Strategic Management, Business, and Economic Review Vol. 1 No. 2 (2026) https://doi.org/10.31098/smber.v1i2.4606

Research Paper

DFAS-ForensicGov: A Doctrinal Framework for Ethical, Explainable, and Sovereign-Sensitive AI Governance in Forensic Accounting

Hasan Mohamed Husain Alaali
DFAS Research Center, Kingdom of Bahrain

Received: June 6, 2026	Revised: August 21, 2026	Accepted: August 25, 2026	Online: August 27, 2026

Abstract

Artificial intelligence (AI) is increasingly used in forensic accounting, but AI-generated outputs can create explainability, accountability, evidence-integrity, and cross-border data-governance risks. This study uses a structured conceptual analysis of forensic-accounting literature, AI-governance and AI-auditing research, professional auditing standards, and institutional governance frameworks. The analysis identifies an operational governance gap between general AI governance principles and the requirements for governing AI-generated forensic outputs before they influence consequential institutional decisions. DFAS-ForensicGov is an integrated architecture comprising the Alaali Forensic Governance Doctrine, Alaali Forensic Explainability Escalation Framework, Alaali Forensic Override Command Chain, Alaali Evidence Integrity Calibration Engine, and Alaali Forensic Integrity Ledger. The doctrine establishes a governance sequence in which evidence-pathway conditions are assessed before and throughout AI-assisted processing, while consequential use is governed by explainability, escalation, authorized human intervention, and traceable documentation. It positions AI outputs as investigative inputs requiring accountable human judgment rather than self-validating evidence. The framework is conceptual and has not yet been empirically validated. Future pilot implementation and comparative testing must validate its practical effectiveness.

Keywords: *Artificial Intelligence Governance, Forensic Accounting, Conceptual Framework, AI Auditing, Explainable AI, Human Oversight, Evidence Integrity, Digital Sovereignty*

INTRODUCTION

Artificial intelligence (AI) is increasingly embedded in forensic accounting activities, including anomaly detection, transaction monitoring, document review, relationship mapping, fraud-risk prioritization, and large and heterogeneous financial dataset analysis. These applications can improve the speed and analytical reach of investigations, particularly where conventional manual procedures cannot efficiently process complex financial evidence. Recent research also indicates that AI capability, legal considerations, and ethical safeguards are closely connected to the effectiveness of cyber-forensic accounting practices (Huy & Phuc, 2025).

However, the use of AI in forensic investigations creates a governance problem that is distinct from the general use of AI in ordinary business processes. A forensic output may affect disciplinary action, regulatory reporting, litigation strategy, asset recovery, reputational outcomes, or criminal referral. In such settings, an AI-generated risk score or anomaly flag cannot be treated merely as a technical recommendation. Its evidentiary basis, interpretability, reliability, legal admissibility, and the identity of the person who authorizes the action become institutional questions.

Existing AI governance frameworks provide important foundations. The NIST AI Risk Management Framework identifies validity, reliability, safety, security, accountability, transparency, explainability, privacy, and fairness as central characteristics of trustworthy AI (National Institute of Standards and Technology [NIST], 2023). Likewise, AI-auditing scholarship emphasizes that organizations need structured mechanisms to assess whether AI systems conform to technical, legal and ethical requirements rather than relying on informal assurance alone

Corresponding author's email: founder@dfas.ai, Alternative Email: hasan.mohd.alaali@gmail.com

(Mökander, 2023). Although these frameworks are valuable, they are not designed specifically around the evidentiary and procedural conditions of forensic accounting investigations.

The same limitation appears in professional auditing standards. The International Standard on Auditing 240 establishes responsibilities relating to fraud, while ISA 315 addresses the identification and assessment of material-misstatement risks (International Auditing and Assurance Standards Board [IAASB], 2021a, 2021b). These standards remain essential foundations for professional skepticism, documentation, audit evidence, and risk assessment. However, they do not provide an integrated operational model for responding when an AI-generated forensic output is insufficiently explainable, produces a potentially harmful false positive, requires human override, or involves data processing across jurisdictions with different privacy and evidentiary rules.

This gap is significant because forensic accounting involves a higher threshold of institutional defensibility than routine analytical work. An investigator must be able to explain not only what an AI system identified but also why the output was considered reliable enough to influence a decision, what additional evidence was examined, who retained authority to challenge or reject the result, and whether the processing pathway complied with applicable legal and regulatory requirements. Research on responsible AI has repeatedly identified a broader gap between high-level principles and enforceable organizational practices (Morley et al., 2020). In forensic accounting, the need to preserve evidence integrity, procedural fairness, professional accountability, and the legal defensibility of investigative decisions intensifies that gap.

Cross-border investigations add a layer of complexity. Financial evidence may be held in cloud environments, processed by external technology providers, or transferred across jurisdictions with different data protection requirements, regulatory expectations, and rules concerning evidence access or admissibility. Therefore, governance cannot be limited to model accuracy or bias testing. It must also address the conditions under which data may be processed, how intervention decisions are documented, and how institutional authority is exercised when technological efficiency conflicts with legal, evidentiary, or sovereign requirements.

This study addresses this problem by proposing DFAS-ForensicGov: a conceptual governance doctrine for AI-enabled forensic accounting. Its central contribution is not another general set of AI ethics principles. Instead, an operational governance architecture is developed that places defined control mechanisms between AI-generated analysis and consequential forensic action. The doctrine comprises a core governance anchor and four complementary mechanisms: an explainability escalation framework, a formal human-override command chain, an evidence-integrity calibration mechanism, and an optional traceability ledger. These components govern when AI outputs may be relied upon, when additional review is required, who may intervene, and how governance decisions are preserved for audit, regulatory, or judicial scrutiny.

Accordingly, this study has three objectives. First, the operational governance requirements created by AI-assisted forensic investigations are identified. Second, a structured doctrine that translates explainability, human oversight, evidence integrity, traceability, and jurisdiction-sensitive compliance into institutional mechanisms is developed. Third, it positions the doctrine against existing AI governance and auditing frameworks through comparative analysis and illustrative forensic scenarios. The study is guided by the following questions:

1. What governance gap remains between existing AI governance and auditing frameworks and the specific requirements of AI-enabled forensic accounting?
2. How can explainability, human intervention, evidence integrity, and jurisdiction-sensitive compliance be translated into an operational forensic-governance architecture?
3. How does DFAS-ForensicGov complement existing professional and regulatory frameworks without replacing them?

LITERATURE REVIEW

Artificial Intelligence in Forensic Accounting

Artificial intelligence is increasingly being used to support forensic accounting through anomaly detection, transaction screening, document analysis, relationship mapping, and the prioritization of potentially suspicious activities. These capabilities are particularly relevant when investigations involve high transaction volumes, unstructured documents, multiple entities, or complex cross-border financial relationships. AI can increase investigative capacity by directing human attention toward patterns that conventional procedures may not readily identify.

Recent scholarship also identifies digitalization and corporate governance as central emerging streams in forensic accounting research, reinforcing the need to examine technological capability alongside institutional control arrangements (Cicchini et al., 2026). Recent empirical research connects AI capability with cyber-forensic accounting performance and demonstrates that legal and ethical considerations remain central to the effectiveness of AI-enabled forensic work (Huy & Phuc, 2025). However, improved analytical performance alone does not establish that an AI output should be relied upon in a consequential investigation.

A system may identify a suspicious pattern without providing sufficient explanation for an investigator, supervisor, regulator, or court to understand its basis. In forensic settings, this limitation is significant because a false positive may trigger unjustified scrutiny, reputational harm, or inappropriate disciplinary action, whereas a false negative may allow misconduct to remain undetected. Therefore, AI-assisted detection must be distinguished from evidence sufficient to support institutional action.

Forensic accounting research is increasingly linking investigative effectiveness with internal control, governance, and institutional accountability. Alzoubi (2025), for example, identified a close relationship between forensic accounting and corporate-governance effectiveness. More broadly, recent accounting research calls for a structured research agenda that addresses the implications of artificial intelligence (AI) for accounting practice, professional judgment, and institutional decision-making (Stratopoulos & Wang, 2025). However, the literature does not adequately specify how AI-generated investigative outputs should be governed before they affect institutional decisions. The unresolved issue is not whether AI should support forensic analysis but under what operational conditions an organization may rely on, challenge, suspend, recalibrate or reject an AI-generated forensic output.

From AI Ethics Principles to Operational Governance

The wider AI-governance literature has established transparency, accountability, fairness, explainability, privacy, robustness, and human oversight as central requirements for trustworthy AI. The OECD AI Principles promote AI that respects human rights and democratic values while requiring responsible stewardship by governments and AI actors (Organisation for Economic Co-operation and Development [OECD], 2024). Similarly, the (National Institute of Standards and Technology [NIST], 2023), Artificial Intelligence Risk Management Framework identifies explainability, accountability, privacy, fairness, validity, reliability, safety, and security as interconnected characteristics of trustworthy AI governance.

These principles are necessary but do not create an operating procedure for consequential decision settings. Morley et al. (2020) identified a persistent gap between high-level AI ethics principles and the practical tools required to embed those principles in organizational decision-making. More fundamentally, AI governance cannot be reduced to a model's technical properties alone. The governance significance of an output depends on the institutional setting in which it is

used, the authority assigned to human reviewers, the evidence available for contestation, and the consequence of the decision that may follow. As argued by Selbst et al. (2019), abstracting technical systems from their social and institutional contexts can obscure the actual sources of harm and accountability failure.

Algorithmic auditing research provides a more operational direction. Raji et al. (2020) proposed an end-to-end model of internal algorithmic auditing based on documentation and lifecycle assessment. Mökander (2023) conceptualized AI auditing as a structured assessment of whether an AI system complies with technical, legal, or ethical requirements. These contributions establish the importance of traceable governance, independent assessment, and lifecycle controls. However, their primary concern is whether an AI system can be assessed or governed in general terms; they do not provide a forensic-accounting operating architecture for deciding how a particular AI-generated output should move, or be prevented from moving, toward consequential institutional action.

This limitation is significant in forensic accounting. The relevant question is not simply whether an AI system meets general governance expectations. The relevant question is whether a specific output may be relied upon in an investigation that could affect an individual, organization, regulator, or legal process. This requires an operational response to four connected questions: whether the output is sufficiently explainable for its intended use, whether escalation is required, who has authority to intervene, and whether the evidence pathway is legally and institutionally defensible.

DFAS-ForensicGov addresses this operational problem by translating dispersed governance requirements into a forensic decision architecture. It does not replace existing AI governance, auditing, or regulatory frameworks. Its contribution is to specify how explainability assessment, escalation, human intervention, evidence-pathway control, and traceability are connected when AI-generated analysis may influence consequential institutional action.

Explainability, Human Oversight, and Accountability

Explainability is particularly important when AI is used to generate outputs that may influence consequential decisions. If investigators cannot understand the material basis, relevant limitations, uncertainty, and evidentiary context of an AI-generated finding, they cannot exercise meaningful professional judgment over that finding. However, explainability should not be treated as a guarantee of reliable or legitimate decision-making. Interpretability tools may assist users, but their practical value depends on whether users can understand, appropriately calibrate, and act upon the information they provide (Kaur et al., 2020). In forensic accounting, therefore, explainability is a decision condition: it determines whether an output can proceed toward reliance, requires additional review, or must be withheld from consequential use.

Human oversight performs a distinct function. Explainability concerns whether a reviewer has a sufficiently usable basis for understanding and questioning an output. Human oversight concerns whether an authorized person has the competence, independence and institutional authority to decide what should happen after the assessment. The existence of a human reviewer is not, by itself, a meaningful oversight. Research on AI-assisted decision-making demonstrates that unless the decision environment is designed to encourage active evaluation and justified intervention, human users may over-rely on algorithmic recommendations (Buçinca et al., 2021). Consequently, human oversight requires more than nominal approval: it requires defined authority to challenge, suspend, modify, reverse, or reject an AI-generated output.

The European Union's Artificial Intelligence Act reflects this distinction by requiring high-risk AI system deployers to interpret and use outputs appropriately, alongside human-oversight measures that enable authorized persons to understand, monitor, intervene in, or stop systems

where necessary (European Parliament & Council of the European Union, 2024). Although the Act is not a forensic-accounting standard, its logic is directly relevant: high-consequence AI use requires both usable information about system limitations and identifiable authority capable of acting on that information.

DFAS-ForensicGov separates these functions to avoid treating explainability and oversight interchangeably. A-FEEF is an assessment and escalation mechanism. It classifies the explainability condition of an AI-generated forensic output and determines whether the output may proceed with a routine review, require additional corroboration, or be escalated. A-FOCC is an authority and an intervention mechanism. It determines who may review, challenge, modify, suspend, or reject an escalated output and requires the rationale, evidence, timing, and responsible authority to be documented. Therefore, A-FEEF answers, "Does this output meet the conditions for its intended use?" A-FOCC answers, "Who has the authority to act when it does not, and what intervention is required?"

Existing auditing standards reinforce the importance of professional judgment, fraud-risk assessment, and adequate and appropriate audit evidence. ISA 240 and ISA 315 establish responsibilities concerning fraud and risk assessment, but they do not specify how an organization should govern an AI-generated fraud signal that is technically complex, partially explainable or inconsistent with contextual evidence (IAASB, 2021a, 2021b). This creates a specific implementation gap between general audit responsibilities and AI-enabled investigative practice. DFAS-ForensicGov addresses this gap by connecting explainability assessment to authorized intervention rather than assuming that either technical transparency or human presence alone establishes accountable governance.

Evidence Integrity and Jurisdiction-Sensitive Governance

AI-enabled forensic investigations often rely on cloud services, external vendors, shared databases, and data transfers across jurisdictions. These arrangements can create data protection, confidentiality, regulatory authority, chain of custody, and evidentiary admissibility risks. Consequently, evidence integrity is not limited to whether data have been altered; it also concerns whether data were accessed, processed, documented, and retained in a manner consistent with applicable legal and institutional requirements.

Current AI-governance frameworks recognize data governance and record-keeping as important requirements. The NIST AI RMF emphasizes organizational risk management across the AI lifecycle, while the EU AI Act requires relevant documentation, logging, transparency and human-oversight arrangements for high-risk systems (NIST, 2023; European Parliament & Council of the European Union, 2024). However, neither framework is designed to govern the specific evidentiary lifecycle of forensic accounting: from AI-assisted identification to human review and to potential use in internal disciplinary, regulatory, or judicial processes.

This limitation is particularly important in cross-border investigations. A technically accurate AI output may still be institutionally unusable if the underlying data were processed through a non-compliant infrastructure, if the processing pathway cannot be documented, or if legal restrictions prevent the resulting evidence from being relied upon. Therefore, forensic AI governance must address both the analytical validity and institutional legitimacy of the evidence pathway. This requirement is consistent with Alaali's (2026) sovereign-sensitive governance perspective for AI-enabled financial systems.

Synthesis and Research Gap

The literature establishes four relevant foundations for AI-enabled forensic accounting. First, AI can strengthen investigative capability through anomaly detection, transaction screening,

document analysis, relationship mapping, and the prioritization of suspicious activity. Second, responsible-AI scholarship identifies explainability, accountability, transparency, human oversight, and documentation as core requirements for the use of trustworthy AI. Third, AI-auditing research supports lifecycle assessment, structured conformity evaluation, and traceable governance controls. Fourth, professional auditing standards, financial-crime frameworks, and AI regulation recognize the importance of fraud-risk assessment, sufficient evidence, data governance, record-keeping and accountable human control.

Therefore, the research gap is not the absence of relevant governance principles or the lack of adoption of AI in forensic accounting. The unresolved problem lies at the point of transition between an AI-generated forensic output and a consequential institutional decision. When an AI system identifies a suspicious pattern, assigns a risk score, or recommends investigation, an organization must determine whether the output is sufficiently explainable for its intended use, whether the evidence pathway is legally and institutionally defensible, who has authority to intervene, and how the resulting decision must be documented.

Existing bodies of guidance address these questions in a fragmented form. AI-governance frameworks establish broad trustworthy AI requirements; AI-auditing literature supports assessment and documentation; auditing standards govern professional judgment and evidence; and financial-crime frameworks address risk and institutional cooperation. However, they do not provide a forensic-specific operating architecture that connects AI-assisted detection, explainability assessment, authorized human intervention, evidence-pathway legitimacy, traceability and final institutional authorization.

Thus, the central theoretical gap is an operational-governance gap: the absence of a mechanism-based architecture for governing the movement from analytical output to institutionally consequential action. DFAS-ForensicGov addresses this gap by translating established requirements into differentiated and connected mechanisms. A-FEEF governs explainability-based escalation, A-FOCC allocates authority for human intervention, A-EICE assesses the legality and institutional defensibility of the evidence pathway before and throughout AI-assisted processing, and A-FIL preserves the governance record for consequential use. Together, these mechanisms establish a sequenced pathway from AI-generated analysis to accountable institutional authorisation.

The doctrinal foundation builds on the DFAS AI Governance Doctrine Compendium, while the DFAS Governance Deployment Roadmap informs the institutional implementation dimension (Alaali, 2025a, 2025b). The next section explains the conceptual research method used to construct and assess the proposed architecture.

RESEARCH METHOD

This study adopts a structured conceptual research design to develop DFAS-ForensicGov as an operational governance doctrine for AI-enabled forensic accounting. This study does not statistically test causal relationships, estimate the predictive accuracy of an AI model, or report findings from a field experiment. Its purpose is theory development: to identify governance problems that are insufficiently addressed by existing literature and professional frameworks, specify relevant constructs and mechanisms, and integrate them into a coherent operational architecture.

This design follows the model-building approach to conceptual research. Conceptual articles require an explicit research design that identifies how theoretical ingredients are selected, analyzed, and combined to produce a defensible contribution (Jaakkola, 2020). Accordingly, the study develops DFAS-ForensicGov through theory synthesis and adaptation. Theory synthesis is used to integrate insights from AI governance, AI auditing, forensic accounting, auditing standards,

and cross-border data governance. Theory adaptation is used to translate these insights into AI-assisted forensic investigations' specific institutional setting.

Source Identification and Selection

The conceptual analysis drew on four source categories: peer-reviewed research on AI governance, explainability, accountability, human oversight, AI auditing, forensic accounting, cyber-forensic accounting, and AI-assisted decision-making; professional auditing and assurance standards addressing fraud responsibilities, risk assessment, audit evidence, professional judgment, and documentation; institutional AI-governance frameworks and regulatory instruments addressing trustworthy AI, transparency, documentation, accountability, human oversight, and data governance; and applied forensic, legal, and governance literature addressing evidence integrity, internal control, institutional accountability, data processing, cloud infrastructure, and cross-border investigative risks.

The review concentrated primarily on literature and institutional materials published or substantially updated between 2019 and 2026. This period was selected because it captures the expansion of AI-governance scholarship, the development of operational AI-auditing approaches, and the emergence of major regulatory instruments addressing high-consequence AI use. Earlier sources were retained only where they provided foundational conceptual guidance relevant to the development of theory or professional auditing responsibilities.

Sources were identified through structured searches of Scopus, Web of Science, and Google Scholar, supplemented by searches of official repositories maintained by recognized standard-setting bodies, regulators, and intergovernmental organizations. The search strategy combined three concept blocks: AI and governance; forensic, auditing, or investigative settings; and institutional control conditions. The principal Boolean search logic was as follows:

> ("artificial intelligence" OR AI OR algorithmic) AND (governance OR accountability OR explainab* OR transparency OR "human oversight" OR "algorithmic auditing" OR "AI auditing")
>
> This core string was combined, where relevant, with
>
> ("forensic accounting" OR "forensic investigation" OR audit* OR "financial crime" OR "audit evidence" OR "AI-assisted investigation")
>
> and:
>
> ("evidence integrity" OR "traceability" OR "data governance" OR "cross-border data transfer" OR "sovereignty" OR "institutional accountability").

The search strings were adapted to the syntax of each database and search engine. Reference lists of highly relevant academic articles, auditing standards, regulatory instruments, and institutional frameworks were also reviewed to identify foundational and directly relevant materials.

Each potentially relevant source was coded using a structured screening record. The coding fields were source category, publication or update date, authority status, substantive topic, addressed governance construct, relevance to consequential forensic decisions, contribution to a proposed mechanism, and screening outcome. The sources were classified as retained, contextual only, or excluded.

A retrospective screening audit was conducted to document the source-selection pathway. The searches produced 56 records. Six duplicate records were removed, leaving 50 unique records for the initial screening. At this stage, 21 records were excluded because they were outside the scope of forensic governance, lacked sufficient scholarly or institutional authority, or were general commentary, vendor guidance, or non-verifiable materials. Therefore, 29 sources were flagged as potentially relevant and assessed in full. After a substantive assessment, eight of these sources were rejected because they did not directly contribute to the defined governance problem, a proposed mechanism, a boundary condition, or the doctrine's comparative positioning. The final synthesis retained 21 sources. Accordingly, 8 of the 29 flagged sources were rejected at full-text assessment, representing a rejection proportion of 27.6%.

Table 1. Source Identification and Screening Process

Review stage	Number of sources	Decision or basis
Records identified through structured academic and institutional searches	56	Initial retrieval set
Duplicate records removed	6	Repeated or substantially identical
Unique screened records	50	Title, abstract, authority, and screening of topical-relevance
Exclusion of records at initial screening	21	Outside scope, insufficient authority, general commentary, vendor guidance, or non-verifiable material
Sources flagged as potentially relevant and fully assessed	29	Direct relevance to the governance problem or a proposed mechanism
Sources rejected after the substantive assessment	8	No direct contribution to the problem, mechanism, boundary, or comparative assessment of the doctrine
Retained sources in the final conceptual synthesis	21	The final evidentiary corpus
Full-text rejection proportion	27.6%	8 rejected ÷ 29 flagged sources

The final conceptual synthesis retained 21 sources. These included 12 peer-reviewed academic sources, 6 professional, regulatory, or intergovernmental instruments, and 3 prior DFAS doctrinal publications solely used to establish the origin and conceptual development of the proposed architecture.

The retained sources were selected because they directly and demonstrably contributed to the definition of the governance problem, the design of a proposed mechanism, a boundary condition, or the doctrine's comparative positioning. A source identified during screening was rejected from the doctrinal evidence base where it focused solely on technical model performance without governance relevance; addressed AI adoption without implications for decision authority, accountability, evidence integrity, or institutional use; duplicated substantially similar conceptual claims without adding analytical value; lacked sufficient scholarly, institutional, or professional authority; or could not be verified through a reliable source. General commentary, vendor guidance, and non-verifiable materials were not used as a basis for the core mechanisms of the doctrine.

The screening record required a documented decision for every considered source. A source was retained where it met the relevance and authority criteria and made a distinct contribution to the problem definition, mechanism design, boundary conditions, or comparative analysis. A source was classified as contextual only when it informed the general background but did not directly support a doctrine component. A source was excluded if it failed the relevance, authority,

verification, or nonduplication criteria. This procedure ensured that sources initially flagged as potentially relevant were not automatically incorporated into the conceptual synthesis.

The analysis did not treat any single framework, discipline, or prior DFAS publication as sufficient evidence for a proposed mechanism to reduce the risk of selective interpretation. Each proposed mechanism must satisfy three decision rules. First, a clearly identified governance problem arising when an AI-generated output may influence a consequential forensic decision must be addressed. Second, the underlying requirement had to be supported across at least two independent source categories or by one binding authoritative instrument together with directly relevant scholarly or professional support. Third, the mechanism had to perform a distinct operational function not already performed by another doctrine component. The proposed mechanisms that did not satisfy these conditions were rejected, merged, or retained only as contextual observations rather than being incorporated into the final architecture.

Explainability escalation, for example, was examined through AI-governance and human-oversight literature; intervention authority through accountability, auditing, and regulatory materials; and evidence-integrity calibration through forensic, data-governance, and jurisdictional sources. The proposed mechanisms were then assessed against four conceptual evaluation criteria: conceptual coherence, operational specificity, institutional compatibility, and contextual adaptability.

This study is not presented as a systematic literature review, meta-analysis, or exhaustive bibliometric mapping. It is a transparent, problem-driven conceptual synthesis. Its objective is not to estimate the prevalence of a research theme but to identify, compare, and integrate the theoretical and institutional requirements necessary to construct a governance architecture for a defined forensic accounting problem.

Analytical Procedure

The analysis proceeded in four stages. First, the study identified the governance requirements that arise when AI-generated outputs influence consequential forensic decisions. The literature and institutional frameworks consistently highlighted the relevant requirements of explainability, human oversight, accountability, documentation, data governance, and legal compliance.

Second, these requirements were examined against forensic accounting characteristics. This stage assessed the insufficiency of general AI-governance principles when AI outputs may affect fraud allegations, internal disciplinary action, regulatory referral, litigation, or the handling of potentially admissible evidence. The analysis identified four central operational problems: insufficient explainability, unclear intervention authority, uncertain evidence-integrity conditions, and weak traceability of governance decisions.

Third, each problem was translated into a dedicated governance mechanism. The explainability problem informed the Alaali Forensic Explainability Escalation Framework, the intervention authority informed the Alaali Forensic Override Command Chain, the evidence and jurisdictional risk informed the Alaali Evidence Integrity Calibration Engine, and the governance traceability informed the optional Alaali Forensic Integrity Ledger. Then, these mechanisms were integrated under the Alaali Forensic Governance Doctrine as the doctrinal anchor of DFAS-ForensicGov.

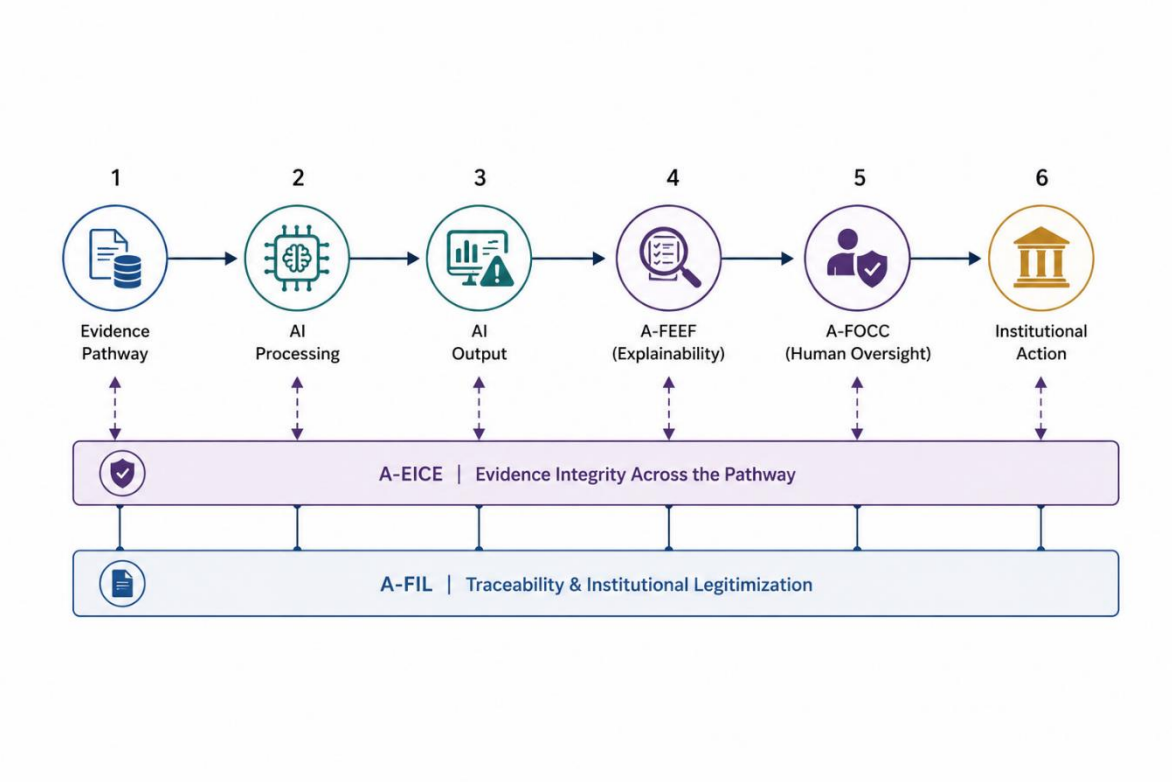

Figure 1. DFAS-ForensicGov Integrated Governance Architecture

A-EICE operates across the evidence pathway, while A-FIL preserves traceability and institutional accountability throughout the governance sequence.

Fourth, the proposed architecture was assessed through two analytical tests. The first was comparative positioning against existing auditing, financial-crime, and AI-governance frameworks. The second test used illustrative scenarios involving opaque fraud signals, potential false positives, and cross-border data-processing risks. These scenarios do not constitute empirical case-study evidence. They are analytical demonstrations used to examine whether the proposed mechanisms produce a coherent sequence from AI output to human review, intervention, evidence calibration, and documented institutional action.

Criteria for Conceptual Evaluation

The proposed doctrine was evaluated against four criteria. First, conceptual coherence: Each mechanism must address a distinct governance problem without duplicating the function of another mechanism. Second, operational specificity: The doctrine must identify actionable triggers, responsible authority, intervention logic, and documentation requirements rather than relying on abstract ethical principles. Third, institutional compatibility: The doctrine must complement, rather than replace, existing auditing standards, AI-governance frameworks, and applicable legal requirements. Fourth, contextual adaptability: the architecture must be able to adapt to different organizational capacities and jurisdictional requirements without losing its core governance logic.

These criteria reflect the central requirements of conceptual contribution: clear constructs, specified relationships among constructs, explanatory logic, and identified boundary conditions (Whetten, 1989). They also respond to the requirement that conceptual models provide a transparent account of how their theoretical elements were selected and combined (Jaakkola, 2020).

Scope and Boundary Conditions

DFAS-ForensicGov is designed for organizations that use artificial intelligence to support forensic accounting or related financial-investigation activities. It is not a substitute for applicable law, professional auditing standards, forensic procedures, data-protection obligations, or judicial evidentiary rules. Nor does it determine whether a particular AI system is legally compliant or technically accurate.

Its role is narrower and more practical: it provides a governance layer for controlling how forensic outputs generated by AI are reviewed, escalated, challenged, recalibrated, authorized, and documented before they influence consequential institutional action. Therefore, the doctrine is most applicable where AI supports, but does not replace, accountable human judgment.

FINDINGS AND DISCUSSION

Core Finding: The Governance Gap Is Operational

The conceptual analysis identified a specific operational gap in AI-enabled forensic accounting. Existing frameworks recognize that AI systems should be transparent, accountable, and supported by appropriate documentation. However, they do not specify how these requirements should operate inside a forensic investigation when an AI-generated output may lead to a consequential institutional decision.

This gap is not primarily technical. It concerns the transition from AI-generated analysis to institutional action. An AI system may identify a suspicious transaction, classify an individual as high risk, or recommend further investigation. However, before that output affects a disciplinary decision, regulatory referral, asset-freezing process, or litigation strategy, the organization must answer four governance questions:

1. Is the output sufficiently explainable for the intended decision?
2. Who has the authority to challenge, suspend, modify, or reject the output?
3. Has the underlying evidence been processed under legally and institutionally acceptable conditions?
4. Is there a traceable record of governance decisions made throughout the process?

Existing AI-governance frameworks provide relevant principles to these questions. For example, the NIST AI RMF identifies accountability, explainability, transparency, privacy, and validity as important characteristics of trustworthy AI (NIST, 2023). The EU AI Act also requires appropriate transparency, logging, and human oversight for high-risk AI systems (European Parliament & Council of the European Union, 2024). However, neither framework provides a forensic-specific mechanism that connects these principles to investigative authority, evidence integrity, and decision escalation.

DFAS-ForensicGov addresses this gap by treating governance as an embedded operational layer rather than a retrospective compliance exercise. Its function is to control whether and how forensic outputs generated by AI may influence consequential decisions. This approach extends earlier work on accountability, sovereignty, and operational control in AI-enabled financial systems to forensic accounting (Alaali, 2026).

DFAS-ForensicGov Architecture

The analysis produced an integrated architecture comprising one doctrinal anchor and four connected governance mechanisms. It governs the transition from AI-assisted detection to consequential institutional action: A-FGD establishes the governing boundary, A-FEEF assesses

explainability and triggers escalation, A-FOCC assigns authority for human intervention, A-EICE assesses the legitimacy of the evidence pathway, and A-FIL preserves the traceable record of consequential governance decisions.

The first component is the Alaali Forensic Governance Doctrine (A-FGD). It establishes the central boundary of the architecture: AI may support forensic analysis, but it cannot displace identifiable human responsibility for consequential investigative decisions. A-FGD rests on four principles:

1. **Explainability as an operational requirement**: AI outputs must be sufficiently understandable for their intended investigative use.
2. **Enforceability by design**: Governance requirements must be embedded in roles, triggers, decisions, and documentation.
3. **Sovereign-sensitive governance**: Data processing and evidence use must remain compatible with applicable jurisdictional, legal, and institutional conditions.
4. **Adaptive governance**: Controls must be recalibrated when the AI system, evidence environment, consequence level, or legal context changes.

These principles establish the logic of the doctrine but do not alone govern an investigation. A-FGD builds on the DFAS AI Governance Doctrine Compendium's enforceability, accountability, and sovereign-awareness principles (Alaali, 2025a).

The second component is the Alaali Forensic Explainability Escalation Framework (A-FEEF). A-FEEF determines whether an AI-generated output is sufficiently explainable for its intended use and establishes the required level of review before consequential reliance.

Operational Explainability Classification Rule

A-FEEF classifies each AI-generated forensic output before it is used in an investigative or institutional decision. The assessment applies to the specific output, its intended use, and its consequence level; it is not a general assessment of the underlying AI system's ability to explain in all circumstances.

The classification uses a six-item explainability checklist. Each criterion is assessed as either satisfied (1) or not satisfied (0). A criterion is treated as not satisfied when the required information, documentation, explanation, or evidence is unavailable, incomplete, materially inconsistent, or cannot be independently verified by the authorized reviewer. No partial score is assigned.

The six assessment criteria are as follows:

1. **Purpose clarity**: The output's proposed investigative or institutional use, intended decision, and consequence level are explicitly defined.
2. **Output traceability**: The relevant system or model version, input scope, processing pathway, generation time, and responsible function can be identified and documented.
3. **Reason communication**: A competent reviewer can explain the material factors, analytical logic, and evidentiary basis that produced the output in intelligible terms.
4. **Uncertainty disclosure**: Material assumptions, limitations, confidence constraints, data quality concerns, and known failure conditions are disclosed.
5. **Evidentiary reconciliation:** The output can be compared with underlying records, contextual evidence, and plausible alternative explanations.
6. **Challengeability**: An authorized reviewer can meaningfully question, verify, modify, restrict, suspend, or reject the output and record the rationale for that decision.

For the purposes of A-FEEF, full explainability does not mean that an AI system is infallible or that every element of its computational code is transparent. This means that the output achieves a score of 6/6 for the intended use, with no critical failure. A competent and independent reviewer

must be able to reconstruct why the output was produced, identify its material limitations, assess its evidentiary basis, consider reasonable alternative explanations, and exercise informed professional judgment before relying on the output.

Table 2. A-FEEF Explainability Classification and Governance Response

Explainability level	Scoring and classification rule	Governance response	Permitted Institutional Use
Level 1: Fully explainable	**6/6 criteria were satisfied** and no critical failure was observed.	Routine professional review and documentation. The completed checklist, responsible reviewer, evidence examined, and decision rationale must be recorded in A-FIL.	May support investigative analysis and may be considered in consequential decision-making, subject to ordinary professional judgment, evidentiary requirements, and A-EICE compliance.
Level 2: Partially explainable	**4/6 or 5/6 criteria satisfied**, no critical failure, but one or two material aspects require further interpretation, review, or corroboration.	Additional authorized human review, independent corroborating evidence, and documented justification are required before reliance on the AI-generated output.	May guide investigative activity or prioritize further inquiry, but cannot independently support disciplinary, regulatory, legal, asset-recovery, criminal-referral, or other consequential institutional action.
Level 3: Limited or non-explainable	**0/6 to 3/6 criteria are satisfied**, or any critical failure is identified regardless of the numerical score.	Formal escalation, A-FOCC intervention, documented restriction or suspension, and independent corroboration before any further use are recommended.	May be retained only as an exploratory lead. It cannot independently support a consequential institutional action.

A critical failure exists where the data or processing pathway cannot be reconstructed; the material reasons for the output cannot be communicated; material uncertainty is undisclosed; the output conflicts with available evidence and that conflict remains unresolved; or an authorized reviewer cannot meaningfully challenge the output. A critical failure automatically prevents Level 1 or 2 classification.

A-FEEF functions as a gateway between AI-generated analysis and human decision authority. It does not determine legal admissibility or allocate final institutional responsibility. Rather, it determines the explainability status of an output and the level of governance control required before that output may influence an investigative or institutional decision. This is consistent with the requirement that human oversight should enable informed intervention rather than nominal supervision (European Parliament & Council of the European Union, 2024).

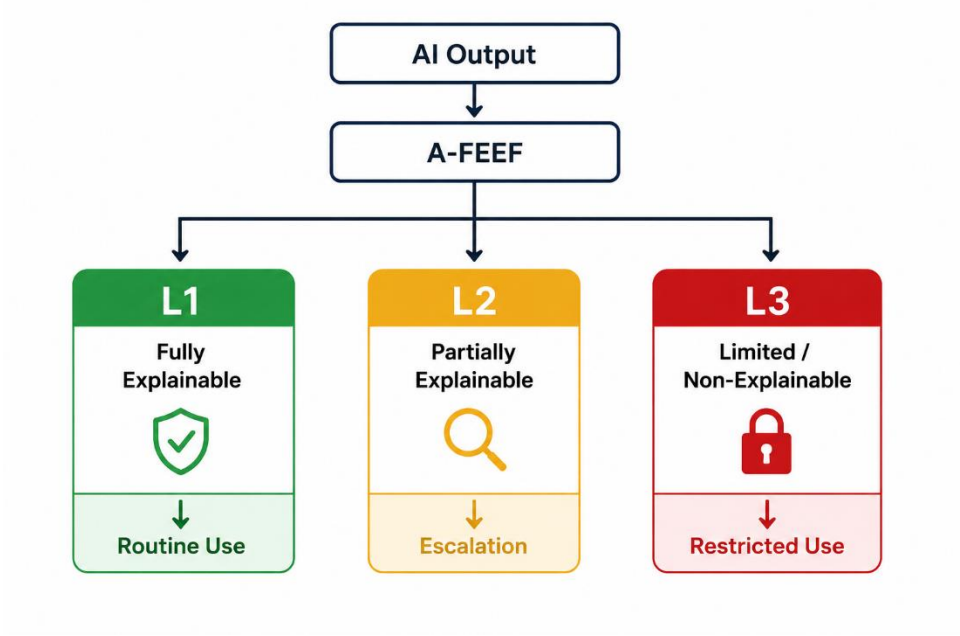

Figure 2. A-FEEF Explainability–Escalation Architecture

A-FEEF links the explainability status of an AI-generated forensic output to the corresponding level of governance control before institutional reliance.

The third component is the Alaali Forensic Override Command Chain (A-FOCC). A-FOCC allocates authority for a designated investigator, supervisor, committee, or other authorized function to review, modify, suspend, reject, or restrict reliance on an AI-generated output. It is distinct from A-FEEF: A-FEEF assesses and escalates the output, whereas A-FOCC determines who acts on that escalation.

Every A-FOCC intervention must record the governance rationale, responsible authority, considered evidence, timing, decision, and any condition for further review or resumed use. This conversion of human oversight into assigned and reviewable institutional authority is consistent with the lifecycle accountability requirements identified in AI-auditing research (Raji et al., 2020).

The fourth component is the Alaali Evidence Integrity Calibration Engine (A-EICE). A-EICE governs whether the evidence pathway remains legally, jurisdictionally, and institutionally defensible. It applies before AI processing is authorized and throughout the investigative lifecycle, covering data access, processing, storage, transfer, retention, and use. It does not assess the accuracy, explainability, or intervention authority of the model.

An AI output may be analytically useful but institutionally unusable if the relevant data were processed through an impermissible environment, transferred in breach of applicable requirements, accessed without sufficient authority, or retained in a way that compromises confidentiality, privacy, chain of custody, or evidentiary integrity. Therefore, A-EICE extends governance beyond model-level control to the legitimacy of the evidence pathway, consistent with sovereign-sensitive governance in AI-enabled financial systems (Alaali, 2026).

The fifth component is the Alaali Forensic Integrity Ledger (A-FIL). A-FIL preserves a reconstructable governance record that links the AI output, explainability assessment, escalation status, intervention authority, evidence-pathway assessment, and final institutional decision.

A-FIL may be simplified proportionately for low-consequence exploratory analysis that does not influence a formal investigation or institutional decision. However, traceability is mandatory when an output is escalated, overridden, relied upon in a formal investigation, or used to support disciplinary, regulatory, litigation, asset-recovery, criminal-referral, or other consequential action.

The record must enable a later reviewer to establish the output, why it was considered or rejected, who exercised authority, whether the evidence pathway was accepted, and how the final decision was authorized.

The mechanisms are complementary but non-substitutable. A-FEEF cannot establish evidence-pathway legitimacy; A-FOCC cannot explain an opaque output; A-EICE cannot determine whether an authorized decision maker should rely on a particular output; and A-FIL cannot correct a defective assessment or intervention decision. Together, they provide a controlled and reviewable pathway from AI-generated forensic analysis to consequential action.

Integrated Operational Sequence

The four mechanisms operate as an integrated governance cycle rather than as isolated or strictly one-directional controls. Evidence-integrity and jurisdictional assessment cannot occur only after an output is generated by an AI system. Where data access, processing, storage, transfer, or use may be legally restricted or institutionally improper, A-EICE assessment must begin before AI processing is authorized and continue throughout the investigative lifecycle.

Therefore, the sequence combines an evidence-pathway control layer with an AI-output decision-control layer.

Before processing forensic data, A-EICE assesses whether the proposed evidence pathway is permissible and institutionally defensible. This includes authority to access data, privacy and confidentiality requirements, cloud or vendor arrangements, storage location, cross-border transfer conditions, retention requirements, and potential effects on evidentiary integrity. If these conditions are not satisfied, the proposed AI environment must not process the data until the pathway is corrected, restricted, or formally authorized.

Once an AI system generates an investigative output, the following sequence applies:

1. **A-FEEF** assesses whether the output is sufficiently explainable, contextually interpretable, and reliable for its intended use and consequence level.
2. Where explainability is limited, the output conflicts with available evidence, or it may influence consequential institutional action, **A-FEEF** triggers formal escalation.
3. **A-FOCC** assigns authorized human responsibility to review, challenge, modify, suspend, or reject the output, considering underlying evidence, alternative explanations, model limitations, and the proposed response's proportionality.
4. **A-EICE** applies again as a continuing control to confirm that the evidence pathway remains legally, jurisdictionally, and institutionally defensible before the output is relied upon in disciplinary, regulatory, litigation, asset-recovery, or other consequential action.
5. **A-FIL** records the classification of explainability, escalation trigger, responsible authority, review rationale, intervention decision, evidence-pathway assessment, and final authorization.

A-FIL may be simplified where an AI output is used solely for low-consequence exploratory analysis and does not influence a formal investigation or institutional decision. However, traceability is mandatory once an output is escalated, overridden, relied upon in a formal investigation, or capable of affecting an individual, organization, regulator or legal process.

This sequence separates four questions that are often wrongly merged: whether an AI system detected a pattern, whether that pattern is sufficiently explainable and reliable for its intended use, whether the evidence pathway is legally and institutionally defensible, and whether an authorized person has accepted responsibility for the resulting action. Therefore, an analytically useful output is not necessarily decision-ready or legally usable.

The mechanisms must be embedded in organizational workflows through defined responsibilities, escalation thresholds, pre-processing evidence controls, continuing compliance

checks and mandatory documentation for consequential use. This operational orientation is consistent with the DFAS Governance Deployment Roadmap, which emphasizes structured institutional adoption rather than passive compliance with governance principles (Alaali, 2025b).

Comparison with the Existing Frameworks

DFAS-ForensicGov is a complementary governance layer, not a replacement for professional auditing standards, financial-crime requirements, artificial intelligence (AI) regulation, or general AI risk-management frameworks. ISA 240 and ISA 315 establish fraud, professional-judgment, risk-assessment, and audit-evidence responsibilities; FATF provides financial-crime and cross-border cooperation foundations; and the NIST AI RMF and EU AI Act establish broader AI risk-management, transparency, documentation, and human-oversight expectations (IAASB, 2021a, 2021b; FATF, 2023; NIST, 2023; European Parliament & Council of the European Union, 2024).

The frameworks operate at different levels of governance. DFAS-ForensicGov contributes to the transition from AI-assisted detection to institutional action. Its contribution is not the invention of established principles, but their integration into a sequenced forensic-decision architecture that separates detection validity, decision reliability, evidence-pathway legitimacy, and institutional authorization.

Table 3. Comparative Governance Positioning of DFAS-ForensicGov

Governance function	ISA 240 and ISA 315	FATF Recommendations	NIST AI RMF	EU AI Act	DFAS-ForensicGov
Primary governance level	Professional audit and fraud responsibilities	Financial-crime governance and international cooperation	Enterprise AI risk management	Regulatory governance of AI systems	AI-assisted forensic decision governance
The AI output explainability threshold	Not specified for the AI outputs	Not specified	Recognizes explainability and transparency as trustworthy AI characteristics	Requires transparency and usable information for relevant systems	Classifies explainability and links it to a defined escalation response
Escalation trigger before consequential action	Professional judgment applies, but no AI-specific trigger	Risk-based approach, but no AI-output escalation logic	Risk assessment and management	Human oversight and risk controls, subject to the applicable system classification	A-FEEF establishes escalation when explainability or consequence conditions require additional review.
The authority to modify, suspend, or reject an AI	Not operationalized as an AI override structure	Not operationalized as an AI override structure	Encourages accountable governance without a forensic	Requires human-oversight capacity but does not	A-FOCC assigns responsibilities for authorized

output			command chain	prescribe forensic decision authority	intervention, suspension, modification, rejection and documentation
Governance of evidence pathway and jurisdictional conditions	Addresses audit evidence, not AI data processing pathways	Supports cross-border cooperation, but not AI evidence processing	Broadly addresses data and lifecycle risk	Establish relevant data governance, documentation, and compliance obligations	A-EICE evaluates whether data access, processing, storage, transfer, and use remain defensible
Traceability of forensic governance decisions	Documentation requirements generally apply	Record-keeping and cooperation requirements generally apply	Lifecycle documentation and governance practices	Logging and documentation requirements for the relevant systems	A-FIL records escalation, review, override, evidence-calibration decision, and final institutional authorization
Core Contribution to Forensic AI Governance	Foundations of Fraud and Audit Evidence	Financial crime and foundations of cooperation	Trustworthy AI risk management foundations	Regulatory Obligations for AI Governance	Integrating relevant requirements into a sequenced pathway from AI output to accountable institutional action

DFAS-ForensicGov does not claim jurisdiction over the legal, technical, or professional domains governed by other frameworks. Its added value lies in specifying how their relevant requirements are translated into a coherent operational sequence when an AI-generated forensic output may influence a consequential institutional outcome.

Illustrative Application

These scenarios are analytical illustrations rather than empirical case evidence. They demonstrate how explainability, intervention authority, evidence-pathway legitimacy, and traceability operate together, including circumstances that require AI use restriction or suspension.

Consider using an AI system to identify potential procurement fraud. Before processing financial records, communications, and vendor data, A-EICE assesses whether the organization has authority to access the data and whether the proposed processing environment satisfies confidentiality, retention, and evidence-integrity requirements. Subsequently, the system flags several executives as high risk and recommends immediate suspension, but provides only a general

risk score without identifying the relative contribution of contracts, communications, payment patterns, or relationship indicators.

Under DFAS-ForensicGov, the output is insufficient evidence for disciplinary action. A-FEEF classifies the output as Level 3 because its reasoning and evidentiary basis are insufficiently explainable for a high-consequence decision. This triggers A-FOCC, requiring an authorized investigator, supervisory committee, or designated decision maker to suspend reliance on the recommendation pending review of the underlying transactions, contextual evidence, alternative explanations, and potential model or data-quality weaknesses.

A-EICE operates concurrently during the review. The evidence pathway may be institutionally defective if relevant records were processed outside an approved environment, accessed without sufficient authority, or retained inconsistently with applicable requirements. A-FOCC may then require the output to be withdrawn from consequential use, require independent corroboration from lawfully processed evidence, or suspend further use of the system until the pathway is corrected. A-FIL records the classification of explainability, escalation, evidence-pathway assessment, intervention rationale, and final decision. The output may remain useful as an exploratory lead but cannot independently support suspension, referral, or another consequential action.

The second scenario concerns cross-border data processing. An investigation team proposes the use of a cloud-based AI platform hosted outside the jurisdiction in which the relevant financial records were obtained. Before processing, the A-EICE assesses the permissibility of cross-border transfer, storage location, vendor access, confidentiality controls, and processing authority. If these conditions are unresolved or noncompliant, data must not be processed through that environment merely because it offers analytical advantages.

Assume that a permitted subset of data is lawfully processed and that the system identifies potentially relevant transaction patterns. A-FEEF may classify the output as Level 2, where it is partially explainable but requires corroboration. A-FOCC then requires an authorized reviewer to determine whether the output may guide additional investigation within defined limits or whether it must be suspended. A-EICE continues to assess compliance as new data, jurisdictions, vendors, or investigative purposes become involved. If the processing pathway later becomes non-compliant, A-FOCC may suspend processing, require migration to an approved environment, or prohibit reliance on the output in a regulatory, disciplinary, or legal process.

The scenarios demonstrate that the mechanisms are connected but non-substitutable. A-FEEF and A-FOCC cannot remedy evidence obtained through an unlawful or institutionally indefensible pathway, while A-EICE cannot sufficiently explain an opaque output for consequential use. Where assessments conflict, a more restrictive condition is applied. An output cannot proceed to consequential institutional action unless it is sufficiently explainable for its intended use, subject to authorized human review, and derived through an evidence pathway that remains legally and institutionally defensible.

Discussion

DFAS-ForensicGov shifts governance from high-level principles to a mechanism-based approach for AI-assisted forensic decisions. Its central safeguard is that an AI-generated output may be analytically useful as an investigative lead while remaining insufficient for disciplinary, regulatory, legal, or other consequential action. Institutional decisions must rest on accountable human judgment, corroborating evidence, and a defensible evidence pathway rather than on an AI-generated score or recommendation alone.

The doctrine also introduces sovereign-sensitive governance as a design requirement. AI governance in forensic accounting cannot be separated from the conditions under which evidence

is accessed, processed, transferred, retained, and relied upon. A-EICE extends governance beyond algorithmic performance toward the complete evidence pathway's institutional defensibility, consistent with the broader need to integrate accountability, sovereignty, and operational control in AI-enabled financial systems (Alaali, 2026).

However, the doctrine is not a frictionless or universally applicable solution. Its mechanisms are distinct but interdependent. An explainability assessment may trigger human intervention, and a review may also identify an evidence-pathway problem. Therefore, organizations need clear protocols to prevent duplicated review, conflicting authority, or uncertainty about which control takes priority.

Implementation creates institutional costs. Organizations may need to assign decision authority, train investigators, modify documentation procedures, review vendor and cloud arrangements, and establish auditable intervention records. These requirements may be proportionate to high-consequence investigations but burdensome for smaller firms or units with limited legal, technical, or forensic capabilities. A rigid application of every control to every AI-assisted activity could delay legitimate investigations and turn governance into a procedural burden.

Therefore, implementation should be proportionate to the consequence level of AI-supported activity. Low-consequence exploratory analysis may require lighter documentation and supervisory review, whereas formal explainability assessment, authorized intervention, continuing evidence-pathway evaluation, and mandatory traceability are required for outputs capable of affecting an individual, organization, regulator or legal process. Thresholds, role assignments, and documentation requirements must be calibrated to the legal authority, risk profile, investigative capacity, and jurisdictional environment of the organization.

The sequence may fail or require suspension where no authorized reviewer has sufficient technical or forensic competence; data access or cross-border processing is unlawful from the outset; investigators lack independence from decision makers; urgent action is required before full review; or the AI system, vendor, or data environment cannot provide adequate documentation. In such circumstances, a formal governance process does not legitimize continued reliance on AI. The appropriate response may be to restrict the system to exploratory use, require independent review, move processing to an approved environment, or suspend its use for the relevant purpose.

Implementation should proceed through controlled pilots that test escalation routes, role clarity, review delays, and the reconstructability of the evidence pathway under scrutiny. The doctrine does not claim to eliminate error, bias, legal uncertainty, or institutional failure. Its narrower purpose is to ensure that uncertainty is recognized, assessed by appropriate authorities, documented and prevented from silently becoming a consequential institutional action.

CONCLUSIONS

This study developed DFAS-ForensicGov as a conceptual governance doctrine for AI-enabled forensic accounting. It addresses an operational-governance gap: existing AI-governance, auditing, regulatory, and financial-crime frameworks recognize transparency, human oversight, accountability, documentation, risk management, and data governance, but do not provide a unified forensic-accounting architecture for governing the movement from an AI-generated output to a consequential institutional decision.

DFAS-ForensicGov responds through an integrated architecture comprising A-FGD, A-FEEF, A-FOCC, A-EICE, and A-FIL. Together, these mechanisms establish a controlled pathway in which AI outputs are assessed for explainability, escalated where necessary, reviewed by authorized humans, evaluated against evidence-pathway conditions, and documented for consequential institutional use.

The contribution of the doctrine is operational rather than merely normative. It does not claim to invent principles of explainability, accountability, human oversight, or evidence integrity. Its contribution is to connect them within a sequenced forensic-decision architecture that separates detection validity, decision reliability, evidence-pathway legitimacy, and institutional authorization.

The framework distinguishes analytical usefulness from decision legitimacy. An AI-generated pattern may be useful as an exploratory investigative lead while remaining insufficient to support disciplinary action, regulatory referral, litigation strategy, asset recovery, or criminal complaint. Therefore, DFAS-ForensicGov should be understood as a complementary governance layer: it does not replace professional auditing standards, legal obligations, forensic procedures, data protection requirements, judicial evidentiary rules, or sector-specific AI regulation.

The framework remains conceptual and requires empirical validation. Its practical effectiveness depends on institutional capability, investigator competence, legal alignment, technical-system quality, and the integrity of the evidence environment.

LIMITATION & FURTHER RESEARCH

This study has several limitations. First, DFAS-ForensicGov is a conceptual doctrine. Its mechanisms are derived through structured conceptual analysis and comparative reasoning rather than through field implementation, experimental testing, or longitudinal evidence. Therefore, the study demonstrates conceptual coherence and operational logic, but it does not establish that the doctrine improves fraud detection, reduces false positives, strengthens evidence admissibility, or produces superior investigative outcomes in practice.

Second, the doctrine does not provide a universal legal test for evidence admissibility or data processing compliance. The legal requirements concerning privacy, cross-border data transfer, investigative authority, disclosure, and evidentiary use differ across jurisdictions. A-EICE is designed to require jurisdiction-sensitive assessment; it does not replace legal counsel, regulatory guidance, or national law.

Third, the framework has not yet been tested across different organizational settings. Audit firms, regulatory authorities, internal audit functions, financial institutions, and public investigative bodies differ materially in governance maturity, technological capacity, staff expertise, risk tolerance, and legal authority. These differences may affect implementation feasibility, cost, and institutional design.

Fourth, the doctrine intentionally focuses on the governance of AI-assisted forensic outputs rather than on the development of technical models. It does not prescribe a particular machine-learning method, explainability technique, data architecture, or cybersecurity standard. Its effectiveness will partly depend on the technical quality, documentation, and reliability of the AI systems it is used with.

Therefore, further research should proceed in five directions:

1. **Pilot implementation**: The framework is tested in audit firms, internal investigation units, regulatory agencies, and financial-crime compliance functions.
2. **Scenario-based validation**: controlled simulations involving false positives, opaque outputs, conflicting evidence, and cross-border data restrictions are used to test escalation and override decisions.
3. **Measurement of governance-performance**: Develop and validate indicators such as the Forensic Explainability Ratio, Override Traceability Index, Evidence Integrity Compliance Rate, and Sovereign Sensitivity Score.
4. **Cross-jurisdictional comparison**: Examine how the doctrine must be adapted across legal systems, regulatory environments, and data-governance regimes.

5. **Comparative evaluation**: AI-assisted investigations using DFAS-ForensicGov are compared with investigations operating under ordinary internal control, audit, or AI-governance procedures.

Therefore, the next stage of development should move from conceptual architecture to empirical validation. The long-term value of DFAS-ForensicGov will depend on whether future research demonstrates that AI-enabled forensic investigations can improve the quality, fairness, traceability, and institutional defensibility of forensic investigations.

REFERENCES

Alaali, H. M. H. (2025a). *DFAS doctrine compendium: Philosophical, strategic, and systemic frameworks for AI governance, crisis intervention, and algorithmic control*. SSRN. https://doi.org/10.2139/ssrn.5367894

Alaali, H. M. H. (2025b). DFAS-GDR: Governance deployment roadmap—Implementation frameworks and adoption protocols for the DFAS Governance Convergence Doctrine. *Journal of Islamic Banking, Economics and Policy, 2*(1), 43–65. https://doi.org/10.63740/bjffnd45

Alaali, H. M. H. (2026). AI governance in financial systems: A PGCC and emerging markets convergence framework for accountability, sovereignty, and operational control. *Journal of Countries Studies*, 1–58. https://doi.org/10.22059/jcountst.2026.412773.1476

Alzoubi, A. B. (2025). Maximizing internal control effectiveness: The synergy between forensic accounting and corporate governance. *Journal of Financial Reporting and Accounting, 23*(1), 404–416. https://doi.org/10.1108/JFRA-03-2023-0140

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), Article 188, 1–21. https://doi.org/10.1145/3449287

Cicchini, D., Lombardi, R., & Bianchi, M. T. (2026). Exploring forensic accounting: Insights and critical analysis. *Corporate Governance*. Advance online publication, 1–22. https://doi.org/10.1108/CG-07-2025-0474

European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Financial Action Task Force. (2023). *The FATF recommendations: International standards on combating money laundering and the financing of terrorism & proliferation*. https://www.fatf-gafi.org/en/publications/Fatfrecommendations/Fatf-recommendations.html

Huy, P. Q., & Phuc, V. K. (2025). Managing performance of cyber forensic accounting with the effect of artificial intelligence in legal and ethical considerations. *Applied Artificial Intelligence, 39*(1), Article 2513740. https://doi.org/10.1080/08839514.2025.2513740

International Auditing and Assurance Standards Board. (2021a). *International Standard on Auditing 240: The auditor's responsibilities relating to fraud in an audit of financial statements*. International Federation of Accountants.

International Auditing and Assurance Standards Board. (2021b). *International Standard on Auditing 315 (Revised 2019): Identifying and assessing the risks of material misstatement*. International Federation of Accountants.

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review, 10*(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., & Vaughan, J. W. (2020). Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Article 3376219). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376219

Mökander, J. (2023). Auditing of AI: Legal, ethical and technical approaches. *Digital Society, 2*, Article 49. https://doi.org/10.1007/s44206-023-00074-y

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science*

and Engineering Ethics, 26, 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

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

Organisation for Economic Co-operation and Development. (2024). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449). https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372873

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287598

Stratopoulos, T. C., & Wang, V. X. (2025). Artificial intelligence and accounting research: A framework and agenda. *International Journal of Accounting Information Systems, 56*, Article 100760. https://doi.org/10.1016/j.accinf.2025.100760

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review, 14*(4), 490–495. https://doi.org/10.5465/amr.1989.4308371

APPENDIX

DFAS-FEP Authorship Classification and Compliance

This manuscript complies with the DFAS-FEP Protocol, a structured authorship governance framework introduced by the author.

DFAS-FEP (Dynamic Financial Applied Meta-Science – Future Engine Prototype) provides an ethical protocol to ensure transparent and accountable authorship in AI-assisted financial research. It defines and enforces the distinction between fully human-generated content and AI-supported elements to uphold integrity, intellectual responsibility, and traceability across all manuscript components.

This classification directly supports the ethical architecture of the DFAS-IFRS Code of Ethics, which applies DFAS-FEP standards to AI-influenced financial disclosures.

For reference, see the published DFAS-FEP Protocol:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5260517

Table 4. (DFAS-FEP) Authorship Classification Table for This Manuscript (DFAS-ForensicGov)

Manuscript Component	DFAS-FEP Class	Justification
DFAS-ForensicGov Core Governance Doctrine	Class 1	The author fully conceptualized the doctrine governing AI-generated forensic outputs before consequential institutional action.
Alaali Forensic Governance Doctrine (A-FGD)	Class 1	The author independently developed A-FGD and its four principles—Explainability as an Operational Requirement, Enforceability by Design, Sovereign-Sensitive Governance, and Adaptive Governance.
Alaali Forensic Explainability Escalation Framework (A-FEEF)	Class 1	A-FEEF and its three-level architecture—Fully Explainable, Partially Explainable, and Limited or Non-Explainable—together with the corresponding escalation requirements were fully human-authored.
Alaali Forensic	Class 1	The author developed the formal architecture for authorized

Override Command Chain (A-FOCC)		human review, modification, suspension, restriction, or rejection of AI-generated forensic outputs, including intervention documentation requirements.
Alaali Evidence Integrity Calibration Engine (A-EICE)	Class 1	The author independently conceptualized A-EICE and its continuing jurisdiction-sensitive governance of data access, processing, storage, transfer, retention, use, privacy, regulatory compliance, and evidentiary integrity.
Alaali Forensic Integrity Ledger (A-FIL)	Class 1	A-FIL was independently developed as the traceability mechanism for recording explainability assessments, escalations, interventions, evidence-pathway decisions, and final authorisation. Its application may be proportionately simplified for low-consequence exploratory analysis, but for escalated, overridden, formal, or consequential institutional use, traceability is mandatory.
Integrated Operational Governance Sequence	Class 1	The author independently designed the integrated sequence in which A-EICE assesses the evidence pathway before AI processing and continues throughout use; AI Output is assessed through A-FEEF; A-FOCC assigns intervention authority; A-EICE confirms continuing legitimacy; A-FIL preserves consequential governance decisions; and Institutional Action follows only where the required conditions are satisfied. The author-developed separation between detection validity, decision reliability, evidence-pathway legitimacy, and institutional authorization.
Conceptual Development and Evaluation Architecture	Class 1	The author structured the four-stage analytical procedure and evaluation criteria of conceptual coherence, operational specificity, institutional compatibility, and contextual adaptability for the development and assessment of the doctrine.
Forensic Application and Boundary Logic	Class 1	The author developed the application of the doctrine to opaque fraud signals, false positives, consequential decisions, and cross-border data-processing risks, together with its boundary as a complementary governance layer rather than a substitute for law or professional standards.
Future Governance Measurement Architecture	Class 1	The author identified the proposed forensic explainability ratio, override traceability index, evidence integrity compliance rate, and sovereign sensitivity score as future governance-performance measurement directions.
Narrative Structure and Analytical Exposition	Class 2	AI assistance was limited to CLF, clarity enhancement, and structural presentation.
Tables, Figures, and Comparative Presentation	Class 2	The underlying concepts and governance relationships were designed by the author; AI assistance was restricted to format and presentation.
Literature Review and Cross-Referencing	Class 2	Literature selection and substantive interpretation were author-directed; AI assistance was limited to citation and reference alignment.
Appendices and Compliance Documentation	Class 2	AI assistance was limited to formatting, structural consistency, and controlled editorial refinement.

Overall Manuscript Classification: Class 2 – AI-Assisted, Author-Validated
Core Doctrine, Governance Mechanisms, and Operational Architecture: Class 1 – Fully Human-Authored

Additional Clarifications

Human-Originated Content (Class 1):

All substantive contributions, including DFAS-ForensicGov, A-FGD, A-FEEF, A-FOCC, A-EICE, A-FIL, the integrated operational governance sequence, conceptual evaluation architecture, forensic application logic, and proposed governance-performance indicators, were independently conceptualized and developed by the author.

AI-Assisted Content (Class 2):

AI tools were only used for non-substantive linguistic refinement, formatting, presentation consistency, and citation alignment. No AI tools were used to originate the doctrine or its substantive governance mechanisms.

Integrity Safeguards:

All AI-assisted content was manually reviewed and validated using DFAS-FEP v1.6, and the authorship classification and traceability documentation were maintained.