 08 / 2024; AMLA; FATF Recommendation 16; BCBS operational risk standards; the EU AI Act); the academic literature in Section 2; and industry reports [Celent, 2023; Bank for International Settlements (BIS) Committee on Payments and Market Infrastructures (CPMI), 2022; SWIFT Institute, 2022] for operational characteristics the academic sources do not document.

The derivation proceeds in three steps: the taxonomy (Section 3) classifies exception types against regulatory exposure and decision intensity; the failure modes (Section 4) follow from intersecting that taxonomy with the decision environment analysis; and the EDR layers (Section 5) are derived from the failure modes, with the proportionality matrix calibrated against the taxonomy. Table 1 maps inputs to outputs. The limitations are those of a conceptual design – no empirical validation, analytical rather than observed scenarios, a corpus bounded by what is public – with expert review and institutional pilot identified in Section 7 as the next step.

This paper makes two distinct contributions. Firstly, it theorises payment exception handling as a high-intensity hybrid decision environment – a governance context that reveals accountability gaps invisible to general AI traceability frameworks precisely because it combines time pressure, tacit expertise, escalation complexity and AI-mediated informational asymmetry in ways that general frameworks do not model. Secondly, it proposes the EDR as a governance instrument that extends decision traceability from a general accountability principle into an operationally implementable institutional architecture, grounded in the specific regulatory and fiduciary obligations of Swiss private banking.

Table 1. Conceptual derivation: from sources to framework outputs

Input source	Framework output
Regulatory instruments (FINMA, AMLA, FATF, BCBS, EU AI act)	Governance exposure categories in the exception taxonomy (Section 3)
Payment operations and operational risk literature	Exception trigger types and decision-intensity classification (Section 3)
AI accountability and explainable AI literature	Decision-traceability requirements and Layer 2 capture criteria (Sections 2.2, 5)
Decision environment analysis (time pressure, tacit expertise)	The four governance failure modes (Section 4.3)
Failure-mode analysis	The four EDR layers (Section 5.2)
Decision-intensity assessment across exception types	The proportionality matrix (Section 5.3)

Source(s): Author's own work

2. Literature review and theoretical positioning

2.1 *Operational risk and payment exception management*

The operational risk literature establishes that payment exceptions are a recognised source of operational loss, reputational exposure and regulatory sanction. The Basel Committee's operational risk framework [Basel Committee on Banking Supervision (BCBS), 2021] categorises exceptions arising from inadequate processes, system failures and human factors as core operational risk events, subject to loss event documentation, root cause analysis and remediation tracking. This framing – which has shaped institutional practice for two decades – treats exceptions as failure events to be measured and reduced.

The payment operations and financial services operations literatures extend this view by characterising exceptions as workflow disruptions requiring exception management systems: queuing mechanisms, resolution protocols, escalation pathways and STP recovery metrics [Bank for International Settlements (BIS) Committee on Payments and Market Infrastructures (CPMI), 2022; SWIFT Institute, 2022]. Empirical work on operational process quality in financial services has documented the relationship between exception volume, manual intervention rates and operational risk exposure (Cummings, 2012; Chorafas, 2001). More recent work on automation and human-system interfaces in regulated environments has highlighted the governance implications of decision-support tools that augment rather than replace human judgement – a line of inquiry directly relevant to the AI-assisted exception environments this paper examines.

The governance content of the exception decision – what was decided, on what informational basis, through what reasoning process – receives almost no systematic attention in either the operational risk or the payment operations literatures. It is this dimension that the present paper addresses.

2.2 *AI accountability, human oversight and explainable AI*

The concept of decision traceability in AI-assisted environments builds on two established bodies of scholarship that have not previously been systematically integrated. The accountability literature – originating in Bovens' (2007) relational conception of accountability as the obligation to explain and justify decisions to a forum with the authority to evaluate them – establishes that accountability requires not merely the existence of a decision-maker but the demonstrable connection between decision outcomes and the reasoning that produced them. Roberts' (1991) distinction between individualising and

socialising forms of accountability further establishes that AI systems, which distribute decision influence across human and algorithmic actors, create structural misalignments with accountability frameworks that presuppose individual human responsibility.

Three strands are drawn together here: the relational account of accountability originating with Bovens (2007); the algorithmic accountability literature that extends it to AI-mediated decisions; and the explainable AI scholarship that bears on whether AI contributions can be captured in a form fit for governance. The aim is to specify the requirement that Layer 2 of the EDR is later designed to meet.

The algorithmic accountability literature (Diakopoulos, 2016; Kroll *et al.*, 2017; Mittelstadt *et al.*, 2016) extends this framework to AI-mediated decision systems, establishing that meaningful accountability for algorithmic decisions requires institutional structures capable of reconstructing the decision pathway – not merely auditing model outputs. Wachter *et al.* (2017) grounded this requirement in the GDPR's right to explanation, while Doshi-Velez and Kim (2017) addressed the interpretability conditions under which AI contributions to decisions can be assessed. What these contributions share is a recognition that technical explainability of AI models is insufficient for institutional accountability purposes: the governance requirement is to document how AI outputs were incorporated into human decisions, not only how AI outputs were generated.

A distinction from the explainable AI literature is useful here. What Layer 2 requires is explainability of AI outputs to the operator who acts on them – structured confidence indicators, data sources and rationales an auditor can later reconstruct – not interpretability of the model to its developers (Rudin, 2019; Lipton, 2018). Whether deployed systems produce outputs of this form varies by system class and vendor; the framework specifies the requirement and leaves that as an empirical matter (Doshi-Velez and Kim, 2017).

Decision traceability, as developed in this paper, designates the institutional capacity to document, reconstruct and govern the full decision pathway – including AI-generated inputs, human engagement with those inputs and the evidentiary record linking them to authorised outcomes. This framing extends Bovens' accountability dimensions to the specific conditions of hybrid human-AI decision environments, where the informational basis of human judgement is partially constituted by AI analysis. The sociotechnical governance literature reinforces this framing: Selbst *et al.* (2019) demonstrated that AI governance cannot be achieved at the level of the model alone but requires governance of the full decision environment in which model outputs are used – an insight whose implications for accountability design remain underdeveloped in financial services contexts. Prior conceptual work (Author, 2025; Author, 2026a) has formalised this as a governance architecture; the present paper extends it by identifying a property of payment exception environments that the general framework does not capture.

That property – which constitutes this paper's standalone theoretical contribution, independent of its application to payment operations – is the governance asymmetry between decision velocity and documentation capacity. General accounts of AI accountability in hybrid decision environments treat documentation as a tractable governance requirement: they specify what must be documented without analysing the conditions under which contemporaneous documentation is structurally suppressed. Payment exception handling makes those conditions visible. The combination of operational time pressure, high exception volume and tacit expert judgement creates an environment where documentation quality degrades precisely at the decision moments of highest governance significance. AI integration intensifies this asymmetry rather than creating it: when AI outputs are incorporated into time-pressured expert judgments, they are systematically more likely to disappear from the governance record the more reliable and experienced the operator. This

inversion – where governance record quality is inversely correlated with decision quality – is the structural problem the EDR framework addresses. It is also, this paper argues, a theoretical property of hybrid AI decision environments under time and expertise constraints that the existing accountability literature has not adequately modelled.

The hybrid human-AI decision environment literature (Cummings, 2012; Santoni de Sio and van den Hoven, 2018) establishes that meaningful human oversight in AI-assisted decision contexts requires more than the formal existence of a human decision-maker. It requires that human actors genuinely engage with AI outputs rather than routinely ratifying them, that the informational basis of AI contributions is transparent, and that institutional accountability structures capture the interaction between human judgement and algorithmic input. In payment exception contexts, these requirements translate into specific documentation obligations that existing governance frameworks do not systematically impose.

2.3 *Financial crime, anti-money laundering and sanctions case management*

AML and sanctions case management are the existing regime closest to decision-level documentation: Anti-Money Laundering Act (AMLA) and Financial Action Task Force (FATF) obligations require recording the factual basis for clearing or escalating a flagged transaction (Demetis, 2018; Maxwell and Vannieuwenhuyse, 2018). But it applies only to the financial-crime subset, and was designed around human analysis with no structured provision for the AI-generated inputs now informing adjudication. It is a precedent for decision-level documentation, not an instrument built for AI-assisted decisions across the full range of exceptions.

2.4 *Regulatory governance of AI and operational resilience*

Swiss private banking payment operations are governed by a layered regulatory framework. FINMA Circular 2023 / 1 on operational risks and resilience establishes the institutional requirements for process documentation, control evidence and operational risk management. The AMLA and its implementing ordinances impose specific documentation requirements on enhanced due diligence decisions, including the requirement to document the factual basis for releasing transactions subject to heightened scrutiny.

FINMA Guidance 08 / 2024 on artificial intelligence at supervised institutions identifies five governance areas for AI deployment, including accountability and responsibility, traceability of model development and the proportionate governance of AI risk. These are substantive expectations, and the Guidance is the most developed AI supervisory instrument in Swiss financial regulation. It establishes how AI systems must be developed, validated, monitored and documented. What it does not specify is how their outputs are captured at the moment they shape human judgments in live workflows: its traceability requirement attaches to the model, not the decision. That decision layer is the gap the EDR targets.

At the international level, the FATF Recommendation 16 on wire transfers and its Swiss implementation through the Swiss National Bank's payment system requirements impose correspondent banking due diligence obligations that generate exception conditions requiring documented resolution. The FSB's 2024 report on AI in financial services explicitly notes the governance challenge posed by AI-mediated decision processes in compliance-sensitive workflows, without prescribing operational solutions (FSB, 2024).

The regulatory architecture thus creates clear accountability obligations for payment exception decisions – particularly in AML, cross-border, and fiduciary contexts – without providing operational governance instruments adequate to the AI-assisted decision environments in which those decisions are increasingly made.

Across the four streams the same gap recurs: operational risk treats exceptions as loss events, not decisions to be evidenced; the accountability and XAI literatures specify what traceability requires but not how to capture it under operational pressure; the financial-crime regime documents only its own subset, without separating AI from human contribution; and the regulatory instruments presuppose decision-layer fulfilment they do not operationalise. A decision-level instrument spanning the full range of payment exceptions is what this paper sets out to provide.

3. A taxonomy of payment exceptions in Swiss private banking

3.1 The exception landscape

Not all payment exceptions are equivalent in their operational properties or governance implications. An exception arising from a missing beneficiary SWIFT code is resolved through a simple data correction; it involves minimal decision content and generates limited governance exposure. An exception arising from a cross-border compliance flag on a transfer to a high-risk jurisdiction, by contrast, involves substantive human judgement about AML risk, client mandate scope and regulatory compliance – judgement that is potentially subject to regulatory scrutiny, client dispute or ex post accountability review.

Treating these exceptions as equivalent for governance purposes – as most current exception management frameworks do – produces systematic governance under-investment in the decision-intensive cases and disproportionate documentation burden in the routine ones. A governance framework adequate to the AI-assisted exception environment requires a taxonomy that distinguishes exception types by their governance properties.

The following taxonomy is developed from analysis of Swiss private banking payment operations practice, cross-referenced with FINMA operational risk guidance and AML documentation requirements. It classifies payment exceptions across two dimensions: exception trigger (the operational reason for non-automated processing) and decision intensity (the degree of substantive human judgement required for resolution). “The resulting taxonomy is presented in [Table 2](#)”.

3.2 Exception taxonomy

The governance implications of this taxonomy are significant. High decision-intensity exceptions – sanctions hits, cross-border compliance flags, client mandate ambiguities and regulatory holds – constitute a minority of exception volume in most institutions but carry the overwhelming share of governance exposure. These are the exceptions where the adequacy of the human decision, the sufficiency of the informational basis and the completeness of the accountability record are subject to regulatory examination, legal challenge or client dispute.

High-intensity exceptions are also where AI-assisted analysis is most actively deployed – and where its output is most likely to shape the human decision. The governance challenge identified in this paper is sharpest precisely in this segment.

3.3 The AI integration dimension

AI-assisted tools are being integrated into payment exception workflows at the triage and pre-screening stages – observable in vendor product developments, regulatory guidance attentiveness to AI deployment and industry publications over the past three years. The precise rate and depth of this integration across Swiss private banking institutions is less clearly documented, and likely varies substantially between systemically important banks with dedicated AI governance programmes and smaller private banks still assessing

Table 2. Taxonomy of payment exceptions in Swiss private banking

Exception type	Trigger category	Decision intensity	Primary governance exposure
Data deficiency	Missing or malformed instruction fields	Low	STP failure, client communication
Format non-compliance	SWIFT / SEPA formatting errors	Low	Technical – minimal accountability exposure
Sanctions screening hit	SDN / EU / HM treasury list match	High	AML, regulatory reporting obligation
Cross-border compliance flag	Jurisdictional risk indicator, EDD trigger	High	AML, cross-border documentation, FINMA scrutiny
Client mandate ambiguity	Instruction inconsistent with mandate on file	High	Fiduciary, client dispute, MiFID/LSFin suitability
Threshold override	Amount exceeds automated release limit	Medium-High	4-eyes principle, audit trail requirement
Correspondent bank query	Correspondent requests additional information	Medium	External accountability, delay exposure
Duplicate detection	Potential duplicate instruction identified	Medium	Operational risk, client communication
Liquidity / timing constraint	Insufficient funds or timing conflict	Low-Medium	Operational – limited governance exposure
Regulatory hold	Court order, regulatory freeze, embargo	High	Legal, regulatory – formal documentation required

Source(s): Author's own work

deployment cases. The paper's governance argument does not depend on a specific adoption rate. It rests on the simpler observation that wherever AI-generated analysis is used in exception workflows, its contribution to human decisions requires governance infrastructure that most institutions have not yet built – not because AI assistance is problematic, but because its operational integration has consistently outpaced the governance design needed to make it accountable.

The tools themselves range considerably in sophistication. Natural language processing systems extract and structure instruction content; sanctions screening engines generate probabilistic match scores with explanatory rationales; compliance analytics platforms surface client-specific risk indicators and historical exception patterns. In more advanced deployments, generative AI produces pre-populated resolution recommendations – synthesised assessments that present the operator with a suggested course of action rather than raw data. Each configuration presents a governance challenge of different character and intensity, and one acknowledged limitation of the EDR framework is that it does not differentiate between these deployment types in technically precise terms. That differentiation is a direction for future research.

From a governance perspective, however, each layer of AI assistance creates an additional dimension of accountability complexity. When an operator approves the release of a cross-border payment flagged for EDD, the governance record must capture not only her approval decision but the AI-generated analysis she reviewed, the compliance data surfaced by the screening tool and the specific aspects of the AI recommendation she accepted, modified, or overrode. Without this, the accountability record is formally present but substantively incomplete – it documents that a human decided, but not how the human was informed.

4. Structural features of the exception decision environment

4.1 Time pressure and cognitive load

Payment exception handling occurs under operational time pressure that systematically constrains the quality and depth of human decision-making. Value date commitments, correspondent banking cut-off times and client expectations create temporal boundaries within which exceptions must be resolved – typically within hours, often within minutes for high-priority transfers. Operations teams managing exception queues of 30 to 50 items simultaneously face cognitive load conditions that increase the reliance on heuristics, institutional knowledge and – increasingly – AI-generated pre-assessments.

This is consistent with the human factors literature: documentation quality degrades as cognitive load rises (Wickens, 2008), and operators relying on automated support accept its output with diminishing scrutiny as it proves reliable (Parasuraman and Manzey, 2010). Records produced under pressure are therefore unlikely to reflect the operator's actual engagement with AI inputs, even when her decision is sound – a structural feature, not a lapse in diligence (Coglianese and Lehr, 2017).

This operational reality has a direct governance implication. The quality of the documentation produced in high-pressure, high-volume exception environments is systematically lower than would be produced in reflective conditions. Brief notes, standard-form entries and checkbox approvals substitute for the substantive records that accountability requirements demand. The governance failure is not individual negligence: it is a structural feature of the decision environment that exception management systems have never been designed to address.

4.2 The expertise-documentation inversion

A systematic pattern in exception handling governance deserves explicit attention: what may be termed the expertise-documentation inversion. The tacit knowledge literature establishes that expert practitioners embed their knowledge in rapid, non-articulated judgement that resists explication (Polanyi, 1966; Dreyfus and Dreyfus, 1986). Applied to documentation behaviour in compliance environments, this implies that experienced operations officers – those with the deepest contextual knowledge and the most reliable judgement – are precisely the least likely to produce detailed documentation of their reasoning. Their expertise manifests as immediacy of judgement; the decision is experienced as obvious, not as a proposition requiring written justification.

Less experienced operators, by contrast, may produce more detailed documentation precisely because they are less certain of their judgement and more likely to record the steps they followed. The governance record therefore systematically under-represents the reasoning of the most expert decision-makers and over-represents the process compliance of the least expert.

AI assistance amplifies this inversion. When an experienced operator receives an AI-generated exception summary that aligns with her own assessment, she is likely to confirm the recommendation with minimal additional documentation – her expertise tells her the AI is right, and she proceeds accordingly. The governance record captures her confirmation, not her judgement. What the AI recommended, why she agreed, and what she would have decided in the AI's absence are invisible to the accountability record. In practice, the operator who produces the least documentation is often the one making the best decision: seniority and documentation quality can move in opposite directions. Existing governance frameworks were not designed to capture this configuration, and the sections that follow develop one that is.

4.3 Four governance failure modes

Drawing on the exception taxonomy and decision environment analysis, four governance failure modes can be identified in AI-assisted payment exception handling. The taxonomy is presented with some tidiness that the operational reality does not always provide: in practice, failure modes overlap, stack, and compound in ways that resist clean categorisation. The four-mode structure is analytically useful, not exhaustive.

The four modes follow from intersecting the exception taxonomy (Section 3) with the decision environment analysis (Sections 4.1–4.2). Each marks a point where the record loses information reconstruction later needs: the AI input that shaped a judgement, the reason for departing from it, the distribution of judgement across actors and the decay of the record over time. They are neither mutually exclusive nor claimed to be exhaustive – only sufficient to ground the framework in Section 5.

4.3.1 Failure mode 1 – the invisible recommendation. An AI system generates a resolution recommendation. The operator, under time pressure and consistent with her own assessment, accepts it. The governance record reflects that she approved the payment – but contains no trace of what the AI recommended, on what basis, or at what confidence level. If a regulatory examiner later asks what analysis preceded the approval, the institution has no answer.

4.3.2 Failure mode 2 – the unexplained override. Here the operator disagrees with the AI. She has contextual knowledge – a long-standing client relationship, a recent conversation with the relationship manager, awareness of a historical precedent – that the AI system could not access. She overrides the recommendation and approves the payment. The governance record notes that an override occurred. It captures nothing of why. The accountability record is, in formal terms, complete; in substantive terms, it is empty.

4.3.3 Failure mode 3 – the undocumented escalation chain. A high-intensity exception moves up the chain: operations officer to compliance team, compliance team to head of operations, head of operations to relationship manager. AI-generated analysis is circulated at each handoff. Each actor forms a judgement – partial, sequential, interdependent. The final authorisation is recorded. The decision contributions of the intermediate actors, and the AI inputs each of them received, are not.

4.3.4 Failure mode 4 – the retrospective rationalisation. Two years after the decision, an audit or regulatory examination raises questions. The operations team reconstructs a rationale. The account is coherent, plausible, and consistent with good practice. It is also, almost certainly, partly fictional – constructed from institutional memory and professional inference rather than from contemporaneous record. The AI inputs that shaped the original decision are gone. The reconstruction proceeds as if they never existed.

5. The exception decision record: a governance framework

5.1 Framework design principles

The EDR framework operationalises key principles of the integrated governance model (Author, 2025) in the specific context of payment exception handling. It is proposed as a structured governance instrument for decision traceability in AI-assisted payment exception handling – not a new documentation system but a governance architecture: a set of design principles and information requirements that exception management workflows must satisfy to produce accountability records adequate to the AI-assisted decision environment.

The EDR framework is grounded in decision traceability theory and operational risk governance scholarship. The essential departure from conventional audit trail approaches requires some precision, because the claim that the EDR is genuinely different from existing

systems – rather than merely better labelled – is central to the paper’s contribution and warrants direct engagement.

Four existing documentation mechanisms are relevant comparators. Audit logs capture system-level events: who accessed what, when, with what outcome. They are designed for IT security and access control purposes, not decision governance. Workflow case management notes – the standard exception resolution record in private banking operations platforms – capture the human decision outcome and, at best, a brief free-text rationale. They were not designed to record AI-generated inputs and have no structured field for doing so. Maker-checker trails record dual authorisation: they document that two people approved a decision, not what either of them knew when they approved it. Model governance logs, increasingly required under AI governance frameworks including FINMA Guidance 08 / 2024, document AI model development, validation and performance – they operate at the system level, not the decision level. None of these mechanisms, individually or in combination, produces a record that links a specific AI output, at a specific moment, to the specific human judgement it preceded.

The EDR’s design principles – contemporaneity, proportionality, completeness and reconstructability – are not design preferences but responses to specific failures in existing documentation practice. Contemporaneity addresses retrospective rationalisation. Proportionality addresses documentation burden asymmetry. Completeness addresses the invisible recommendation and unexplained override failure modes. Reconstructability addresses evidentiary durability. The structure follows from the failure modes; the apparent symmetry is a product of the analysis rather than its starting point. Whether these principles can be operationalised within existing exception management infrastructure, or require new systems investment, is an implementation question the framework deliberately leaves open. The answer will vary by institutional context, and not all institutions will find one that allows full implementation within existing constraints.

In operational terms:

- Contemporaneity: documentation must be generated at the time of the decision, not reconstructed after it. AI-generated inputs must be captured at the moment they are presented to the human operator, not summarised retrospectively.
- Proportionality: documentation intensity must be calibrated to decision intensity. Low-intensity exceptions – data deficiency, format non-compliance – require minimal governance records. High-intensity exceptions – sanctions hits, cross-border compliance flags, client mandate ambiguities – require comprehensive accountability documentation.
- Completeness: the governance record must capture both the AI contribution and the human judgement. An approval record that documents only the human decision without the AI input that preceded it is structurally incomplete for accountability purposes.
- Reconstructability: the governance record must enable a supervisor, auditor, or regulatory examiner who was not present at the time of the decision to reconstruct the decision pathway – what information was available, what analysis was provided, what judgement was applied and why.

These four principles can be stated as design propositions, which makes the framework testable proposition by proposition rather than as a whole:

-
- P1. High decision-intensity exceptions require contemporaneous capture of AI inputs at the point they are presented to the operator; retrospective reconstruction is inadequate for these categories.
 - P2. Documentation intensity must be proportional to governance exposure, so that routine categories carry minimal records and high-intensity categories carry the full requirement.
 - P3. Multi-actor escalation chains require stage-specific records of both the AI contribution presented and the human judgement applied at each handoff; a single end-of-chain authorisation is incomplete.
 - P4. Evidentiary anchoring must be tamper-evident but technology-neutral; the durability requirement is mandatory for high-intensity exceptions, while the specific mechanism is an institutional choice.
-

A clarification on what the EDR is. It is a governance architecture – a set of requirements AI-assisted exception systems must satisfy – operationalised through design specifications, leaving open whether existing platforms can be modified or new infrastructure is needed. The author is aware of no commercially deployed platform that captures all four layers as specified; the framework makes that gap explicit without claiming no system could be adapted towards it.

5.1.1 What the exception decision record adds beyond existing documentation mechanisms. Institutions already maintain audit logs, workflow case notes, maker-checker trails, AML case management records, model governance documentation and operational risk event registers. It is fair to ask whether the EDR is a relabelling of these. It is not, and the distinction matters for the contribution claimed here.

Audit logs record system access – who did what, when – not the informational basis of a decision. Workflow case notes capture what the operator chose to write, not what she was shown, with no structured field for AI inputs; their quality depends on operator time, so the records most needed are the least likely to be substantive. Maker-checker trails show that two people authorised a decision, not what either knew. AML and sanctions case records [AMLA, [Financial Action Task Force \(FATF\), 2023](#)] are the closest analogue but cover only that subset and do not separate AI from human analysis. Model governance documentation (FINMA Guidance 08 / 2024, EU AI Act) records system properties at the model level, not the decision. Operational risk registers are retrospective loss records, selective by threshold.

The narrower novelty claim follows. None of these mechanisms, alone or together, produces a decision-level record that binds a specific AI output to the operator who received it, the engagement that followed, and a durable evidentiary chain. The EDR does not displace existing documentation; it occupies a layer the others leave empty.

5.2 EDR structure: the four-layer model

The EDR framework organises the decision record into layers, each corresponding to a distinct element of the exception decision pathway. The layered structure reflects the sequential logic of exception handling – classification precedes AI analysis, which precedes human judgement, which requires evidentiary anchoring – rather than a preference for architectural symmetry: “The four-layer EDR structure is presented in [Table 3](#)”.

5.3 Proportionality calibration

“[Table 4](#) presents the EDR proportionality matrix across exception categories”. The EDR framework applies proportionality through a decision-intensity matrix that maps exception

Table 3. The four-layer EDR framework

EDR layer	Content	Purpose	Governance function
Layer 1 Exception Classification	Exception type (per taxonomy), trigger source, flag rationale, initial risk score, STP failure reason	Establishes the governance starting point: what triggered the exception and why automated processing failed	Proportionality calibration: determines documentation intensity for subsequent layers
Layer 2 AI Contribution Record	AI system identifier and version, analysis generated (summary, risk assessment, recommendation), confidence indicators, data sources consulted, timestamp of AI output delivery; model owner; prompt or input query submitted; documents retrieved; limitations or caveats disclosed to the operator; applicable policy version; confirmation that the operator viewed the output	Creates a durable, timestamped record of the AI-generated inputs presented to the human operator at the time of the decision	Closes the invisible recommendation failure mode; enables reconstruction of AI contribution
Layer 3 Human Decision Record	Operator identity and role, decision outcome (approve / modify / hold / reject / escalate), engagement record (modifications to AI recommendation, additional information requested, counterparty consulted), decision rationale; decision type classified as acceptance of the AI recommendation, modification, override, escalation or independent decision reached after AI review; rationale captured in structured rather than free-text form	Documents the human judgement applied to the exception, including the operator's engagement with AI recommendations	Closes the unexplained override failure mode; establishes personal accountability
Layer 4 Evidentiary Anchor	Timestamp chain, authorisation sequence (including 4-eyes records where required), document references, escalation log, immutable record identifier; retention period; access rights and access audit trail; legal hold flag; tamper-evidence controls (hashing, cryptographic timestamping, or distributed ledger anchoring); cross-system linkage references to payment, KYC, sanctions and relationship manager systems	Creates a tamper-resistant record of the complete decision pathway, linked to the authorisation infrastructure	Supervisory defensibility; audit evidence quality; temporal integrity

Source(s): Author's own work

Table 4. EDR proportionality matrix

Exception category	Required EDR layers	AI contribution capture	Evidentiary anchor required
Sanctions screening hit	All four layers (full)	Full capture – mandatory	Yes – immutable record
Cross-border compliance flag	All four layers (full)	Full capture – mandatory	Yes – immutable record
Client mandate ambiguity	All four layers (full)	Full capture – mandatory	Yes – immutable record
Regulatory hold	All four layers (full)	Full capture – mandatory	Yes – immutable record
Threshold override	Layers 1, 2, 3, 4	Abbreviated capture	Yes – standard audit trail
Correspondent bank query	Layers 1, 3, 4	Optional capture	Standard audit trail
Duplicate detection	Layers 1, 3, 4	Optional capture	Standard audit trail
Data deficiency	Layers 1, 3	Not required	Standard workflow log
Format non-compliance	Layer 1 only	Not required	System log sufficient
Source(s): Author's own work			

types to documentation requirements. High-intensity exceptions require all four EDR layers to be fully populated; medium-intensity exceptions require Layers 1, 3 and 4 with abbreviated Layer 2 capture; low-intensity exceptions require Layer 1 and minimal Layer 3 records only.

5.4 Technical infrastructure implications

The EDR framework does not prescribe a specific technical architecture for its implementation. The governance requirements it articulates are compatible with a range of technical approaches – from enhanced exception management workflow systems to more sophisticated AI-integrated governance platforms that automatically capture AI outputs at the point of delivery and anchor them to the human decision record (Author, 2025).

What the framework does require is that exception management systems be designed with Layer 2 capture – AI contribution recording – as a first-class function rather than an afterthought. Current exception management platforms in Swiss private banking were designed before AI-assisted analysis was integrated into exception workflows; their audit trail functionality captures human actions but has no mechanism for recording the AI-generated inputs that precede those actions. Addressing this gap requires deliberate infrastructure investment, not merely policy enhancement.

The evidentiary durability requirement of Layer 4 implies that EDRs for high-intensity exceptions must be stored in a manner that precludes retrospective alteration – either through immutable logging systems, cryptographic timestamping, or distributed ledger mechanisms that provide tamper-evident audit trails. This aligns with the validation layer of the integrated governance model, which emphasises tamper-evident governance infrastructures as a structural governance requirement, not a technical option (Author, 2025). The choice of technical implementation is an institutional decision; the governance principle of evidentiary durability is not optional for exceptions with material regulatory or fiduciary exposure.

The four-layer architecture is therefore intended to do something existing documentation practices do not. Existing approaches record authorisation; the EDR records the pathway to authorisation. For high-intensity exception categories – sanctions hits, cross-border compliance flags, mandate ambiguities – that gap matters. In practice, an institution either has the record or it does not.

6. Illustrative scenarios

6.1 Scenario 1 – cross-border payment flag and EDD clearance

A high-net-worth client instructs a wire transfer of CHF 850,000 to a counterparty in a jurisdiction identified by the institution's compliance system as requiring enhanced due diligence under FINMA's cross-border payment policy. The automated processing engine flags the instruction for manual review and routes it to the compliance-enabled exception queue.

The AI-assisted screening tool generates a structured exception summary: the counterparty jurisdiction is classified as elevated risk under the institution's current country risk matrix; the transaction amount exceeds the standard EDD threshold by 70%; the beneficiary name generates a partial match with a politically exposed person (PEP) database entry – insufficient confidence for an automatic hold but sufficient for a mandatory human review flag. The AI tool suggests: "EDD review required before release. PEP match probability: 34%. Recommend obtaining confirmatory documentation from relationship manager and updating counterparty risk classification".

The operations officer reviews the AI summary, consults the client file, contacts the relationship manager and obtains a written confirmation that the beneficiary is the client's

wholly-owned holding company – a matter the AI system could not have known from its available data. She approves the release.

The governance challenge in this scenario is reconstruction: two years later, under supervisory examination, the institution must demonstrate not only that the payment was approved but that the approval was grounded in adequate due diligence. Under the EDR framework, the classification record (Layer 1) captures the cross-border flag and jurisdiction risk score at the time of processing. Critically, Layer 2 preserves the complete AI-generated exception summary – including the PEP match probability and the recommended course of action – timestamped at the moment it was presented to the operator. This is the record that establishes what the operator was shown before she decided. Layer 3 then documents her engagement with that analysis: the additional information she obtained, the basis on which she assessed the PEP concern as resolved, and her approval rationale. The Layer 4 anchor ties the relationship manager's written confirmation to the decision record with an immutable timestamp chain.

The harder case sits just below the surface. Had the relationship manager's confirmation been verbal rather than written, the Layer 4 anchor would rest on the operator's own note of the call – the kind of fragile, self-generated record the framework is meant to displace. The EDR narrows the difficulty rather than dissolving it: it requires the verbal confirmation to be logged contemporaneously and attributed.

The resulting record is qualitatively different from the single-line note it replaces. A regulatory examiner reviewing it can follow the decision pathway from initial flag to approved release, with the AI contribution and the human response both visible and attributable.

6.2 Scenario 2 – client mandate ambiguity and AI override

A discretionary portfolio management client's investment guidelines specify that foreign currency transfers above CHF 500,000 require an explicit written instruction from the client. A payment instruction of CHF 620,000 in USD arrives via the relationship manager's email – embedded in a longer message about portfolio rebalancing and not in the standard instruction format.

The AI-assisted instruction processing system extracts the payment instruction from the email, classifies it as a foreign currency transfer, identifies the amount threshold trigger and flags the instruction for exception review. The AI system notes: 'Instruction received via non-standard channel. Format does not comply with mandate documentation requirements. Recommend requesting formal written instruction before processing.' Confidence: High.

The operations officer disagrees. She recognises that the instruction is embedded in a message from a senior relationship manager who is known to the team and that the client in question has previously provided instructions in this format without dispute. She judges that the AI recommendation, while formally correct, is inconsistent with established practice for this client and that requesting a formal instruction would create unnecessary friction. She approves processing.

Scenario 2 raises a governance question that is structurally distinct from the EDD case. There is no sanctions or AML exposure here – the regulatory stakes are lower, but the fiduciary ones are real. If a client later disputes that the USD transfer was authorised, or claims the instruction format was inadequate, the institution's position depends entirely on its ability to show that the non-standard format was consciously considered, not overlooked. That is a harder evidentiary case to make without contemporaneous documentation.

Under the EDR framework, Layer 2 preserves the AI recommendation *verbatim* – high-confidence classification, specific mandate provision flagged, recommended course of

action. Layer 3 requires the operator to document her override rationale: the established client precedent, the relationship manager's identity, her judgement that the instruction's substance satisfied the mandate requirement. Layer 4 creates an immutable record marking the departure from AI recommendation. Together, these records convert a tacit professional judgement into a documentable institutional act.

Where this tests the framework hardest is what the override rests on. The operator may be right precisely because she holds tacit knowledge of the client's established practice that the institution has never recorded. Layer 3 can capture that she invoked precedent; it cannot, alone, establish the precedent was real absent any prior record. The framework converts tacit judgement into a documented assertion – an improvement, but not a manufactured evidentiary history.

The override here may well be the right call. The operator's judgement may be better than the AI's precisely because she has contextual knowledge the system lacked. That is not the problem the EDR addresses. The problem is that without the framework, her reasoning stays in her head – and when scrutiny comes, months or years later, there is no record that reasoning existed at all.

6.3 Scenario 3 – sanctions screening and escalation chain

An instruction for an international transfer generates a partial match against the EU Consolidated Sanctions List. The sanctions screening engine generates a match report: the beneficiary name shares three of four name components with a listed entity; jurisdiction overlap exists; confidence score: 58%. The instruction is automatically held and routed to the compliance officer for review.

The compliance officer reviews the match report, consults the AI-assisted KYC database and determines that the beneficiary is a different legal entity from the listed party – the name similarity is coincidental, the jurisdictions diverge at the subsidiary level and the beneficial ownership structure clearly distinguishes the two entities. She clears the sanctions match but escalates the transaction to the head of compliance for a second opinion, as internal policy requires for any sanctions flag regardless of outcome.

The head of compliance reviews the compliance officer's analysis and the AI match report, concurs with the clearance determination and authorises the release. The entire escalation chain involves three decision-makers, two AI-generated analytical outputs (the initial match report and a supplementary KYC summary) and a four-eyes authorisation.

Applying the EDR framework to this scenario requires it to accommodate a multi-actor decision chain rather than a single operator's judgement. The compliance officer's initial clearance analysis constitutes a Layer 3 record in its own right, linked to the AI match report (Layer 2) that initiated the review. When the transaction is escalated, the supplementary KYC analysis generates a second Layer 2 record – distinct from the first because it was generated at a different stage and presented to a different decision-maker. The head of compliance's concurrence, with its own Layer 3 rationale, closes the chain and Layer 4 binds the entire sequence into a single timestamped decision record with four-eyes authorisation captured at the relevant stage.

The undocumented escalation chain failure mode identified in Section 4 is, in practice, most common in sanctions-related exceptions precisely because the multi-actor review structure creates multiple handoff points where AI contributions can fall out of the governance record. Without EDR-structured documentation, each stage of the escalation produces its own partial record; the full decision pathway is reconstructable only through effort, and the AI contributions at each stage are likely to be unrecorded. The EDR

framework transforms what is structurally a fragmented accountability trail into a coherent and defensible record.

7. Implications

7.1 *For operations management*

The EDR framework has direct implications for how exception management workflows are designed, staffed and monitored. Operations managers face the challenge of implementing proportional governance documentation without increasing the cognitive and temporal burden on operators already working under high time pressure. Three operational design principles follow from the framework.

Firstly, AI contribution capture must be automated, not manual. Requiring operators to document AI inputs adds documentation burden at the point of highest operational pressure. Exception management systems must be designed to automatically capture and record AI outputs at the moment of presentation, creating Layer 2 records without operator action. This is a system design requirement, not a training or policy requirement.

Secondly, Layer 3 documentation must be structured, not free-text. Free-text notes are the standard governance record in current exception management practice; they are inadequate for the AI-assisted decision environment because they do not systematically capture the engagement between operator judgement and AI recommendation. Structured Layer 3 forms – with specific fields for AI recommendation engagement, override rationale and additional information obtained – produce governance records that are both faster to complete and more substantively informative than free-text alternatives.

Thirdly, exception governance quality should be monitored as an operational metric. Exception resolution time and exception rate are standard operational KPIs; EDR completeness and quality should be added to this monitoring framework, with particular attention to high-intensity exception categories where governance failure exposure is greatest.

7.2 *Implementation challenges*

The EDR framework is a governance architecture, not an off-the-shelf product. Institutions seeking to implement it will encounter obstacles that the framework's design principles do not resolve. These deserve direct treatment rather than relegation to a future research agenda.

The first is cost. Layer 2 capture – automated recording of AI outputs at point of delivery – requires modification of existing exception management platforms. In institutions where these platforms are vendor-managed or deeply embedded in broader workflow infrastructure, the cost and timeline of such modifications may be substantial. A 2023 Celent analysis of exception processing infrastructure in European private banking noted that legacy workflow systems at many mid-sized institutions lack extensible APIs suitable for AI output capture, implying either costly platform upgrades or workaround solutions that undermine the contemporaneity principle. The framework does not resolve this tension; it identifies the requirement and leaves the implementation path to institutional determination.

The second is internal resistance. Operations teams asked to adopt structured Layer 3 documentation will reasonably ask whether enhanced governance requirements extend processing time. In high-volume exception environments, even a modest per-exception documentation overhead compounds across the day. The proportionality principle is designed to minimise this burden for low-intensity exceptions; but high-intensity exceptions – precisely the ones under greatest governance pressure – carry the full documentation requirement. The honest answer is that the EDR framework adds some documentation burden at the decision moments where operators are already under the most time pressure. This tradeoff is defensible, but it should be acknowledged rather than obscured.

One qualification on proportionality deserves statement. The intensity calibration in Table 4 is set by the institution, but a supervisor may not accept that calibration: an examiner may regard a category the institution has treated as medium-intensity as warranting fuller documentation. Institutional calibration is therefore best understood as a defensible baseline, open to supervisory challenge, rather than a unilateral determination of what governance the decision requires.

Vendor-supplied AI complicates this. Where tools are third-party, the institution may lack direct access to the outputs, prompt histories or source-data lineage that Layer 2 presupposes; some platforms expose these through APIs, others only through proprietary interfaces. The framework treats Layer 2 capture as a governance obligation that vendor contracts should secure. Where a vendor will not provide that access, the resulting gap is one the institution must document and own.

The third concerns multi-actor escalation chains – the governance scenario corresponding to Failure Mode 3. In practice, escalation decisions in Swiss private banking often involve informal communication channels: a phone call to the relationship manager, a corridor exchange with the compliance officer, a message through an internal chat platform. The EDR framework requires that these interactions be captured in the Layer 3 escalation log. Whether operators and their counterparts will consistently document informal consultations – particularly under time pressure and without clear supervisory consequence for non-documentation – is a behavioural question the framework cannot answer. The risk is that the escalation log becomes a post-hoc formality rather than a contemporaneous record, which would replicate rather than resolve the retrospective rationalisation failure mode. This tension has no clean resolution within a governance architecture alone; it requires cultural change and supervisory attention.

Finally, the framework has not been empirically tested in live institutional deployment. The scenarios presented in Section 6 are analytical constructs; they illustrate the framework's logic under specified conditions, not observed behaviour under operational constraints. How the EDR performs when operators are interrupted, when AI systems produce ambiguous or conflicting outputs, when escalation chains fragment under organisational pressure, or when institutions prioritise transaction throughput over documentation completeness – these are open questions. The broader context is that several institutions in the Swiss and European private banking space are beginning to engage with structured AI contribution logging in exception workflows, though without published accounts; this suggests the governance gap the framework addresses is recognised in practice, even if the proposed solution remains untested. Formal pilot implementation would substantially strengthen the evidential base.

The absence of empirical validation raises an obvious question: why adopt an unvalidated framework? The answer is not that the EDR is proven, but that the alternative – waiting for a validated solution before acting – is itself a governance choice with costs. Institutions operating AI-assisted exception workflows today are already accumulating accountability gaps. The EDR framework is proposed not as a final answer but as a structured starting point: a governance architecture coherent enough to implement partially, test iteratively and refine against operational reality. It should also be noted that the framework may fail in high-volume environments where exception throughput pressure is severe: proportionality mitigates this risk for routine categories, but does not eliminate it for the high-intensity exceptions that matter most.

7.3 For compliance architecture

Compliance functions in Swiss private banking face a governance gap that FINMA Guidance 08 / 2024, for all its importance, does not close. The Guidance establishes model

governance requirements; it does not establish decision governance requirements for AI-assisted operational workflows. The EDR framework addresses this gap at the institutional level, providing a governance instrument that compliance functions can implement in advance of regulatory crystallisation.

There is a proactive case for EDR implementation that goes beyond compliance risk avoidance. Institutions that build decision traceability capability before regulatory requirements crystallise enter those requirements with lower implementation costs – the infrastructure is already there. They also enter supervisory examinations with a stronger position: examiners reviewing AI-assisted decision processes find documented governance rather than reconstructed rationale. They accumulate the evidentiary record that matters most in adverse scenarios: when a client disputes a transaction, when a regulator investigates a pattern, or when an internal audit questions a decision chain.

Compliance functions should treat the EDR framework not as an additional compliance layer but as a rationalisation of existing documentation obligations. The information required by the four-layer model is substantially the information that FINMA, AML obligations and fiduciary requirements already mandate – it is simply not currently captured in a coherent, AI-aware governance architecture.

Two further implications follow. For vendors, Layer 2 capture should be a product feature of AI-integrated exception systems, not an institutional workaround. For the wider system, the AML and sanctions exceptions discussed here bear on the integrity of the financial system and the effectiveness of sanctions regimes, giving better traceability here a public-interest dimension as well as a supervisory one.

There is also a client-facing dimension the framework touches without resolving. If a client disputes a held or rejected payment, EDR records can support a fuller account than current documentation allows. The framework does not require disclosure to clients, and that boundary remains a separate question – but the capability, once built, is available for it.

7.4 *For the regulatory trajectory*

The regulatory trajectory for AI governance in payment operations is moving in the direction that the EDR framework anticipates. The EU AI Act's requirements for high-risk AI systems – including logging obligations, human oversight mandates and accountability documentation – will apply to AI systems deployed in consequential financial decision workflows for institutions with EU market exposure. The EBA's ([European Banking Authority, 2025](#)) consultation on revised internal governance guidelines signals an evolution towards more explicit decision-layer governance requirements.

The framework also has implications for supervisors and internal auditors. Regulatory examinations of AI-assisted payment operations currently face a structural evidential problem: they can observe outcomes and test controls, but they cannot reconstruct the AI contributions that preceded human decisions unless institutions have captured them contemporaneously. Decision-layer documentation is, in this sense, also a precondition for effective supervisory examination, not only an institutional interest.

8. Conclusion

Payment exception handling in Swiss private banking is a governance-intensive activity that current accountability frameworks are not designed to address adequately in the AI-assisted decision environment. The integration of generative AI and AI-supported analysis tools into exception workflows creates a new class of governance challenge: the undocumented AI-mediated decision, in which the accountability record captures what a human decided but not how the human was informed by the AI systems that shaped the decision.

This paper has made three contributions to addressing this challenge – though the weight of each will depend on the reader's prior commitments. Firstly, it developed a taxonomy of payment exceptions in private banking that distinguishes exception types by their governance properties, particularly decision intensity and regulatory exposure, providing a principled basis for proportional governance investment. Secondly, it identified four structural governance failure modes in AI-assisted exception handling. These failure modes are not individual governance failures; they are structural features of exception decision environments that AI integration amplifies, and the number four reflects the analysis rather than a design target. Thirdly, it proposed the Exception Decision Record as a governance instrument – a layered framework for creating contemporaneous, proportional, complete and reconstructable accountability records. Whether practitioners find the framework useful will ultimately depend less on its internal logic than on whether it captures something true about the institutions they work in.

The EDR framework is not a theoretical construct: it is a governance architecture derived from the operational realities of Swiss private banking and grounded in decision traceability theory, operational risk scholarship and the regulatory requirements that private banking institutions already face. Its implementation does not require regulatory mandate; it requires institutional recognition that adequate governance of AI-assisted decision-making is a capability that must be built, not a compliance obligation to be satisfied after the fact. That said, the framework has limitations that should be acknowledged. It is conceptual, not empirically validated. The implementation challenges identified in Section 7.2 are real, not rhetorical. And the four-layer model, however logically derived, will require adaptation in institutional environments that differ from the Swiss private banking context on which it is calibrated.

The broader implication of this paper extends beyond payment exception handling to the full landscape of AI-assisted operational decisions in private banking. The governance challenge identified here – AI inputs that shape human decisions but are not captured in accountability records – is structurally present in suitability assessment, AML alert review, KYC decision-making and corporate action processing. The EDR framework, adapted to the specific decision environment of each workflow, provides a generalisable governance architecture for the hybrid human-AI decision environments that are becoming the operational norm in regulated private banking.

A brief illustration shows the generalisation is substantive rather than asserted. Take AML alert review. Where payment exception handling involves AI-generated resolution recommendations, alert review involves AI-generated typology scores – probabilistic assessments that a transaction or pattern fits a money-laundering typology – followed by human adjudication. The decision-layer problem is the same: the analyst's conclusion is recorded (alert closed, escalated or reported) while the AI input that informed it is not. Layer 2, adapted, would record the typology score, the features the system surfaced and the confidence level at the moment of presentation; Layer 3 would record whether the analyst accepted, modified, overrode or escalated. The same architecture applies; only the decision class differs.

As the regulatory environment for AI governance in financial services continues to develop – through FINMA's evolving guidance, the EU AI Act's progressive application and the FSB's ongoing analytical work – institutions that have invested in decision traceability infrastructure will be better positioned to demonstrate governance adequacy, respond to supervisory scrutiny and maintain the fiduciary accountability that the private banking relationship demands. Waiting for regulatory mandate before building that capability is a choice institutions will find difficult to explain when the mandate arrives.

Ethics statement

This paper is a conceptual study. No primary data collection was conducted. No ethics approval was required.

Note on AI use

In accordance with Emerald's author guidelines, the intellectual content, argumentation, framework design and all substantive writing in this manuscript represent the original scholarly work of the author. AI assistance was used solely for copy-editing and language revision of the author's original draft text.

References

- Author (2025), "An integrated AI-blockchain governance framework for operational resilience in private banking", *Working Paper Version 1.3*, doi: [10.5281/zenodo.17952003](https://doi.org/10.5281/zenodo.17952003).
- Author (2026a), "Decision traceability and governance challenges of generative AI in private banking operations, manuscript under review".
- Bank for International Settlements (BIS) Committee on Payments and Market Infrastructures (CPMI) (2022), "Improving the speed, cost, transparency and access of Cross-Border payments: Consultative report", Bank for International Settlements, Basel, available at: www.bis.org/cpmi/publ/d211.pdf (accessed 26 March 2026).
- Basel Committee on Banking Supervision (BCBS) (2021), "Principles for operational resilience, bank for international settlements", Basel, available at: www.bis.org/bcbs/publ/d516.pdf (accessed 26 March 2026).
- Bovens, M. (2007), "Analysing and assessing accountability: a conceptual framework", *European Law Journal*, Vol. 13 No. 4, pp. 447-468, doi: [10.1111/j.1468-0386.2007.00378.x](https://doi.org/10.1111/j.1468-0386.2007.00378.x).
- Bovens, M., Goodin, R.E. and Schillemans, T. (Eds) (2014), *The Oxford Handbook of Public Accountability*, Oxford University Press, Oxford, doi: [10.1093/oxfordhb/9780199641253.001.0001](https://doi.org/10.1093/oxfordhb/9780199641253.001.0001).
- Celent (2023), *Exception Processing in Private Banking: Benchmarks, Automation Gaps and Operational Risk Implications*, Celent Research Report, Boston, MA.
- Chorafas, D.N. (2001), *Managing Operational Risk: Reducing Business Uncertainties*, Palgrave Macmillan, London.
- Coglianese, C. and Lehr, D. (2017), "Regulating by robot: administrative decision making in the machine-learning era", *Georgetown Law Journal*, Vol. 105 No. 5, pp. 1147-1223.
- Cummings, M.L. (2012), "Automation and accountability in decision support system interface design", *Journal of Technology Studies*, Vol. 32 No. 1, pp. 23-31, doi: [10.21061/jots.v32i1.a.4](https://doi.org/10.21061/jots.v32i1.a.4).
- Demetis, D.S. (2018), "Fighting money laundering with technology: a case study of bank X in the UK", *Decision Support Systems*, Vol. 105, pp. 96-107, doi: [10.1016/j.dss.2017.11.005](https://doi.org/10.1016/j.dss.2017.11.005).
- Diakopoulos, N. (2016), "Accountability in algorithmic decision making", *Communications of the ACM*, Vol. 59 No. 2, pp. 56-62, doi: [10.1145/2844110](https://doi.org/10.1145/2844110).
- Doshi-Velez, F. and Kim, B. (2017), "Towards a rigorous science of interpretable machine learning", arXiv Preprint arXiv:1702.08608, available at: <https://arxiv.org/abs/1702.08608> (accessed 26 March 2026).
- Dreyfus, H.L. and Dreyfus, S.E. (1986), *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*, Free Press, New York, NY.
- European Banking Authority (2025), "Consultation paper on draft revised guidelines on internal governance (EBA/CP/2025/20)", European Banking Authority, Paris, available at: www.eba.europa.eu (accessed 26 March 2026).

- Financial Action Task Force (FATF) (2023), “Guidance on beneficial ownership for legal persons”, FATF, Paris, available at: www.fatf-gafi.org/en/publications/Fatfrecommendations/Guidance-beneficial-ownership-legal-persons.html (accessed 26 March 2026).
- Financial Stability Board (2024), “The financial stability implications of artificial intelligence, FSB, Basel”, available at: www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/ (accessed 26 March 2026).
- FINMA (2023), *Circular 2023/1: Operational Risks and Resilience — Banks*, Swiss Financial Market Supervisory Authority, Bern, available at: www.finma.ch/en/regulation/finma-circulars/ (accessed 26 March 2026).
- FINMA (2024), *Guidance 08/2024: Artificial Intelligence at Supervised Institutions*, Swiss Financial Market Supervisory Authority, Bern, available at: www.finma.ch/en/news/2024/06/20240626-mm-wegleitung-ki/ (accessed 26 March 2026).
- Kroll, J.A., Huey, J., Barocas, S., Felten, E.W., Reidenberg, J.R., Robinson, D.G. and Yu, H. (2017), “Accountable algorithms”, *University of Pennsylvania Law Review*, Vol. 165 No. 3, pp. 633-705, available at: https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/ (accessed 26 March 2026).
- Lipton, Z.C. (2018), “The mythos of model interpretability”, *Queue*, Vol. 16 No. 3, pp. 31-57, doi: [10.1145/3236386.3241340](https://doi.org/10.1145/3236386.3241340).
- Maxwell, W. and Vannieuwenhuyse, A. (2018), “Robots replacing human compliance officers? Automated decision-making and financial regulation”, *Revue Internationale Des Services Financiers*, No. 4, pp. 22-31.
- Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016), “The ethics of algorithms: mapping the debate”, *Big Data and Society*, Vol. 3 No. 2, pp. 1-21, doi: [10.1177/2053951716679679](https://doi.org/10.1177/2053951716679679).
- Parasuraman, R. and Manzey, D.H. (2010), “Complacency and bias in human use of automation: an attentional integration”, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 52 No. 3, pp. 381-410, doi: [10.1177/0018720810376055](https://doi.org/10.1177/0018720810376055).
- Polanyi, M. (1966), *The Tacit Dimension*, Doubleday, Garden City, New York, NY.
- Power, M. (2007), *Organized Uncertainty: Designing a World of Risk Management*, Oxford University Press, Oxford.
- Roberts, J. (1991), “The possibilities of accountability”, *Accounting, Organizations and Society*, Vol. 16 No. 4, pp. 355-368, doi: [10.1016/0361-3682\(91\)90027-C](https://doi.org/10.1016/0361-3682(91)90027-C).
- Rudin, C. (2019), “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead”, *Nature Machine Intelligence*, Vol. 1 No. 5, pp. 206-215, doi: [10.1038/s42256-019-0048-x](https://doi.org/10.1038/s42256-019-0048-x).
- Santoni de Sio, F. and van den Hoven, J. (2018), “Meaningful human control over autonomous systems: a philosophical account”, *Frontiers in Robotics and AI*, Vol. 5, p. 15, doi: [10.3389/frobt.2018.00015](https://doi.org/10.3389/frobt.2018.00015).
- Selbst, A.D., Boyd, D., Friedler, S.A., Venkatasubramanian, S. and Vertesi, J. (2019), “Fairness and abstraction in sociotechnical systems”, in *Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency (FAT*)*, ACM, New York, NY, pp. 59-68, doi: [10.1145/3287560.3287598](https://doi.org/10.1145/3287560.3287598).
- SWIFT Institute (2022), “Correspondent banking in transition: Compliance, efficiency and the future of Cross-Border payments, SWIFT institute working paper No. 2022-001”, Society for Worldwide Interbank Financial Telecommunication, Brussels.
- Wachter, S., Mittelstadt, B. and Russell, C. (2017), “Counterfactual explanations without opening the black box: automated decisions and the GDPR”, *Harvard Journal of Law and Technology*, Vol. 31 No. 2, pp. 841-887, available at: <https://jolt.law.harvard.edu/assets/articlePDFs/v31/>

[Counterfactual-Explanations-without-Opening-the-Black-Box-Sandra-Wachter-et-al.pdf](#)
(accessed 26 March 2026).

Wickens, C.D. (2008), “Multiple resources and mental workload”, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 50 No. 3, pp. 449-455, doi: [10.1518/001872008X288394](#).

Journal of
Financial
Regulation and
Compliance

Further reading

Author (2026b), “Regulatory expectations and institutional gaps in AI decision traceability: a documentary analysis of prudential frameworks in Swiss private banking, working manuscript”.

Basel Committee on Banking Supervision (BCBS) (2023), “Principles for the sound management of operational risk, bank for international settlements”, Basel, available at: [www.bis.org/bcbs/publ/d532.pdf](#) (accessed 26 March 2026).

European Parliament and Council of the European Union (2024), Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence (EU AI Act), *Official Journal of the European Union*, L 1689, Brussels.

Corresponding author

Hervé Racordon can be contacted at: hervé.racordon@gab-bag.com