AMBER-I: Accounting, Management, Business, and Economic Research International Journal
Vol 1 No 2 October 2026
Open Access: https://publikasi.polije.ac.id/amberi/index

DFAS-AAP: AI accountability protocol in financial modelling

Hasan Mohamed Husain Alaali[1*]
[1]DFAS Research Center, Kingdom of Bahrain
email: Founder@dfas.ai; hasan.mohd.alaali@gmail.com
ORCID: https://orcid.org/0009-0007-2489-9435

Article Info :
Received:
17-08-2026
Revised:
10-09-2026
Accepted:
12-10-2026

ABSTRACT

The growing adoption of artificial intelligence in financial modelling created accountability challenges concerning model traceability, human intervention, internal control, and financial reporting. This study developed the DFAS-AAP (AI Accountability Protocol) to integrate these functions within financial decision processes. A conceptual protocol-development approach was applied by combining accountability theory, agency theory, and internal-control principles with financial reporting and artificial intelligence governance requirements. The architecture was assessed through theoretical integration, regulatory mapping, and illustrative financial scenarios. The protocol comprised five interconnected components addressing model classification and responsibility attribution, explainability escalation, reporting-related documentation, role-based override authority, and control integrity with decision traceability. It incorporated the Alaali AI Override Chain of Command and the Alaali Explainability Escalation Framework. Illustrative applications in valuation, environmental, social, and governance risk scoring, and credit stress analysis showed how these mechanisms interacted. The analysis indicated that accountability in finance required the integration of explainability, human oversight, intervention authority, documentation, and auditability rather than reliance on each mechanism independently. The protocol provided a structured conceptual basis for this integration, although empirical institutional validation remained necessary.

Keywords : AI governance; financial modelling; accountability; explainability; internal control; financial reporting

©2022 Authors.. This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.
(https://creativecommons.org/licenses/by-nc/4.0/)

INTRODUCTION

Artificial intelligence (AI) is increasingly embedded in financial valuation, credit assessment, risk modelling, investment analysis, ESG evaluation, and financial reporting (Brynjolfsson & McAfee, 2021; Bussman et al., 2021; Srivastava & Callen, 2020). These applications can improve analytical speed, scalability, and the capacity to process complex datasets, but they also alter the structure of responsibility within financial decision-making. When an AI-generated output contributes to a material financial decision, accountability may be distributed across model developers, users, managers, auditors, technology providers, and institutional control functions (Mökander & Floridi, 2022; National Institute of Standards and Technology [NIST], 2023). This creates a governance problem that extends beyond model accuracy or technical explainability: institutions must be able to determine who is responsible for reviewing an output, when human intervention is required, who has authority to override the system, how that intervention is documented, and how the resulting decision can subsequently be audited (Deloitte, 2023a; Public Company Accounting Oversight Board [PCAOB], 2023).

Existing research has addressed important components of this problem. Explainable AI research has examined interpretability and the ability to understand model outputs (Doshi-Velez & Kim, 2017; Rudin, 2019), while algorithmic accountability research has emphasized transparency, traceability, and institutional responsibility (Kroll et al., 2017; Kroll et al., 2022; Wieringa, 2020). Accountability theory provides a basis for examining answerability and responsibility, whereas agency theory highlights information asymmetry and monitoring problems between principals and agents (Bovens, 2007; Jensen & Meckling, 1976). Internal-control frameworks such as COSO provide established mechanisms for allocating responsibilities, monitoring activities, documenting controls, and responding to risk

(Committee of Sponsoring Organizations of the Treadway Commission [COSO], 2013; Deloitte, 2023b). Regulatory developments have similarly increased expectations concerning human oversight, documentation, risk management, and transparency in AI-enabled systems (European Commission, 2021; Hacker, 2023; International Organization for Standardization [ISO], 2022; NIST, 2023).

These streams nevertheless address different parts of the accountability problem. Explainability does not by itself determine who has authority to intervene in a financial decision (Doshi-Velez & Kim, 2017; Rudin, 2019). Internal controls may establish organizational responsibilities without specifying how those responsibilities should operate when decisions depend on probabilistic or opaque AI outputs (COSO, 2013; Deloitte, 2023b). Regulatory requirements establish governance expectations but do not necessarily provide an institution-level workflow connecting model classification, explainability, escalation, human override, documentation, internal control, and financial disclosure (European Commission, 2021; Floridi & Cowls, 2022; Mittelstadt et al., 2023; Mökander & Floridi, 2022). The resulting gap is therefore not simply a lack of AI governance principles, but a lack of an integrated operational architecture for translating those principles into accountable financial decision processes.

This study addresses that gap through the DFAS-AAP (AI Accountability Protocol), a conceptual governance protocol for AI-enabled financial modelling and decision-making. DFAS-AAP integrates five complementary governance components covering model classification and responsibility attribution, explainability escalation, financial-reporting disclosure, human override authority, and control-integrity logging (Alaali, 2025a, 2025b, 2025c, 2025d). Within this architecture, the Alaali AI Override Chain of Command (A-AOCC) specifies role-based escalation and intervention responsibilities, while the Alaali Explainability Escalation Framework (AEEF) provides a mechanism for increasing human scrutiny when an AI-generated output exhibits characteristics requiring additional review. These mechanisms are intended to connect technical model behaviour with identifiable institutional responsibility rather than treating explainability, oversight, and auditability as separate governance objectives.

The study is guided by the following research question: How can accountability, human override authority, explainability escalation, internal control, and disclosure requirements be integrated into a coherent governance protocol for AI-enabled financial decision-making? The objective is not to establish the empirical effectiveness of DFAS-AAP at this stage, but to develop and analytically examine a structured protocol capable of translating established accountability principles into operational governance mechanisms. Accordingly, the paper distinguishes conceptual development and illustrative application from empirical validation.

The study contributes primarily through conceptual integration and operationalization. First, it connects accountability theory, agency theory, and internal-control principles to the specific governance problems created by AI-enabled financial decisions (Bovens, 2007; COSO, 2013; Jensen & Meckling, 1976). Second, it translates these theoretical foundations into identifiable mechanisms for classification, escalation, intervention, documentation, and auditability. Third, it maps these mechanisms to relevant financial and AI-governance requirements to examine how the proposed architecture could operate within existing institutional and regulatory environments (Basel Committee on Banking Supervision [BCBS], 2022; European Commission, 2021; IFRS Foundation, 2022; ISO, 2022; NIST, 2023). The contribution therefore lies not merely in combining existing governance concepts, but in specifying how previously separated accountability functions can interact within a single financial AI decision chain.

The remainder of the paper develops this argument systematically. The research methods section explains the conceptual development method and the analytical logic used to construct and assess DFAS-AAP. The results and discussion section presents and discusses the resulting architecture, its theoretical foundations, constituent protocols, illustrative applications, regulatory mapping, and implementation implications. The conclusion summarizes the contribution, limitations, and directions for empirical validation.

RESEARCH METHODS

This study adopts a conceptual protocol-development approach. The purpose is to construct and analytically examine an accountability architecture for AI-enabled financial decision-making rather than to estimate statistical relationships or claim empirical validation. The methodological design therefore focuses on theoretical integration, governance-mechanism specification, comparative regulatory mapping, and illustrative application.

The development process began by identifying the principal accountability functions that become problematic when AI-generated outputs enter financial decision chains. These functions include attribution of responsibility, model classification, explainability, escalation, human intervention, override authority, documentation, internal control, auditability, and financial disclosure (European Commission, 2021; Kroll et al., 2017; NIST, 2023; Wieringa, 2020). The conceptual problem was then examined against three established theoretical and institutional foundations: accountability theory, agency theory, and the COSO internal-control framework (Bovens, 2007; COSO, 2013; Jensen & Meckling, 1976). These foundations were selected because they address complementary dimensions of the governance problem. Accountability theory concerns answerability and responsibility; agency theory addresses delegation, information asymmetry, and monitoring; and COSO provides an institutional structure for control activities, information, monitoring, and risk management.

The protocol was constructed by translating these theoretical requirements into operational governance functions. Rather than treating transparency or human oversight as general principles, the development logic requires each governance function to correspond to an identifiable mechanism within the decision process (Floridi & Cowls, 2022; Mittelstadt et al., 2023; Mökander & Floridi, 2022). This resulted in an integrated architecture covering model classification and responsibility attribution, explainability escalation, disclosure-ready documentation, role-based override authority, and control-integrity logging (Alaali, 2025a, 2025b, 2025c, 2025d). The A-AOCC was incorporated to specify the institutional pathway through which intervention authority can be assigned and escalated, while AEEF was incorporated to structure the transition from ordinary model use to enhanced human scrutiny when predefined review conditions are encountered.

The resulting architecture was then assessed through a comparative regulatory and institutional mapping. Relevant requirements from AI governance, financial reporting, prudential oversight, and internal-control frameworks were compared with the functions provided by DFAS-AAP (BCBS, 2022; European Commission, 2021; IFRS Foundation, 2022; ISO, 2022; NIST, 2023). The unit of comparison was the governance function rather than the framework as a whole. The principal comparison dimensions were human oversight, transparency and explainability, decision traceability, internal control, documentation, disclosure, and post-decision accountability. This approach was used to identify areas of correspondence and remaining differences without treating conceptual alignment as evidence of regulatory certification or formal compliance.

The study also uses illustrative scenarios to examine the internal operation of the proposed mechanisms in financial contexts. The scenarios cover AI-assisted valuation, ESG risk scoring, and credit stress analysis (Kotsantonis et al., 2016; Srivastava & Callen, 2020). They are used to demonstrate how classification, escalation, human review, override authority, documentation, and disclosure could interact within a decision workflow. These scenarios are hypothetical demonstrations rather than empirical case studies; they do not constitute evidence that DFAS-AAP improves institutional performance or regulatory outcomes.

The analytical assessment therefore operates at three levels. The first is theoretical coherence, examining whether the proposed mechanisms correspond to the accountability, agency, and internal-control problems identified in the theoretical foundations. The second is operational coherence, examining whether responsibilities, escalation pathways, interventions, and documentation can be connected through a logically identifiable governance sequence. The third is institutional compatibility, examining whether the resulting mechanisms can be related to existing financial reporting, internal-control, and AI-governance expectations (BCBS, 2022; European Commission, 2021; IFRS Foundation, 2022; NIST, 2023).

This methodology establishes a reproducible conceptual logic for DFAS-AAP while maintaining a clear boundary between protocol development and empirical validation. The study does not claim that the proposed architecture has been validated through institutional deployment. Its findings should therefore be interpreted as conceptual and operational propositions that provide a basis for subsequent pilot implementation, comparative evaluation, and empirical testing.

RESULTS AND DISCUSSION

The conceptual development process produced an integrated accountability architecture in which AI governance is treated as a sequence of connected institutional controls rather than as a collection of independent principles (Floridi & Cowls, 2022; Mökander & Floridi, 2022; NIST, 2023).

The resulting DFAS-AAP architecture links model identification and responsibility attribution to explainability review, escalation, human intervention, disclosure, and audit documentation. Accountability is consequently positioned across the AI decision lifecycle rather than being activated only after a problematic decision has occurred (Kroll et al., 2022; Mittelstadt et al., 2023).

The central analytical result is that effective accountability in AI-enabled finance requires coordination among several governance functions. Explainability without intervention authority may reveal why an output is questionable without determining who can act (Doshi-Velez & Kim, 2017; Rudin, 2019). Human oversight without documented escalation may establish nominal responsibility without producing a traceable decision process. Internal controls without AI-specific triggers may monitor organizational processes while failing to identify when probabilistic model behaviour requires additional scrutiny (COSO, 2013; Deloitte, 2023b). Similarly, disclosure without a documented decision trail may describe an outcome without preserving the institutional reasoning that produced it (IFRS Foundation, 2022; PCAOB, 2023). DFAS-AAP addresses these interfaces by linking the individual governance functions within a common decision architecture.

The following sections examine the theoretical foundations of this architecture before presenting its operational components, illustrative applications, regulatory relationships, and implementation implications.

Theoretical Framework

DFAS-AAP is grounded in three complementary foundations: accountability theory, agency theory, and internal-control theory as operationalized through COSO (Bovens, 2007; COSO, 2013; Jensen & Meckling, 1976). These foundations address different but interconnected questions. Accountability theory explains why decisions must remain attributable and reviewable. Agency theory explains why delegation to AI-enabled systems can create information and control asymmetries. COSO provides an institutional basis for embedding oversight, documentation, risk assessment, and monitoring into organizational processes. Their integration provides the theoretical basis for treating AI accountability as an organizational control problem rather than solely as a technical explainability problem.

Accountability Theory

Accountability concerns the relationship between an actor responsible for conduct and a forum capable of requesting information, evaluating that conduct, and imposing consequences or corrective action (Bovens, 2007). Applied to AI-enabled finance, this relationship becomes more complex because the analytical output may originate from a computational system while the authority to accept, reject, modify, or act upon that output remains distributed among human and institutional actors. Accountability therefore cannot be assigned meaningfully to the algorithm itself; it must be preserved through the governance structure surrounding its development, deployment, review, and use (Kroll et al., 2022; Mökander & Floridi, 2022).

Algorithmic accountability literature reinforces the importance of traceability, explanation, and institutional oversight when automated systems contribute to consequential decisions (Kroll et al., 2017; Wieringa, 2020). For financial applications, these requirements are particularly significant because model outputs can affect valuations, lending decisions, risk assessments, capital allocation, and reporting (Bussman et al., 2021; Malhotra, 2022). A governance system must therefore preserve not only the output generated by the model but also the institutional pathway through which that output becomes an authorized financial decision.

DFAS-AAP operationalizes this requirement by distributing accountability across identifiable stages of the AI decision lifecycle. Model classification establishes the characteristics and governance requirements of the system. Explainability escalation identifies conditions requiring additional scrutiny. Role-based override mechanisms determine who is authorized to intervene. Logging and documentation preserve the rationale and sequence of interventions. Disclosure mechanisms connect relevant model and decision information to financial-reporting requirements (Alaali, 2025a, 2025b, 2025d). Accountability is thus treated as an ex ante design requirement and an ex post verification capability.

This approach also distinguishes accountability from explainability. An explainable output may still lack an accountable decision process if no person or institutional function is responsible for evaluating or acting upon the explanation. Conversely, assigning a responsible officer without providing sufficient information about the model output can create formal accountability without substantive

oversight (Kroll et al., 2022; Mittelstadt et al., 2023). DFAS-AAP therefore treats explainability and responsibility as complementary but analytically distinct requirements.

Agency Theory in Autonomous Finance

Agency theory traditionally examines relationships in which a principal delegates authority to an agent whose information, incentives, or actions may differ from the principal's interests (Jensen & Meckling, 1976). Financial institutions already operate through multiple agency relationships among shareholders, managers, employees, auditors, customers, and regulators. AI-enabled decision systems introduce an additional layer of delegation because analytical tasks previously performed or directly interpreted by humans can be partially transferred to computational systems (Gawer, 2023; Yermack, 2018).

AI does not constitute an economic agent in the classical contractual sense. Nevertheless, its use can intensify agency-type problems by increasing information asymmetry between those who develop or supply a system, those who operate it, those who authorize decisions based on its outputs, and those affected by those decisions (Bussman et al., 2021; Malhotra, 2022). A manager may receive a recommendation without understanding the model's internal reasoning; an institution using a third-party system may have limited access to model parameters or training data; and an auditor may need to evaluate a decision produced through a chain involving both automated and human judgments (PCAOB, 2023). The governance problem therefore concerns delegated analytical authority and the institutional capacity to monitor and control its consequences.

DFAS-AAP responds to this problem by preserving human decision authority at defined points in the workflow. The A-AOCC specifies role-based responsibilities for reviewing and overriding outputs, while escalation mechanisms determine when additional human scrutiny is required. The purpose is not to assume that an institution can modify every AI system internally. In closed-source or vendor-controlled systems, institutional authority may instead operate through output acceptance or rejection, human review, escalation, audit logging, procurement conditions, contractual requirements, deployment restrictions, or system suspension (European Commission, 2021; ISO, 2022; NIST, 2023).

This distinction is important because governance authority must correspond to the technical and organizational access actually available to the institution. By locating accountability at the points where institutional actors possess genuine decision or control authority, DFAS-AAP extends agency logic to AI-enabled financial workflows without treating the technology itself as a legally or institutionally accountable agent.

COSO and Internal Control Extensions

The COSO Internal Control—Integrated Framework provides an established structure for designing and evaluating organizational controls through five interrelated components: the control environment, risk assessment, control activities, information and communication, and monitoring activities (COSO, 2013). These components provide a relevant institutional foundation for AI governance because accountable use of AI requires responsibilities, risk identification, intervention mechanisms, information flows, documentation, and continuing monitoring (Deloitte, 2023b).

AI-enabled financial systems nevertheless introduce control conditions that require additional operational specification. Model behaviour may change over time, outputs may be probabilistic rather than deterministic, decision logic may not be fully observable, and institutions may rely on third-party systems over which they have limited technical control (Bussman et al., 2021; NIST, 2023). Conventional control procedures therefore need to be connected to AI-specific events such as anomalous outputs, material deviations, explainability deficiencies, model changes, or decisions requiring human escalation (Malhotra, 2022).

DFAS-AAP applies COSO principles to these conditions by linking AI decision events to identifiable control activities (Alaali, 2025d). Model classification and responsibility assignment contribute to the control environment and risk-assessment process. Explainability and escalation mechanisms establish conditions for additional review. Role-based override pathways function as control activities by defining who can intervene and under what authority. Decision logs and disclosure records support information and communication, while continuing review of model outputs and interventions supports monitoring.

The proposed extension does not replace COSO or claim to modify the formal COSO framework. Rather, it uses COSO as an institutional control foundation and specifies how AI-related

accountability mechanisms can be positioned within that structure. This distinction allows DFAS-AAP to address the operational characteristics of AI-enabled financial decision-making while remaining anchored in established internal-control principles (COSO, 2013; Deloitte, 2023b).

Literature Review

The growing use of artificial intelligence in financial decision-making has generated several related streams of research concerning algorithmic accountability, explainability, model risk, internal control, and regulatory governance (Bussman et al., 2021; Malhotra, 2022; Mökander & Floridi, 2022). Although these streams address important dimensions of AI governance, they remain only partially integrated. This fragmentation is particularly important in finance, where an AI-generated output may affect valuation, credit assessment, risk measurement, financial reporting, or capital allocation and therefore must remain subject to identifiable institutional responsibility (Srivastava & Callen, 2020).

A substantial body of research has focused on algorithmic accountability. Kroll et al. (2017) argue that accountable algorithmic systems require mechanisms capable of supporting scrutiny and verification rather than relying solely on technical transparency. Wieringa (2020) similarly emphasizes that algorithmic accountability is relational and institutional: accountability depends not only on understanding a system but also on identifying who must provide an account, to whom, and under what conditions. This distinction is important for financial applications because responsibility cannot be transferred to an algorithm merely because the algorithm generated the analytical output (Kroll et al., 2022). Human and institutional actors remain responsible for determining how that output enters the decision process.

A second literature stream concerns explainable artificial intelligence (XAI). Doshi-Velez and Kim (2017) and Rudin (2019) demonstrate the importance of interpretability when machine-learning systems influence consequential decisions. Explainability can help users understand model behaviour, identify questionable outputs, and evaluate whether a recommendation is consistent with relevant evidence. However, explainability alone does not specify what happens after a problematic output has been identified. It does not necessarily determine who must review the output, who possesses override authority, when escalation is mandatory, or how the intervention should be documented (Mittelstadt et al., 2023). Consequently, explainability represents an important component of accountability but not a complete governance architecture.

The literature on internal control and model governance introduces a complementary perspective. COSO (2013) establishes an institutional structure based on control environment, risk assessment, control activities, information and communication, and monitoring. These principles provide a foundation for assigning responsibilities and maintaining control over financial processes. AI-enabled systems, however, create additional challenges because model outputs may be probabilistic, model behaviour may evolve, and decision logic may be difficult to observe (Deloitte, 2023b; Malhotra, 2022). The governance problem therefore extends from controlling conventional processes to determining how institutions should identify, escalate, review, and document AI-related decision risks.

Regulatory developments have reinforced these concerns. Contemporary AI governance approaches increasingly emphasize risk management, documentation, transparency, human oversight, and continuing monitoring (European Commission, 2021; Hacker, 2023; ISO, 2022; NIST, 2023). Financial institutions simultaneously operate under established reporting, prudential, audit, and internal-control obligations (BCBS, 2022; IFRS Foundation, 2022; PCAOB, 2023). The resulting governance environment requires institutions to satisfy both AI-specific expectations and existing financial accountability requirements. This creates a coordination problem: AI governance cannot operate independently from the control and reporting structures through which financial institutions already establish responsibility.

The literature therefore reveals a distinction between principle-level governance and operational accountability. Existing approaches provide important principles concerning transparency, fairness, explainability, risk management, and human oversight (Floridi & Cowls, 2022; Mökander & Floridi, 2022), but these principles do not automatically produce an institutional decision pathway. For example, requiring human oversight does not by itself determine which role performs that oversight, what event triggers intervention, what authority the reviewer possesses, how disagreement with the model is resolved, or how the final decision is recorded. Similarly, requiring documentation does not

establish which stages of the AI decision chain must be documented or how those records connect to audit and disclosure requirements.

This problem is especially significant where financial institutions use third-party or closed-source AI systems. In such circumstances, the institution may not control model weights, training data, source code, or internal recalibration (NIST, 2023). Governance mechanisms must therefore operate at the levels where institutional authority actually exists, including input controls, output review, human authorization, escalation, audit logging, procurement requirements, contractual obligations, deployment restrictions, and system suspension (European Commission, 2021; ISO, 2022). An operational accountability architecture must recognize these technical boundaries rather than assuming unrestricted institutional control over the underlying AI model.

A further gap concerns the relationship between pre-decision and post-decision accountability. Governance mechanisms are frequently concentrated either on ex ante model assessment or ex post auditing (Mittelstadt et al., 2023; PCAOB, 2023). Financial AI requires continuity between these stages. Model classification should influence the level of oversight; unusual or insufficiently explainable outputs should activate escalation; escalation should lead to an identifiable reviewer; intervention should be governed by defined authority; and the resulting decision should remain traceable through documentation and audit records. Accountability therefore needs to operate as a lifecycle process rather than as an isolated compliance event.

Against this background, DFAS-AAP addresses a specific conceptual gap: the absence of an integrated finance-oriented architecture connecting model classification, responsibility attribution, explainability escalation, human override authority, internal-control documentation, and disclosure within the same governance sequence. The proposed contribution is not that these individual concepts are absent from the literature. Rather, it lies in specifying how they can be connected operationally so that an AI-generated financial output remains attributable, reviewable, contestable, and auditable throughout the institutional decision process.

The DFAS-AAP Framework: Five Integrated Protocols

DFAS-AAP translates the accountability requirements identified in the theoretical and literature analysis into five interconnected governance protocols (Alaali, 2025a, 2025b, 2025c, 2025d). Each protocol addresses a different stage or function within the AI-enabled financial decision process, while their integration is intended to preserve responsibility from model identification through final decision documentation.

The architecture follows a sequential governance logic. DFAS-FEP establishes model classification and responsibility attribution (Alaali, 2025a). DAIF provides the audit and explainability-escalation layer through which outputs requiring additional scrutiny can be identified. DFAS-IFRS structures relevant model and decision information for financial-reporting and disclosure purposes (Alaali, 2025b). DFAS-Governance establishes the role-based chain through which human intervention and override authority are exercised. DFAS-CICP preserves the control and logging record connecting these activities (Alaali, 2025d).

The five protocols are complementary rather than interchangeable. Classification without escalation does not determine when intervention is required. Escalation without authority does not determine who can act. Override authority without documentation weakens subsequent accountability. Documentation without an internal-control structure may preserve information without establishing whether appropriate procedures were followed. Disclosure without traceability may report an outcome without explaining the governance process underlying it. DFAS-AAP therefore treats these functions as components of a connected accountability chain.

DFAS-FEP (Future Engine Prototype): Model Classification and Authorship Governance

DFAS-FEP provides the initial classification and responsibility layer within DFAS-AAP (Alaali, 2025a). Its function is to establish the identity, characteristics, provenance, and responsible human contributions associated with an AI-enabled model before the model's outputs enter consequential financial decision processes.

The classification function differentiates models according to the nature and extent of computational autonomy involved in their operation. This distinction is relevant because a deterministic model operating according to transparent rules does not create the same governance requirements as an AI-assisted model incorporating statistical learning or a system whose outputs are substantially

generated through less interpretable computational processes. Classification therefore provides a basis for calibrating subsequent oversight requirements to the characteristics and use of the model rather than imposing identical governance procedures on all systems.

The responsibility function records identifiable human and institutional involvement across relevant stages of the model lifecycle. These may include model design, development, validation, deployment, modification, authorization, and material human intervention. The objective is not to attribute authorship or responsibility to the AI system itself, but to preserve the traceability of the human and institutional actors responsible for its development and use (Alaali, 2025a).

Within DFAS-AAP, DFAS-FEP therefore acts as the entry point to the accountability chain. The resulting classification and responsibility information can inform subsequent monitoring, escalation, review, and documentation requirements. Higher-risk or less interpretable applications may require more intensive oversight, while lower-risk and more transparent applications may justify proportionately lighter controls (European Commission, 2021; NIST, 2023). The appropriate governance intensity consequently depends on the model's characteristics, context, and potential decision impact.

DAIF: Dynamic Audit Intelligence Framework

DAIF provides the monitoring and escalation component of DFAS-AAP. Its purpose is to identify circumstances in which ordinary reliance on an AI-generated output should give way to enhanced review and documented human scrutiny.

A central mechanism within DAIF is the Alaali Explainability Escalation Framework (AEEF). AEEF establishes a structured pathway through which defined indicators can trigger progressively greater review. Relevant indicators may include material deviation from expected outputs, unusual variance, insufficient explanation, inconsistency with available financial information, changes in model behaviour, or other institutionally defined risk conditions. The specific triggers should be calibrated to the application and supported by appropriate validation rather than treated as universal thresholds (NIST, 2023).

The escalation process separates anomaly detection from decision authority. A trigger does not automatically establish that an AI output is incorrect, nor does it automatically justify an override. Instead, it changes the governance status of the output by requiring additional examination. The responsible reviewer can then evaluate the model output, supporting information, explanation quality, assumptions, and potential consequences before deciding whether the output should be accepted, subjected to further review, or referred through the override chain.

This distinction limits the risk of treating automated monitoring as a substitute for professional judgment. DAIF functions as a governance mechanism for directing attention and preserving review discipline. Its contribution to accountability lies in establishing an identifiable transition from routine AI use to enhanced scrutiny and ensuring that the reason for escalation and subsequent human response can be documented (Kroll et al., 2017; Wieringa, 2020).

DFAS-IFRS: Disclosure-Ready Model Output Templates

DFAS-IFRS provides the financial-reporting and disclosure interface within DFAS-AAP (Alaali, 2025b). Its purpose is to structure relevant information concerning AI-assisted financial outputs so that material assumptions, judgments, interventions, and model-related information can be considered within applicable financial-reporting requirements (IFRS Foundation, 2022).

This function is particularly relevant when AI contributes to estimates or judgments involving significant uncertainty, such as fair-value measurement, expected credit losses, impairment assessment, or other model-dependent financial reporting processes (Srivastava & Callen, 2020). The use of AI does not replace the institution's responsibility to support and document material judgments. Accordingly, DFAS-IFRS is designed to preserve information concerning relevant assumptions, model characteristics, uncertainty, human interventions, and override rationales where those elements are material to the reporting process.

For fair-value applications involving significant unobservable inputs, the protocol can support the documentation required to evaluate relevant Level 3 judgments under IFRS 13. In credit-risk applications, the same governance logic can support documentation surrounding model-dependent estimates under IFRS 9 (IFRS Foundation, 2022). The precise disclosure obligation remains determined

by the applicable accounting standard and the facts of the reporting context; DFAS-IFRS does not create additional IFRS requirements or imply formal endorsement by the IFRS Foundation.

Within the broader DFAS-AAP architecture, this component connects the internal governance history of an AI-assisted decision with the information required for financial reporting and audit review (Alaali, 2025b; PCAOB, 2023). The objective is to reduce the separation between model governance and reporting governance by ensuring that material human interventions and decision assumptions remain traceable when they affect reportable financial judgments.

DFAS-Governance: Override Chain of Command

DFAS-Governance establishes the institutional authority structure for intervention in AI-enabled financial decisions. Its central mechanism is the Alaali AI Override Chain of Command (A-AOCC), which defines role-based pathways for reviewing, escalating, authorizing, and documenting decisions to depart from an AI-generated output.

The need for an explicit override structure arises because human oversight is ineffective when authority is ambiguous (European Commission, 2021; ISO, 2022). An employee may identify a questionable model output but lack authority to reject it. Conversely, unrestricted override authority can create its own governance risks by allowing individuals to disregard model outputs without adequate justification or review. A-AOCC therefore separates the ability to identify a concern from the authority to approve a material intervention.

The chain of command can be calibrated according to decision materiality, model risk, and institutional structure (BCBS, 2022; NIST, 2023). Lower-impact matters may be resolved at an operational or model-management level, while decisions with greater financial, regulatory, or strategic consequences may require escalation to senior risk, finance, compliance, audit, or governance functions. The architecture does not prescribe identical organizational titles across institutions; rather, it requires the authority attached to each stage to be defined before an intervention occurs.

Every material override should preserve the identity or institutional role of the initiating and approving actors, the reason for intervention, the information considered, the decision taken, and the time at which the action occurred. Where a third-party AI system prevents internal modification of the underlying model, the override may operate at the output or workflow level through rejection of the recommendation, additional human analysis, deployment restrictions, vendor escalation, or suspension of system use (ISO, 2022; NIST, 2023). The mechanism therefore assigns only the authority that the implementing institution can realistically exercise.

DFAS-CICP: Control Integrity and Logging Architecture

DFAS-CICP provides the control-integrity and documentation layer that connects the preceding components of DFAS-AAP (Alaali, 2025d). Its purpose is to preserve a traceable record of material events throughout the AI-enabled decision process and to relate those events to established internal-control responsibilities.

The logging architecture records relevant stages of the decision chain, including model identification, significant inputs where appropriate, generated outputs, escalation events, human reviews, override requests, approvals, decision rationales, and subsequent actions. Time-stamped records allow auditors and control functions to reconstruct the sequence through which a material AI-assisted decision was reached (PCAOB, 2023). The appropriate level of logging should reflect decision materiality, regulatory requirements, confidentiality obligations, data-protection constraints, and the technical capabilities of the system (European Commission, 2021; NIST, 2023).

DFAS-CICP also provides the principal connection between DFAS-AAP and COSO-based internal control (Alaali, 2025d; COSO, 2013). Responsibility assignment contributes to the control environment; AI-related risk identification contributes to risk assessment; escalation and override procedures constitute control activities; decision records support information and communication; and continuing examination of model behaviour and intervention patterns supports monitoring activities. The protocol therefore uses established internal-control logic to organize AI-specific governance events rather than presenting AI governance as an independent control system.

The architecture must also accommodate technical-access limitations. Institutions using external or closed-source AI systems may be unable to log internal model processes or inspect hidden parameters. In these circumstances, DFAS-CICP focuses on information that the institution can legitimately capture and control, including inputs submitted to the system where permitted, outputs

received, human evaluations, escalation events, authorization records, vendor communications, and final institutional decisions (ISO, 2022; NIST, 2023).

Through this function, DFAS-CICP completes the accountability chain established by the other DFAS-AAP components. DFAS-FEP identifies the model and responsible actors; DAIF identifies circumstances requiring enhanced scrutiny; DFAS-Governance determines who possesses intervention authority; DFAS-IFRS connects material information to reporting requirements; and DFAS-CICP preserves the evidence needed to reconstruct and evaluate the resulting decision process (Alaali, 2025a, 2025b, 2025d).

Illustrative Scenarios and Use Cases

The operational logic of DFAS-AAP can be examined through illustrative scenarios representing high-stakes uses of AI in financial decision-making. The scenarios are designed to demonstrate how the five integrated protocols interact when an AI-generated output requires additional scrutiny, human intervention, documentation, or disclosure. They are hypothetical applications rather than empirical case studies or evidence of validated institutional performance.

Each scenario follows the same accountability sequence: an AI-enabled model generates an output; a defined condition activates enhanced review; the relevant institutional actor evaluates the output; any intervention follows the designated authority structure; and the resulting decision and rationale are documented. This sequence illustrates how DFAS-AAP connects technical model behaviour with institutional accountability rather than treating model monitoring, human oversight, and audit documentation as separate activities (Kroll et al., 2017; NIST, 2023; Wieringa, 2020).

AI Valuation Engine Override

Consider a financial institution using an AI-assisted discounted cash flow model to support asset valuation. During a reporting cycle, the model produces a valuation that differs materially from the institution's existing valuation expectations. The magnitude of the deviation activates a DAIF review condition. This trigger does not establish that the model is incorrect; instead, it requires additional examination of the output, assumptions, and supporting information.

Under AEEF, the valuation output is subjected to enhanced scrutiny. The designated valuation specialist examines the assumptions underlying the result and identifies insufficient sensitivity to material changes in relevant macroeconomic conditions. Because the issue may materially affect the valuation decision, the output is escalated through A-AOCC to the institutional role authorized to approve an intervention.

If the authorized reviewer determines that reliance on the original output is not adequately supported, the institution may reject or adjust the model-assisted recommendation using an independently justified valuation judgment. The intervention, supporting rationale, responsible roles, review sequence, and final decision are recorded through DFAS-CICP (Alaali, 2025d). Where the resulting valuation involves significant unobservable inputs or other material judgments relevant to IFRS 13, DFAS-IFRS structures the associated information for consideration within the applicable reporting and disclosure process (Alaali, 2025b; IFRS Foundation, 2022).

The scenario demonstrates the distinction between automated detection and institutional decision authority. A material deviation initiates scrutiny but does not automatically produce an override. The final intervention remains a documented human decision subject to defined authority and reporting requirements.

ESG Risk Scoring Platform

A financial institution may use an AI-enabled ESG risk-scoring platform to analyze corporate disclosures and identify governance-related risks (Kotsantonis et al., 2016). Suppose the system assigns a high governance-risk classification to a portfolio company based on natural-language processing of available disclosures. The classification is materially inconsistent with other information available to the institution and consequently activates enhanced review under DAIF.

AEEF directs attention to the basis of the classification. Human review identifies that the model's interpretation may be affected by insufficient contextual representation of the relevant institutional or regional governance environment. This finding does not automatically invalidate the model output. Instead, it provides a documented basis for escalating the decision to the institutional function responsible for ESG-related review.

Through A-AOCC, the designated authority evaluates the AI-generated assessment together with contextual evidence and determines whether the original classification should be accepted, supplemented, or overridden. Any intervention is recorded through DFAS-CICP (Alaali, 2025d), including the original output, reason for escalation, evidence considered, responsible reviewer, authorization, and final decision.

This scenario illustrates the role of contextual human judgment where an AI model may process information correctly at a technical level but insufficiently account for the environment in which that information should be interpreted. DFAS-AAP does not presume that human judgment is inherently superior to the model; rather, it requires material disagreement between the system and the institution to pass through an accountable and traceable review process (Kroll et al., 2022; Mökander & Floridi, 2022).

Credit Stress Simulation

Consider a financial institution using a model incorporating Monte Carlo simulation and AI-assisted analytical components to evaluate capital resilience under severe market conditions. The model produces a stress scenario indicating an unexpectedly large deterioration in capital adequacy relative to the institution's established scenario assumptions. A predefined DAIF condition activates additional review because of the materiality of the result (BCBS, 2022).

The responsible risk function examines the model assumptions, distributional inputs, scenario construction, and sensitivity of the output to extreme market conditions. If the review identifies assumptions that require reconsideration, the matter is escalated through A-AOCC to the authorized risk-management role. The reviewer may accept the original scenario, request additional analysis, or authorize an adjustment when supported by documented analytical reasoning.

DFAS-CICP records the original output, escalation trigger, review process, assumptions examined, any approved intervention, and the final institutional decision (Alaali, 2025d). Where the analysis contributes to financial-reporting judgments governed by applicable standards, DFAS-IFRS provides a structured mechanism for preserving information relevant to the associated reporting process (Alaali, 2025b; IFRS Foundation, 2022).

The scenario demonstrates that stress-model governance should distinguish between model-generated evidence and managerial intervention. An extreme output should not be adjusted merely because it appears undesirable. Any modification must be analytically justified, authorized, and traceable. This control is particularly important because discretionary adjustment of stress assumptions can itself create governance risk (BCBS, 2022).

Collectively, the three scenarios demonstrate how DFAS-AAP can structure accountability across different financial applications while preserving a common governance sequence. They illustrate the proposed mechanisms but do not establish their empirical effectiveness. Institutional testing would be required to determine whether implementation improves explainability, intervention quality, auditability, or regulatory outcomes.

Regulatory Mapping

DFAS-AAP operates within an existing regulatory and institutional environment rather than as a substitute for established standards. Its governance mechanisms can therefore be assessed by mapping their functions against relevant requirements concerning human oversight, risk management, documentation, internal control, financial reporting, and accountability **(BCBS, 2022; COSO, 2013; European Commission, 2021; Financial Conduct Authority [FCA], 2019; IFRS Foundation, 2022; NIST, 2023)**. The purpose of this mapping is to identify functional correspondence, not to claim certification, regulatory approval, or automatic compliance. Table 1 summarizes this mapping.

Table 1. Regulatory And Institutional Mapping Of DFAS-AAP

Regulatory or Institutional Framework	Relevant Governance Requirement	DFAS-AAP Mechanism	Nature of Alignment
EU AI Act (European Commission, 2021)	Human oversight, risk management, documentation, transparency, and monitoring	DAIF, AEEF, A-AOCC, DFAS-CICP	Functional alignment with oversight, escalation, documentation, and monitoring requirements

Regulatory or Institutional Framework	Relevant Governance Requirement	DFAS-AAP Mechanism	Nature of Alignment
IFRS 7 (IFRS Foundation, 2022)	Disclosure of financial-instrument risks and relevant risk information	DFAS-IFRS, DFAS-CICP	Supports structured documentation of information relevant to applicable disclosures
IFRS 9 (IFRS Foundation, 2022)	Model-dependent judgments concerning expected credit losses and financial instruments	DFAS-IFRS, DFAS-CICP, A-AOCC	Supports traceability of material assumptions, reviews, and interventions
IFRS 13 (IFRS Foundation, 2022)	Fair-value measurement, valuation techniques, inputs, and Level 3 disclosures	DFAS-IFRS, DFAS-CICP	Supports documentation of model-assisted valuation judgments and material interventions
Basel III and associated supervisory expectations (BCBS, 2022)	Risk governance, model oversight, internal controls, and capital-risk management	DAIF, A-AOCC, DFAS-CICP	Provides escalation, human review, and decision-traceability mechanisms
COSO Internal Control—Integrated Framework (COSO, 2013)	Control environment, risk assessment, control activities, information and communication, and monitoring	DFAS-FEP, DAIF, A-AOCC, DFAS-CICP	Positions AI-specific governance activities within established internal-control functions

The mapping indicates that DFAS-AAP primarily operates as an **implementation architecture** connecting governance expectations that otherwise arise from different regulatory and institutional sources. Human oversight requirements are translated into defined escalation and intervention pathways; documentation requirements are connected to decision logs; internal-control principles are linked to AI-specific control events; and financial-reporting requirements are supported through structured preservation of material model and intervention information (BCBS, 2022; European Commission, 2021; IFRS Foundation, 2022; NIST, 2023).

The relationship should nevertheless be interpreted conservatively. Functional alignment with a regulatory requirement does not establish legal compliance, and the applicability of individual requirements depends on jurisdiction, institution, system classification, and use case. Similarly, DFAS-IFRS does not modify IFRS requirements, and DFAS-AAP does not replace COSO, Basel standards, or applicable AI regulation (BCBS, 2022; COSO, 2013; European Commission, 2021; IFRS Foundation, 2022). Its role is to provide an operational governance layer through which institutions may organize responsibilities and evidence relevant to those external requirements.

This distinction also identifies an important area for subsequent validation. Future institutional applications should evaluate whether the proposed mappings remain operationally feasible under specific regulatory environments and whether additional controls are required for particular financial activities, jurisdictions, or AI architectures.

Implementation Framework

Implementation of DFAS-AAP requires translation of its conceptual architecture into institutional roles, procedures, technical controls, and documentation practices. The protocol is therefore designed as a modular governance structure rather than a uniform technical system. The precise implementation should reflect the institution's size, risk profile, AI architecture, regulatory environment, available technical access, and existing control infrastructure (ISO, 2022; NIST, 2023).

Implementation begins by identifying where AI contributes to consequential financial decisions and determining which institutional actors possess responsibility at each stage of the relevant workflow. DFAS-FEP establishes the initial model classification and responsibility record (Alaali, 2025a). DAIF and AEEF establish the conditions under which enhanced scrutiny is activated. A-AOCC assigns escalation and intervention authority. DFAS-CICP determines the information that must be preserved to reconstruct the decision process (Alaali, 2025d), while DFAS-IFRS connects material model-related information with applicable financial-reporting requirements (Alaali, 2025b).

Operational feasibility is a necessary constraint on this architecture. Institutions differ substantially in technical resources, staffing, governance maturity, and access to AI systems. An institution that develops its own models may exercise direct control over model design, validation, parameters, and deployment. An institution using a closed-source external AI service may instead

control only inputs, acceptance of outputs, human review, workflow integration, contractual requirements, procurement decisions, and continued system use (ISO, 2022; NIST, 2023). DFAS-AAP therefore requires governance authority to be assigned according to the control that an institution actually possesses.

Implementation should also remain proportionate to risk. Requiring identical escalation, documentation, and authorization procedures for every AI-assisted decision could create unnecessary governance burden and operational delays. Model criticality, decision materiality, potential impact, explainability limitations, and regulatory significance should therefore determine the intensity of oversight (European Commission, 2021; NIST, 2023). This risk-based approach allows the architecture to preserve accountability without assuming that every AI interaction requires the highest level of institutional review.

Financial Institutions

Financial institutions constitute the principal organizational setting for DFAS-AAP implementation because they directly integrate AI-generated analysis into valuation, lending, risk management, investment, reporting, and other consequential financial processes (Bussman et al., 2021; Deloitte, 2023a). Implementation should occur across the relevant AI lifecycle rather than being limited to post-decision audit (NIST, 2023).

At the model-entry stage, institutions can use DFAS-FEP to classify AI-enabled systems and identify the human and organizational responsibilities associated with their development, procurement, validation, deployment, and use (Alaali, 2025a). This classification provides the basis for determining the level of monitoring and escalation appropriate to each application.

During operation, DAIF can be incorporated into existing model-risk and control processes to identify outputs or conditions requiring additional scrutiny. AEEF then structures the escalation process, while A-AOCC specifies which organizational roles possess authority to review, approve, reject, or escalate a material AI-assisted decision. The relevant roles may differ among institutions and applications; responsibility should therefore be assigned according to actual organizational authority rather than prescribed job titles (BCBS, 2022; NIST, 2023).

DFAS-CICP supports implementation by preserving the decision trail needed for internal control and audit (Alaali, 2025d; COSO, 2013). Institutions should determine which information can legitimately and technically be recorded, including relevant model outputs, escalation triggers, human reviews, intervention rationales, approvals, and final decisions. Where AI-generated analysis contributes to financial-reporting judgments, DFAS-IFRS can structure relevant information for consideration under applicable reporting requirements (Alaali, 2025b; IFRS Foundation, 2022).

Implementation must also account for resource and vendor constraints. Smaller institutions may rely on existing risk, compliance, audit, and finance functions rather than establishing dedicated AI-governance units. Institutions using third-party systems may need to supplement technical controls with contractual requirements, vendor documentation, procurement standards, service-level obligations, independent review, and the ability to restrict or suspend system use (ISO, 2022; NIST, 2023). Governance design should therefore reflect what the institution can realistically control.

The implementation objective is not to maximize the number of controls but to establish a proportionate and traceable accountability chain. For each material AI-assisted financial decision, the institution should be able to identify the system involved, the responsible organizational actors, the conditions requiring escalation, the authority to intervene, the rationale for any intervention, and the evidence preserved for subsequent review (European Commission, 2021; NIST, 2023).

Auditors

Internal and external auditors represent a second implementation layer because AI-enabled financial decisions increasingly affect processes that require assurance over controls, judgments, estimates, and supporting evidence (PCAOB, 2023). Within DFAS-AAP, the auditor's role is not to assume responsibility for the underlying AI decision but to evaluate whether the institution has established and followed an appropriate accountability process.

DFAS-CICP provides the principal audit interface by preserving records of model identification, escalation events, human reviews, override decisions, approvals, and decision rationales (Alaali, 2025d). These records enable auditors to reconstruct the sequence through which a material AI-assisted decision was produced and determine whether relevant control procedures were followed

(COSO, 2013; PCAOB, 2023). The availability of time-stamped and role-specific documentation can also support evaluation of whether intervention authority was exercised consistently with the institution's established governance structure.

Auditors can examine whether model classification under DFAS-FEP corresponds to the level of oversight applied to the system and whether escalation conditions under DAIF and AEEF were appropriately activated and resolved (Alaali, 2025a). Where an override occurred, the audit process can assess whether the intervention was performed by an authorized role under A-AOCC, supported by documented reasoning, and subjected to any additional approval required by the institution's control structure.

Where AI-assisted outputs contribute to financial-reporting judgments, DFAS-IFRS can provide structured information concerning relevant assumptions, model outputs, human interventions, and decision rationales (Alaali, 2025b). This information may support audit procedures concerning estimates and disclosures, but it does not replace the auditor's professional judgment, applicable auditing standards, or the need to obtain sufficient appropriate audit evidence (PCAOB, 2023).

The extent of audit review should remain proportionate to materiality and risk. Auditors may also face technical limitations when institutions rely on proprietary or third-party AI systems. In such cases, assurance may need to focus on the controls surrounding system selection, input governance, output review, human authorization, vendor documentation, and institutional use rather than direct inspection of inaccessible model components (ISO, 2022; NIST, 2023). DFAS-AAP therefore supports auditability by preserving evidence at the points where institutional control and review can realistically be exercised.

Regulators

Regulators and supervisory authorities constitute an external governance layer for AI-enabled financial systems. Their role differs from that of institutions and auditors because regulatory oversight focuses on whether organizations satisfy applicable legal, prudential, reporting, consumer-protection, and governance requirements (BCBS, 2022; European Commission, 2021; Hacker, 2023). DFAS-AAP can support this function by organizing evidence concerning how institutions classify, monitor, review, override, and document material AI-assisted decisions.

DFAS-CICP can provide structured records of significant governance events, including escalation triggers, human interventions, authorization pathways, and final institutional decisions (Alaali, 2025d). Such documentation may assist supervisory review by allowing regulators to examine whether human oversight exists in operational practice rather than solely in formal policies (European Commission, 2021; NIST, 2023). DAIF and AEEF can further demonstrate how the institution identifies circumstances requiring additional scrutiny, while A-AOCC can clarify the allocation of decision authority and escalation responsibilities.

DFAS-IFRS can support traceability where AI-assisted models contribute to financial-reporting judgments subject to applicable IFRS requirements (Alaali, 2025b; IFRS Foundation, 2022). Similarly, DFAS-FEP can provide information concerning model classification and institutional responsibility that may assist regulators in evaluating whether governance intensity is proportionate to the nature and significance of the AI application (Alaali, 2025a; NIST, 2023).

The protocol does not presume that regulators require unrestricted access to proprietary model architecture, source code, or confidential data. Regulatory visibility should remain consistent with applicable legal authority, confidentiality requirements, data-protection obligations, and supervisory practice (European Commission, 2021; ISO, 2022). Where technical access is restricted, governance evidence can instead focus on model documentation, institutional controls, review procedures, intervention records, validation evidence, and decision outcomes.

DFAS-AAP should therefore be understood as a potential mechanism for organizing regulatory evidence rather than as a substitute for regulation or a guarantee of compliance. Its practical value depends on whether its records and governance mechanisms correspond to the requirements applicable to a particular institution, jurisdiction, and AI use case (BCBS, 2022; European Commission, 2021; NIST, 2023). Regulatory pilot applications and supervisory evaluation would be required before conclusions could be drawn regarding its effectiveness as a standardized regulatory governance mechanism.

Journals and Academic Publishers

Although DFAS-AAP is primarily designed for financial institutions, elements of its accountability logic can also be applied to academic research involving AI-enabled financial modelling. Journals and academic publishers increasingly encounter manuscripts in which AI contributes to model development, analytical procedures, data processing, interpretation, or manuscript preparation. These uses create questions concerning authorship, disclosure, reproducibility, and responsibility for AI-assisted research outputs.

Within this context, DFAS-FEP provides the most directly applicable component of DFAS-AAP (Alaali, 2025a). Its classification and authorship-governance logic can support structured documentation of the roles performed by human researchers and AI systems during model development and research production. Such documentation can help distinguish substantive human intellectual contribution from AI-assisted functions while preserving human responsibility for the validity, interpretation, and reporting of the research.

For manuscripts involving AI-generated or AI-assisted financial models, publishers could request information concerning the model used, the purpose of AI involvement, the extent of human validation, material modifications or overrides, and the responsibility of the listed authors for the final analytical output. Where relevant, decision and modification records may also improve reproducibility by clarifying how AI-assisted outputs were evaluated before inclusion in the published research.

The application of DFAS-AAP principles to academic publishing should remain consistent with each journal's authorship, disclosure, research-integrity, and AI-use policies. The protocol does not replace publisher requirements or established authorship standards. Rather, its potential role is to provide a structured accountability mechanism where existing disclosure statements do not capture the full sequence of human and AI involvement in model-based research.

This extension also requires separate evaluation because academic publishing differs institutionally from financial decision-making. The relevant accountability actors, evidentiary requirements, and consequences of AI use are not identical. Accordingly, the application of DFAS-AAP to journals and academic publishers should be treated as a transferable governance use case requiring further validation rather than as an established publishing standard.

Discussion

The analysis indicates that the accountability problem created by AI-enabled financial decision-making cannot be addressed adequately through explainability, human oversight, internal control, or disclosure when these mechanisms operate independently (Floridi & Cowls, 2022; Mökander & Floridi, 2022). The principal contribution of DFAS-AAP is therefore the integration of these functions into a continuous governance sequence connecting model classification, responsibility attribution, explainability escalation, human intervention, documentation, and reporting.

This integration extends the implications of accountability theory in AI-enabled finance. Accountability requires more than identifying an individual who is formally responsible for a decision (Bovens, 2007; Wieringa, 2020). The responsible actor must also have access to sufficient information, possess appropriate authority, operate within a defined review process, and leave evidence that allows the decision to be subsequently examined (Kroll et al., 2017; Kroll et al., 2022). DFAS-AAP translates these conditions into operational mechanisms by connecting DFAS-FEP, DAIF and AEEF, A-AOCC, DFAS-IFRS, and DFAS-CICP across the decision lifecycle (Alaali, 2025a, 2025b, 2025d).

The proposed architecture also extends the application of agency theory to AI-assisted financial workflows. AI systems do not become accountable agents in the conventional sense; responsibility remains with human and institutional actors. However, reliance on AI can increase information asymmetry and create new layers of analytical delegation (Jensen & Meckling, 1976; Gawer, 2023). The governance response is therefore not to assign responsibility to the technology but to establish controls around the points at which institutions select systems, interpret outputs, authorize decisions, intervene in recommendations, and document resulting actions.

A similar implication emerges from the relationship between DFAS-AAP and COSO. AI governance does not necessarily require the replacement of established internal-control principles. Instead, existing control functions can be operationalized for AI-specific decision events (COSO, 2013; Deloitte, 2023b). Model classification can inform risk assessment; escalation and override procedures can function as control activities; decision records can support information and communication; and

continuing examination of model behaviour and interventions can contribute to monitoring. DFAS-AAP consequently positions AI accountability within established institutional control structures while adding mechanisms designed for the characteristics of AI-enabled decisions.

The illustrative scenarios further demonstrate why governance should distinguish between identifying a potentially problematic output and authorizing an intervention. A deviation, explainability deficiency, or unexpected result may justify enhanced scrutiny without establishing that the model is wrong (Doshi-Velez & Kim, 2017; Rudin, 2019). Similarly, human intervention should not automatically be regarded as superior to an AI-generated output. The accountability requirement is that material intervention be supported by evidence, exercised by an authorized actor, and documented sufficiently for subsequent review.

This distinction is particularly important in financial contexts because discretionary overrides can themselves introduce risk. An institution that permits unrestricted human intervention may replace model risk with managerial bias, opportunism, or inconsistent decision-making (Jensen & Meckling, 1976; Malhotra, 2022). A-AOCC therefore functions not simply as a mechanism enabling human control over AI but also as a control over the exercise of human override authority. Accountability applies to both reliance on the model and departure from its output.

The regulatory mapping suggests that DFAS-AAP can provide an operational bridge among governance expectations concerning human oversight, documentation, internal control, risk management, and financial reporting (BCBS, 2022; European Commission, 2021; IFRS Foundation, 2022; ISO, 2022; NIST, 2023). However, functional correspondence should not be interpreted as evidence of legal compliance or regulatory endorsement. Regulatory requirements vary across jurisdictions and applications, and institutional implementation would need to be evaluated against the specific obligations applicable to each use case.

Practical implementation also introduces constraints. Institutions differ in technical capabilities, organizational structures, financial resources, governance maturity, and access to AI systems. Vendor-controlled or closed-source models may prevent direct inspection or modification of internal model processes (Bussman et al., 2021; NIST, 2023). In such cases, accountability must operate through the controls available to the institution, including input governance, output review, human authorization, contractual arrangements, vendor escalation, audit documentation, deployment restrictions, and system suspension (ISO, 2022).

The conceptual nature of the present study constitutes an additional limitation. DFAS-AAP has been developed and examined through theoretical integration, regulatory mapping, and illustrative scenarios rather than through institutional deployment. The analysis can therefore establish conceptual coherence and specify operational relationships, but it cannot demonstrate that adoption improves decision quality, reduces governance failures, increases audit effectiveness, or produces superior regulatory outcomes. These claims require empirical testing.

Positioning DFAS-AAP as an Industry Standard

DFAS-AAP has characteristics that could support future development toward a standardized governance architecture for AI-enabled financial decision-making. Its modular structure identifies specific governance functions, assigns identifiable responsibilities, establishes escalation and intervention pathways, and preserves evidence of material decisions. These characteristics may facilitate consistent implementation across institutions while allowing individual organizations to calibrate controls according to their risk profiles, organizational structures, and regulatory environments (European Commission, 2021; NIST, 2023).

Potential standardization would require a distinction between the conceptual protocol proposed in this study and a formally recognized industry standard. DFAS-AAP is not presently a regulatory standard, certification scheme, or formally accredited assurance framework. Establishing such status would require substantially more evidence and institutional development, including operational specifications, measurable conformity criteria, independent testing, stakeholder consultation, implementation guidance, and external validation (ISO, 2022).

A future standardization process could therefore begin by converting the protocol's conceptual mechanisms into assessable requirements. Model classification criteria would need to be operationally defined; escalation conditions would require application-specific calibration; override authority would need minimum documentation and segregation-of-duty requirements; logging expectations would

require technical and evidentiary specifications; and regulatory mappings would need to be validated within individual jurisdictions.

Pilot implementations would then be necessary to determine whether these requirements are feasible across institutions of different sizes and technological capabilities. Evaluation should consider not only governance benefits but also implementation costs, staffing requirements, training needs, technical integration, vendor dependence, decision delays, documentation burden, and scalability. A governance standard that produces excessive operational friction could undermine adoption even if its conceptual objectives are desirable.

Independent participation would also be essential. Regulators, auditors, financial institutions, technology providers, professional bodies, and academic researchers would need to evaluate whether the proposed requirements are sufficiently clear, proportionate, verifiable, and compatible with existing standards. Such external assessment would reduce dependence on the protocol's originating framework and provide evidence concerning its applicability beyond the context in which it was developed.

Accordingly, DFAS-AAP should presently be positioned as a candidate architecture for future standardization, rather than as an established industry standard. Its potential for standardization derives from the structured and auditable nature of its governance mechanisms, while any future certification or formal standard status remains conditional on empirical validation, independent assessment, and institutional adoption.

Future Research Directions

Future research should prioritize empirical evaluation of DFAS-AAP rather than further conceptual expansion. The first stage should involve pilot implementations within financial institutions to determine whether the proposed governance sequence can be integrated into existing model-risk, internal-control, audit, and financial-reporting processes. Such studies should examine implementation feasibility as well as whether the protocol produces observable improvements in accountability-related outcomes.

Quantitative research could compare governance outcomes before and after implementation using measures such as escalation frequency, time required to resolve flagged outputs, completeness of override documentation, frequency of unauthorized interventions, audit exceptions, and consistency of decision records. These measures would provide evidence concerning whether the proposed mechanisms improve traceability and control rather than merely increasing documentation.

Research should also evaluate the effectiveness of AEEF and A-AOCC independently. AEEF requires testing to determine whether escalation criteria identify genuinely material situations without generating excessive false positives or review burden. A-AOCC should be evaluated for whether role-based intervention improves accountability while avoiding unnecessary decision delays, authority conflicts, or excessive managerial discretion.

Comparative studies across banking, insurance, investment management, fintech, and other financial activities would help determine whether the architecture is transferable or requires sector-specific modifications. Cross-jurisdictional research would similarly be useful because regulatory expectations, institutional structures, data-governance requirements, and supervisory practices differ across markets (BCBS, 2022; European Commission, 2021; Hacker, 2023).

Further research should examine implementation under different technical-access conditions. Institutions operating internally developed AI models may possess substantially greater monitoring and modification capabilities than institutions relying on third-party or closed-source systems (NIST, 2023). Comparing these environments would clarify which DFAS-AAP mechanisms are universally applicable and which depend on particular levels of technical control.

The potential application of selected DFAS-AAP mechanisms to academic publishing also requires independent investigation. Research could examine whether structured AI-use classification, human-validation records, and intervention documentation improve transparency and reproducibility in AI-assisted financial research without creating disproportionate administrative requirements.

Finally, any movement toward formal standardization should be preceded by independent evaluation and stakeholder consultation. Future studies should develop measurable conformity criteria, assess governance costs and benefits, test interoperability with existing standards, and determine whether the protocol can be implemented proportionately across institutions with different levels of technological and organizational maturity (ISO, 2022; NIST, 2023).

CONCLUSION

The increasing integration of AI into financial decision-making creates an accountability problem that extends beyond model performance and technical explainability (Kroll et al., 2017; Mökander & Floridi, 2022; Rudin, 2019). When AI-generated outputs influence valuation, risk assessment, credit decisions, financial reporting, or other consequential activities, institutions must preserve identifiable responsibility for reviewing those outputs, determining when intervention is necessary, authorizing any override, and documenting the resulting decision.

This study addressed the research question of how accountability, human override authority, explainability escalation, internal control, and disclosure requirements can be integrated into a coherent governance protocol for AI-enabled financial decision-making. DFAS-AAP responds by connecting five governance components within a continuous accountability architecture: DFAS-FEP for model classification and responsibility attribution (Alaali, 2025a), DAIF and AEEF for enhanced scrutiny and escalation, DFAS-IFRS for structured reporting-related information (Alaali, 2025b), DFAS-Governance and A-AOCC for role-based intervention authority, and DFAS-CICP for control integrity and decision traceability (Alaali, 2025d).

The analysis indicates that the principal conceptual value of this architecture lies in the connections among these mechanisms. Explainability becomes operationally relevant when it can trigger accountable review; human oversight becomes meaningful when authority is defined; override becomes governable when intervention is justified and documented; and internal control becomes applicable to AI when model-related events can be incorporated into established risk, control, information, and monitoring processes (COSO, 2013; NIST, 2023).

DFAS-AAP should nevertheless be interpreted within the limits of the study. The protocol has been developed through conceptual analysis, theoretical integration, regulatory mapping, and illustrative scenarios rather than empirical institutional testing. The study therefore does not establish that DFAS-AAP improves financial performance, regulatory compliance, audit quality, or decision accuracy. Nor does alignment with existing regulatory and institutional frameworks constitute certification or formal endorsement (BCBS, 2022; European Commission, 2021; IFRS Foundation, 2022; ISO, 2022).

The next stage of research should consequently focus on pilot implementation, measurement of governance outcomes, independent assessment, and comparison across financial sectors and jurisdictions. Such evidence would determine whether the proposed architecture can move from a conceptually integrated accountability protocol toward a validated and potentially standardized governance mechanism for AI-enabled finance.

REFERENCES

Alaali, H. M. H. (2025a). *DFAS-FEP:* A Global Governance Standard for Responsible AI Use and Authorship Integrity in Financial Modelling. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5260517

Alaali, H. M. H. (2025b). *DFAS-IFRS* Code of Ethics: Governing Artificial Intelligence in Financial Reporting, A Global Ethics Overlay for Model Classification, Override Accountability, and Hallucination Prevention. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5268372

Alaali, H. M. H. (2025c). The DFAS-FEP Compliance Protocol (DFAS-CP): Operationalizing Ethical AI Enforcement in Financial Modelling. SSRN. https://ssrn.com/abstract=5315167

Alaali, H. M. H. (2025d). DFAS-CICP: A Dynamic Control & Internal Control Protocol for AI-Embedded Financial Ecosystems. SSRN. DOI: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5367894

Basel Committee on Banking Supervision. (2022). *Principles for the sound management of operational risk*. Bank for International Settlements.

Bovens, M. (2007). Analysing and assessing public accountability: A conceptual framework. *European Law Journal, 13*(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Brynjolfsson, E., & McAfee, A. (2021). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.

Bussman, O., Wagner, A., & Leicht, R. (2021). AI in finance: Risks and challenges. *Journal of Risk Management, 14*(2), 33–50.

Committee of Sponsoring Organizations of the Treadway Commission. (2013). *Internal control—Integrated framework*. COSO.

Deloitte. (2023a). *AI in financial services: Trends, risks, and opportunities*. Deloitte Insights.

Deloitte. (2023b). *AI risks and internal controls: Extending COSO to emerging technologies*. Deloitte Insights.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*. https://arxiv.org/abs/1702.08608

European Commission. (2021). *Proposal for a regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act)*. COM/2021/206 final.

Financial Conduct Authority. (2019). *Algorithmic trading compliance in wholesale markets: FCA supervisory observations*. FCA. https://www.fca.org.uk/publication/multi-firm-reviews/algorithmic-trading-compliance-wholesale-markets.pdf

Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. *Harvard Data Science Review, 4*(1).

Gawer, A. (2023). Platforms, agency, and algorithmic governance: Emerging challenges in digital markets. *Journal of Management Studies, 60*(2), 312–339.

Hacker, P. (2023). Regulating AI in Europe: Between risk regulation and innovation policy. *European Law Journal, 29*(1), 1–24.

IFRS Foundation. (2022). *IFRS standards and emerging technologies: Guidance for preparers*. IFRS Foundation.

International Organization for Standardization. (2022). *ISO/IEC 38507:2022 — Governance implications of the use of artificial intelligence by organizations*. ISO. https://www.iso.org/standard/81230.html

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics, 3*(4), 305–360. https://doi.org/10.1016/0304-405X(76)90026-X

Kotsantonis, S., Pinney, C., & Serafeim, G. (2016). ESG integration in investment management: Myths and realities. *Journal of Applied Corporate Finance, 28*(2), 10–16.

Kroll, J., et al. (2017). Accountable algorithms. *University of Pennsylvania Law Review, 165*(3), 633–705.

Kroll, J. A., Mittelstadt, B. D., & Wachter, S. (2022). Contextual accountability: Beyond the algorithmic transparency fallacy. *Regulation & Governance*. Advance online publication.

Malhotra, Y. (2022). Model risk and governance in AI-driven financial architectures. *Journal of Financial Regulation and Compliance, 30*(1), 55–78.

Mittelstadt, B. D., Wachter, S., & Floridi, L. (2023). From diagnostic to interventionist: Algorithmic impact assessments and the governance of AI. *AI and Ethics, 3*(2), 367–384.

Mökander, J., & Floridi, L. (2022). AI governance: A research agenda. *AI and Ethics, 2*, 521–533.

National Institute of Standards and Technology. (2023). *AI risk management framework (NIST AI RMF 1.0)*. U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

Public Company Accounting Oversight Board. (2023). *Emerging risks in auditing automated decision-making systems*. PCAOB.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Srivastava, A., & Callen, J. L. (2020). Machine learning and financial reporting quality. *Journal of Accounting and Economics, 69*(2–3), 101276.

Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 1–18). https://doi.org/10.1145/3351095.3372833

Yermack, D. (2018). Corporate governance and blockchains. *Review of Finance, 21*(1), 7–31. https://doi.org/10.1093/rof/rfw074