

Digital Trust in Public Funds: An Explainable AI Framework for Continuous Compliance Monitoring of U.S. Federal Grant Subrecipients

Ayesha Arobee

ORCID: 0009-0003-0395-5005

Budget Analyst, Kansas Department for Aging and Disability Services (KDADS), Topeka, Kansas, United States

Correspondence: Ayesha Arobee · ayeshaarobeees@gmail.com

Publication history. First published in the offline edition of Cybersecurity Advances, Volume 1, Issue 1 (December 2024). This publisher-authorized digital archival edition was released on 12 October 2026. The scholarly content is reproduced from the supplied offline-edition manuscript; review-stage notices were removed and journal typography, pagination, metadata, and accessibility were normalized.

© 2024 the author(s). Published by AlphaTech Scholarly Press under the Creative Commons Attribution 4.0 International License (CC BY 4.0). No DOI is displayed unless and until registration is verified.

Background: Federal grants flow through thousands of state agencies and subrecipients, yet oversight remains largely periodic and document-driven. Two 2024 regulatory developments transformed this environment: the revised OMB Uniform Guidance (2 CFR Part 200, effective 1 October 2024) directs pass-through entities toward *risk-based, continuous* subrecipient monitoring and raised the Single Audit threshold to one million dollars, while OMB Memorandum M-24-10 (March 2024) established risk-management expectations for artificial intelligence in federal use — including requirements for explainability and human oversight. Together they create both the mandate and the guardrails for AI-assisted compliance monitoring. **Objective:** We design and specify an explainable AI (XAI) framework — the Grant Risk Assessment and Compliance Engine (GRACE) — for continuous monitoring of subrecipient compliance that couples cross-system anomaly detection with tamper-evident evidence preservation. **Methods:** Following a design-science methodology, we (i) analyzed the 2024 regulatory corpus (2 CFR 200, GAO fraud-risk guidance, NIST AI Risk Management Framework), (ii) derived a cross-system red-flag taxonomy spanning financial, payroll, procurement, and performance-reporting systems, (iii) specified a five-component architecture, and (iv) evaluated it through regulatory-requirement mapping and a fully synthetic case demonstration. **Results:** GRACE maps onto every major obligation of the 2024 Uniform Guidance monitoring cycle; the synthetic walkthrough shows how feature-level attributions (e.g., duplicate-invoice rate +1.82, payroll-to-performance mismatch +1.45) translate into a triaged composite score (89/100 → site review) with a plain-language explanation and an immutable ledger entry for every monitoring event. Tamper-evident logging via hash chains and periodic Merkle-root anchoring gives auditors and pass-through entities a shared, verifiable evidence base. **Conclusions:** Explainable, provenance-preserving anomaly detection can convert subrecipient oversight from episodic document review into continuous, defensible, and auditable practice — a foundational mechanism of digital trust in the stewardship of public funds. Empirical validation on operational data is the required next step and is specified as such.

Keywords: explainable artificial intelligence; anomaly detection; grant compliance monitoring; tamper-evident audit trails; Uniform Guidance; subrecipient oversight; public-sector accountability; digital trust

Contribution statement. This paper makes four contributions: (1) a consolidated account of the 2024 federal regulatory developments that create both a mandate for continuous, risk-based subrecipient monitoring and binding governance expectations for the AI systems that would enable it; (2) a cross-system red-flag taxonomy for grant compliance, derived from GAO fraud-risk guidance and the occupational-fraud literature and expressed as machine-usable features; (3) the GRACE framework, a five-component architecture integrating ensemble anomaly detection, layered explanations, and a tamper-evident evidence ledger, specified for implementability within pass-through entities; and (4) a regulatory-requirement mapping and synthetic case demonstration that make the design's assumptions, data needs, and failure modes explicit and auditable.

1. Introduction

Each year the U.S. federal government awards hundreds of billions of dollars in financial assistance — grants, cooperative agreements, and other awards that largely reach their ultimate beneficiaries only after passing through states, local governments, tribes, universities, and nonprofit organizations acting as subrecipients [1]. The integrity of that downstream flow is the operational

heart of public financial trust: when a subrecipient misspends, double-bills, or misreports, the loss is not merely monetary. It corrodes the contract of confidence between citizens, oversight institutions, and the agencies that steward public funds. The General Accountability Office has repeatedly placed the integrity of federal grants and pandemic-era assistance among the government's highest-risk areas [6][7], and the Association of Certified Fraud Examiners estimates that organizations lose about five percent of revenue to occupational fraud annually, with government among the most exposed sectors [8].

The structural problem is that oversight has not kept pace with the flow of funds. Monitoring under the Uniform Guidance has traditionally been executed as *episodic* artifact review — annual financial reports, periodic desk reviews, and the annual Single Audit — that observes compliance months after transactions occur. The consequences of point-in-time oversight are well documented in pandemic-program oversight findings, where funds moved at unprecedented speed through historically thin monitoring [7]. Continuous monitoring, by contrast, evaluates transactions and behaviors as they accumulate, surfacing anomalies while intervention is still possible.

Two 2024 regulatory developments make this paper timely. First, the Office of Management and Budget's 2024 revision of the Uniform Guidance (2 CFR Part 200), effective 1 October 2024, reorganized subrecipient oversight around *risk and consequence*: it directs pass-through entities to evaluate subrecipient risk in award decisions, tailors monitoring intensity to that risk, and — by raising the Single Audit threshold from \$750,000 to \$1,000,000 — deliberately shifts a larger share of assurance burden onto pass-through entities' own risk-based monitoring [1][2]. Second, OMB Memorandum M-24-10, "Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence" (March 2024), implementing Executive Order 14110, established binding risk-management practices for AI used in federal decision contexts, including impact assessments, testing, mitigation, and meaningful human oversight [3][4]. NIST's AI Risk Management Framework (AI RMF 1.0) supplies the technical vocabulary — govern, map, measure, manage — into which such systems must fit [5]. Read together, these instruments define the regulatory space this paper occupies: oversight *should* become continuous and risk-based, and if AI is used to achieve it, the AI must be explainable, bounded, and human-supervised. An opaque anomaly-detection "black box" that flags subrecipients without reasons would fail M-24-10's expectations and, more fundamentally, would fail the Due Process-like needs of grant administration, where a flagged finding can trigger corrective action, cost disallowance, or reputational harm.

This paper therefore designs, specifies, and evaluates **GRACE — the Grant Risk Assessment and Compliance Engine** — a framework for continuous compliance monitoring of subrecipients built on three pillars: (i) *cross-system anomaly detection* that fuses evidence from federal payment systems (e.g., PMS, ASAP.gov), state and local financial and payroll systems, procurement records, and performance reporting into a single risk picture; (ii) *layered explainability* that converts model outputs into feature-level attributions, counterfactual statements, and plain-language narratives suitable for analysts, subrecipients, auditors, and oversight bodies; and (iii) *tamper-evident evidence preservation* — an append-only, hash-chained ledger with periodically anchored Merkle roots — that makes every scoring event, explanation, and analyst action independently verifiable at audit time.

1.1 Research questions

RQ1: What do the 2024 regulatory developments require of subrecipient monitoring, and what constraints do they place on AI-assisted monitoring systems? **RQ2:** Which cross-system features and red flags are both (a) detectable from data routinely available to pass-through entities and (b) grounded in the fraud and noncompliance literature? **RQ3:** How should an explainable, continuously monitoring architecture be structured so that every automated judgment is auditable, contestable, and preserved as evidence? **RQ4:** What does evaluation of such a framework require, and what are its principal limitations before operational deployment?

1.2 Contributions

We answer these questions with four contributions: the regulatory baseline and its governance constraints (Section 2.1, Table 1); a cross-system red-flag taxonomy expressed as machine-usable features (Section 3.2, Table 2); the GRACE architecture with its tamper-evident evidence ledger (Section 4, Figures 1–2); and an evaluation comprising a complete regulatory mapping plus a synthetic case demonstration with feature-level explanations and triage thresholds (Section 5, Figures 3–4, Tables 3–4). Throughout, we are explicit that this is a design-science artifact: its evaluation here is analytical and synthetic, and empirical validation on operational data is specified as the required next step rather than claimed.

2. Background and Related Work

2.1 The 2024 regulatory baseline

Three instruments define the 2024 environment. **The revised Uniform Guidance** (2 CFR Part 200, 2024 revision) restructures subrecipient oversight around a documented risk evaluation (200.332), requires monitoring arrangements *proportionate to risk*, and streamlines cost-principle and documentation requirements [1]. Critically for this paper, the revision raised the Single Audit threshold to \$1,000,000 (from \$750,000), effective for fiscal years beginning on or after 1 October 2024 — reducing the fraction of subrecipients subject to the annual Single Audit and thereby increasing the assurance responsibility borne by pass-through entities themselves [2]. **OMB M-24-10** requires agencies using AI in decision-making to adopt risk-management practices consistent with the NIST AI RMF, including impact assessments, independent testing, ongoing monitoring, and human review for consequential decisions [3][4]. **GAO's fraud-risk framework** directs program managers to identify, assess, and respond to fraud risks systematically, and GAO's High-Risk Series continues to flag federal grant administration as an area demanding improved monitoring [6][7]. The 2024 regulatory posture is thus doubly prescriptive: it demands better monitoring and, where AI is the enabler, demands trustworthy AI. GRACE is designed to satisfy both simultaneously, which is why its explainability and evidence-integrity components are architectural requirements rather than add-ons.

2.2 Fraud theory and red-flag literature

Cressey's classic fraud triangle — pressure, opportunity, rationalization — remains the organizing theory for red-flag reasoning in financial oversight [9]. The occupational-fraud literature operationalizes it: ACFE's 2024 Report to the Nations finds asset misappropriation dominant, median losses rising with scheme duration, and tips/controls detection gaps concentrated where monitoring is weakest [8]. In public-sector settings, objective procurement red-flag indicators — single-bid concentration, contract-splitting, unusual winner patterns — have been developed into cross-national corruption-risk indices for EU procurement [10], demonstrating that behavioral red flags can be formalized as quantitative features without waiting for audit findings. Federal grant-specific red flags recur across OIG and audit literature: unallowable cost categories, payroll costs unsupported by effort reporting, budget-to-actual variances, duplicate or split invoicing, and late or inconsistent performance reporting [6][7]. Section 3.2 converts these into a taxonomy spanning the systems a pass-through entity already touches.

2.3 Anomaly detection in financial administration

Anomaly detection research offers mature primitives: unsupervised methods such as isolation forests isolate outliers with low sample complexity [11]; the survey literature organizes point, contextual, and collective anomalies relevant to grant data, where "unusual for this subrecipient, this category, this time of year" (contextual) matters as much as global outliers [12]. Applied work in public procurement and benefits administration has demonstrated both the promise and the known failure mode — high false-positive rates that overwhelm small oversight staffs — motivating threshold design, ensembles that combine rule-based red flags with learned models, and explanation-driven triage [10][12].

2.4 Explainable AI and AI governance

Feature-attribution methods (LIME [13]; SHAP [14]) and counterfactual explanations [15] provide the technical substrate for explanation; however, the "explainability" literature's central caution applies directly to compliance contexts: post-hoc explanations of opaque models can be plausible-but-unfaithful, and for consequential decisions, Rudin argues for inherently interpretable models or carefully validated explanations [16]. GRACE therefore treats explanations as *auditable artifacts* — generated, stored in the evidence ledger, and validated against known-ground-truth cases — rather than ephemeral UI decorations. On the governance side, NIST AI RMF 1.0 [5] and its Generative AI Profile [17] define the functions and outcomes (including explainability and human oversight) that OMB M-24-10 incorporates by reference for federal AI use [3]; ISO/IEC 42001 provides a certifiable management-system wrapper [18]. Kroll's analysis of "accountable algorithms" frames the legal-administrative requirement: decisions subject to review need reasons that can be examined, contested, and reconciled with the record [19].

2.5 Tamper-evident logging and audit evidence

The evidence-integrity requirement has a mature cryptographic lineage: Merkle's hash-tree construction [20], Certificate Transparency's publicly auditable append-only logs [21], and long-standing NIST guidance on security-relevant log management and audit controls [22][23]. Applied to grant compliance, the design goal is narrower than a public blockchain: an *append-only, hash-chained* ledger within the pass-through entity's control, with periodic signed Merkle-root anchoring, suffices to make any post-hoc alteration of monitoring evidence detectable by anyone holding the chain of roots — including external auditors performing Single Audit support or oversight reviews.

3. Methodology

The study follows a design-science methodology in four components.

Regulatory corpus analysis (RQ1). We analyzed the 2024 Uniform Guidance revision [1][2], OMB M-24-10 [3], EO 14110 [4], the NIST AI RMF [5][17], GAO fraud-risk guidance [6][7], and audit-evidence requirements (including record-retention obligations of 2 CFR 200.334-337) to extract (a) obligations that continuous monitoring must satisfy and (b) constraints that an AI component must satisfy. Extracted requirements were normalized into a requirement register used for the mapping evaluation (Table 1).

Red-flag taxonomy derivation (RQ2). Candidate red flags were harvested from three source families — GAO fraud-risk framework practices [6], occupational-fraud scheme typologies [8], and grant-specific OIG/audit findings [7] — then filtered by two inclusion criteria: (i) computable from data routinely accessible to a pass-through entity (federal payment systems, state financial and payroll systems, procurement records, performance reports, audit findings), and (ii) attributable to a documented fraud or noncompliance pattern. The result is a cross-system taxonomy of 14 features across four system domains — financial systems, payroll and procurement systems, performance and reporting systems, and portfolio context (audit history, profile, and score trajectory) — as compiled in Table 2; each feature is annotated with its source pattern and its anomaly type (point, contextual, or collective) following [12].

Framework design and specification (RQ3). The architecture was constructed iteratively under three constraints derived from RQ1: explainability as an architectural requirement (M-24-10 [3]; AI RMF [5]); evidence integrity sufficient for audit reliance (hash chaining [20][21]; 2 CFR 200.334-337 [1]); and human-in-the-loop triage proportional to risk (2 CFR 200.332 [1]). Design decisions were evaluated against alternatives (e.g., inherently interpretable models per [16]; distributed ledgers versus hash-chained append-only logs) and are reported with rationale in Section 4.

Evaluation design (RQ4). Two evaluation instruments are used. First, a *regulatory-requirement mapping*: every GRACE component is traced to the specific requirements it addresses (Table 1), providing construct validity against the normative baseline. Second, a *synthetic case demonstration*: a fully synthetic subrecipient case (S-114) was constructed from the taxonomy — duplicate/split invoicing, payroll-to-performance mismatch, unallowable cost-category usage — and processed through the specified pipeline to illustrate the intended analyst experience (Figure 3, Table 3). The synthetic data is labeled as such throughout; no operational grantee data was used, and no claims of field performance are made. Limitations of this design are stated in Section 7.

4. The GRACE Framework (RQ3)

4.1 Architecture overview

GRACE comprises five components organized in three planes (Figure 1). The **data plane** integrates, at defined frequencies, transactional and summary evidence from federal payment systems (e.g., Payment Management System, ASAP.gov), state and local financial/payroll/procurement systems, Federal Financial Reports and performance narratives, prior Single Audit and OIG findings, and public records. The **intelligence plane** computes red-flag features (Table 2), scores each subrecipient with an ensemble — an unsupervised isolation-forest component over normalized features [11], a supervised red-flag-rule component encoding the taxonomy [6][8], and drift monitors for population shift — and produces the **explanation layer**: per-subrecipient SHAP-style feature attributions [14], counterfactual statements ("raising duplicate-invoice rate below 2% would reduce the score to the desk-review band") [15], and templated plain-language narratives. The **assurance plane** writes every event — scores, explanations, analyst actions, and case outcomes — to the **tamper-evident evidence ledger** (Figure 2), routes triage through human case management with cadence proportional to risk (Figure 4), and implements governance (bias/drift monitoring, model-change control, oversight reporting) aligned to the AI RMF's Govern-Map-Measure-Manage functions [5].

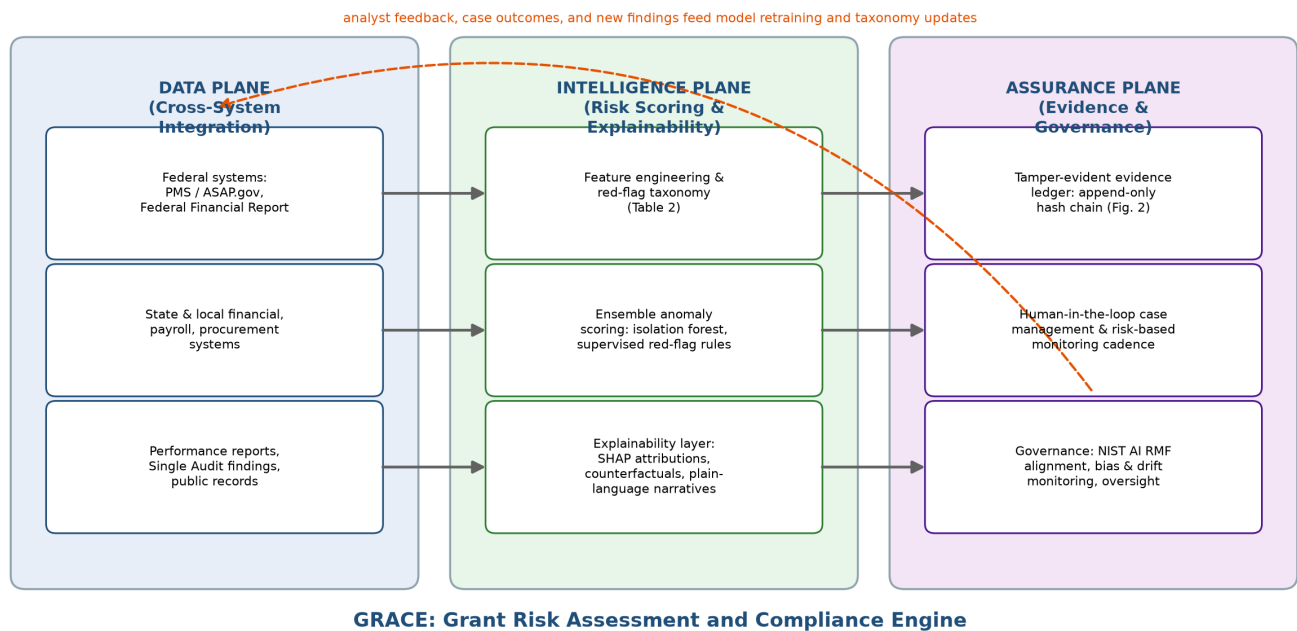


Figure 1. GRACE architecture: data, intelligence, and assurance planes with the feedback loop from case outcomes to model retraining and taxonomy updates.

4.2 Component design decisions and rationale

Ensemble scoring rather than a single model. Rule-based red flags provide precision on known patterns but miss novel ones; unsupervised anomaly detection catches novelty but produces high false-positive rates in low-prevalence regimes [11][12]. The ensemble scores each event with both and combines them into a composite in [0,100]; thresholds (default 30/70) are governance parameters, not model constants, and are calibrated per pass-through entity against analyst workload capacity.

Layered explanations rather than a single narrative. Three explanation strata serve four audiences (Table 4): feature attributions for analysts debugging the score [14]; counterfactuals for subrecipients seeking remediation paths [15]; templated narratives for auditors and oversight bodies, each narrative pinned to its evidence records. Consistent with [16], GRACE mandates *explanation validation*: attributions must be checked against seeded synthetic cases with known ground truth (as in Section 5.2) and, once deployed, against confirmed case outcomes — an explanation that systematically fails to track the evidence is treated as a defect, not a nuisance.

Tamper-evident ledger rather than a distributed ledger. Full blockchain deployment is unnecessary for a pass-through entity's assurance needs. GRACE's ledger is an append-only log in which each record embeds the hash of its predecessor (hash chain), and periodic Merkle-tree roots are signed and retained — optionally anchored to an external transparency service [20][21]. Any alteration of a historical record breaks every subsequent hash; auditors holding the sequence of roots can recompute and verify the entire chain at review time without trusting the monitoring team's database. This directly supports audit-evidence reliability expectations and the record-retention obligations of 2 CFR 200.334-337 [1][22][23], while keeping deployment within ordinary public-sector infrastructure.

Human-in-the-loop triage rather than automated adverse action. GRACE never issues sanctions, disallowances, or award decisions autonomously. Scores route cases to analyst review (Figure 4): below 30, events auto-close with the explanation retained; 30–70 trigger desk review within five business days; above 70 trigger enhanced review with subrecipient engagement and, where warranted, site review — consistent with the risk-based, tailored monitoring of 200.332 [1] and the human-oversight expectations of M-24-10 [3].

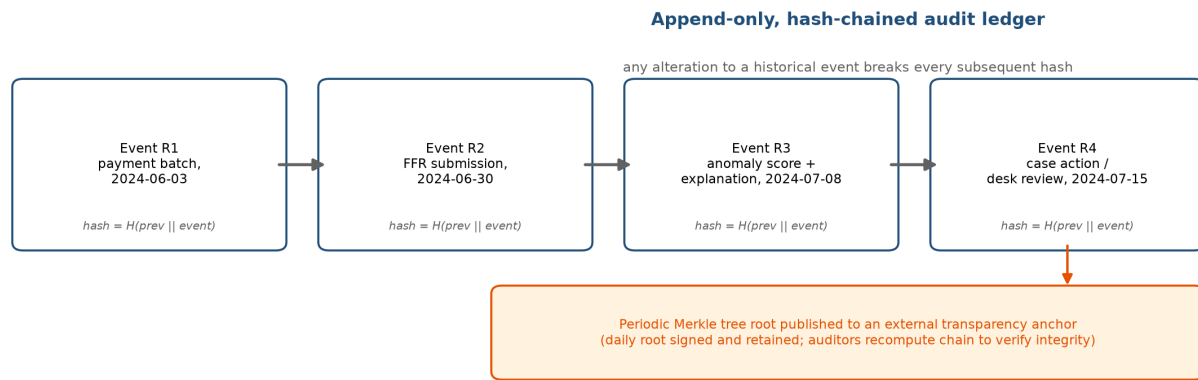


Figure 2. Tamper-evident evidence ledger: append-only hash chain over monitoring events with periodic signed Merkle-root anchoring [20][21].

Requirement (2024 corpus)	Source	GRACE component addressing it
Risk evaluation of subrecipients; monitoring tailored to risk	2 CFR 200.332 [1]	Cross-system data plane; ensemble risk scoring; cadence workflow (Fig. 4)
Monitoring actions documented and demonstrable; records retained	2 CFR 200.332; 200.334-337 [1]	Tamper-evident evidence ledger (Fig. 2)
Explainability, testing, and human oversight where AI is used	OMB M-24-10 [3]; NIST AI RMF [5]	Explanation layer; human-in-the-loop triage; governance module
AI impact assessment and ongoing monitoring	M-24-10 [3]; AI RMF Measure [5]	Drift and bias monitoring; explanation validation protocol
Increased assurance burden from \$1M Single Audit threshold	2 CFR 200 Subpart F [1][2]	Ledger exports and narrative explanations for audit support
Systematic fraud-risk assessment and response	GAO-15-593SP [6]	Red-flag taxonomy (Table 2); documented triage thresholds

Table 1. Mapping of 2024 regulatory requirements to GRACE components [1][2][3][5][6].

System domain	Feature (example)	Anomaly type	Source pattern
Financial systems	Duplicate / split invoice rate	point + collective	procurement splitting [8][10]
Financial systems	Budget-line utilization variance	contextual	budget misstatement [7]
Financial systems	Drawdown-to-expenditure lag	contextual	cash-management misuse [1]
Financial systems	Fixed-asset purchase concentration (FY-end)	contextual	unallowable equipment costs [7]
Payroll & procurement	Payroll-to-performance mismatch	contextual	unsupported payroll/effort [7]
Payroll & procurement	Personnel-cost month-end spikes	contextual	effort misreporting [8]
Payroll & procurement	Single-vendor concentration	point	collusion / favoritism [10]
Payroll & procurement	Threshold-adjacent purchase clustering	collective	split purchases [10]
Performance & reporting	Output decline versus spend growth	contextual	performance misreporting [7]
Performance & reporting	FFR / report lateness frequency	point	reporting noncompliance [1]
Portfolio context & composite	Prior finding recurrence	point	repeat noncompliance [6]

Portfolio context & composite	Spending spike outside service area	contextual	beneficiary eligibility risk [7]
Portfolio context & composite	Administrative-cost ratio shift	contextual	cost shifting [8]
Portfolio context & composite	Prior-score trajectory	collective	accumulating-risk pattern [12]

Table 2. Cross-system red-flag taxonomy for grant compliance, expressed as machine-usable features.

5. Evaluation (RQ4)

5.1 Regulatory-requirement mapping

Table 1 maps the 2024 regulatory corpus to GRACE components. The mapping shows that the framework is *over-determined* by the regulations in a productive sense: each GRACE component is traceable to at least one specific obligation, and no component exists purely for technical interest. The evidence ledger and explainability layer carry the heaviest regulatory load, jointly satisfying M-24-10's transparency and oversight expectations while producing the documentation trail that risk-based monitoring under 200.332 requires a pass-through entity to be able to demonstrate [1][3][5].

5.2 Synthetic case demonstration

To make the analyst experience concrete — and to validate the explanation pipeline against known ground truth — we constructed a fully synthetic subrecipient case. **Subrecipient S-114** (synthetic) is a mid-sized nonprofit receiving pass-through funds across three programs. The synthetic profile was seeded with three engineered patterns: invoice-splitting to stay under procurement-threshold review limits (17% duplicate/split rate), payroll charges growing 40% year-over-year while reported client-outputs declined, and recurring equipment purchases clustered near fiscal year-end. Benign features were seeded to balance the profile (punctual FFRs, in-budget utilization).

Figure 3(a) shows the feature-attribution explanation for S-114's composite score: the three seeded patterns dominate (+1.82, +1.45, +0.97), the benign features pull negative, and the composite triages to 89/100 — the enhanced-review band (Figure 3(b)). Table 3 presents the accompanying plain-language narrative as an analyst would receive it, and the counterfactual statements a subrecipient would receive when contesting or remediating. Two properties matter for trust: the narrative cites only features present in the evidence ledger, and every element — score, attribution, narrative, analyst disposition — is a ledger record, re-verifiable at audit time (Figure 2).

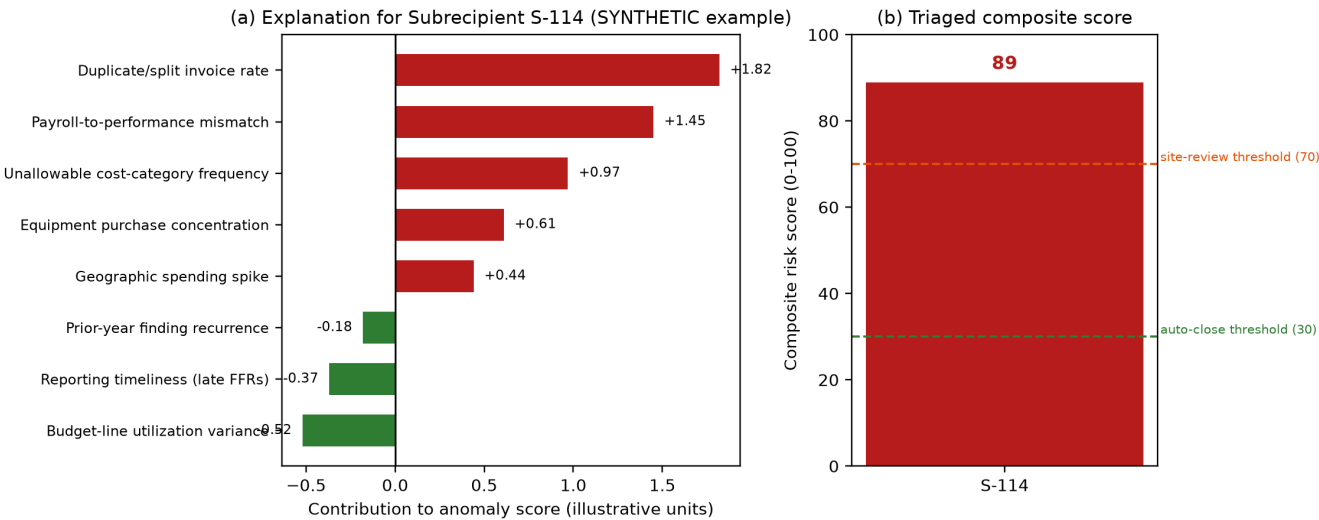


Figure 3. Synthetic case S-114. (a) Feature-level contributions to the anomaly score; (b) composite score against triage thresholds. All data are synthetic and constructed for illustration; no operational subrecipient data was used.

5.3 Triage and workload feasibility

Figure 4 summarizes the intended operational cadence. The design target is that 75–90% of scored events auto-close below threshold, concentrating scarce analyst attention on the 30–70 and >70 bands. This mirrors the empirical structure of alert-driven oversight in adjacent domains and is the principal defense

against the false-positive failure mode documented for unsupervised anomaly detection [11][12]. Thresholds are explicit, documented governance parameters — reviewable by oversight bodies and tunable to entity capacity — rather than hidden model behavior, which is what makes the triage itself auditable.

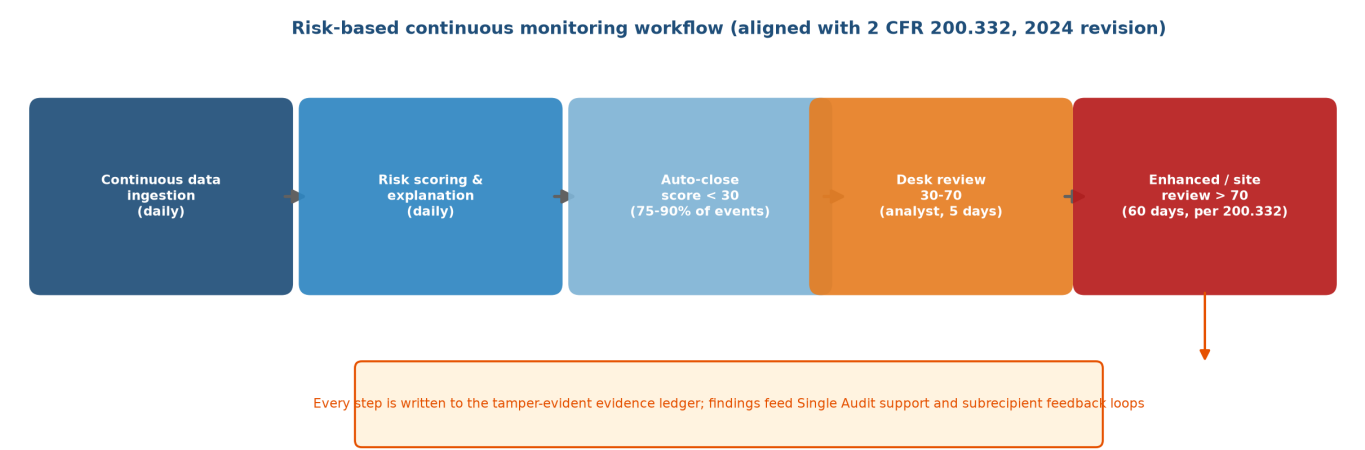


Figure 4. Risk-based continuous monitoring workflow aligned with 2 CFR 200.332 (2024 revision): every stage writes to the evidence ledger.

5.4 Explanation-audience mapping

Table 4 maps the three explanation strata to their audiences and contestability properties. The design principle is that explanations must be *actionable for their audience*: an analyst needs to know why the model moved; a subrecipient needs to know what to change; an auditor needs to know that the record is complete and unaltered; an oversight body needs aggregate, defensible summaries of how the system behaved [14][15][16][19].

Field	Content (illustrative)
Composite score	89/100 - enhanced review band
Top drivers	Duplicate/split invoice rate 17% (matched peer median 3%); payroll cost growth +40% while reported outputs -12%; equipment purchases clustered at fiscal year-end
Contextualization	Scored against matched peer group (nonprofit, 3 programs, \$2-5M annual pass-through)
Counterfactual statement	"Reducing the duplicate/split invoice rate below 2% would place S-114 in the desk-review band"
Evidence links	Ledger records R-3311 to R-3329 (hash-verified)
Required action	Enhanced review within 60 days; subrecipient engagement; no automated adverse action

Table 3. Synthetic case S-114: analyst-facing finding narrative (illustrative; synthetic data).

Explanation stratum	Primary audience	Purpose	Contestability property
Feature attributions (SHAP-style)	Monitoring analysts	Debug and triage the score	Validated against seeded cases and confirmed case outcomes [14][16]
Counterfactual statements	Subrecipients	Concrete remediation path	Recomputable on request; changes track the ledger [15]
Plain-language finding narrative	Auditors, oversight bodies	Record and accountability	Pinned to hash-verified ledger records [19][21]

Table 4. Explanation strata, audiences, and contestability properties.

6. Discussion

6.1 From episodic review to continuous, defensible oversight

The 2024 Uniform Guidance revision asks pass-through entities to *demonstrate* risk-based monitoring, not merely assert it [1][2]. GRACE's contribution to that demonstration is structural: continuous scoring with retained explanations and tamper-evident provenance produces, as a by-product of normal operation, the documentation that episodic oversight struggles to reconstruct after the fact. In digital-trust terms, the framework relocates trust from the diligence of individual reviewers to verifiable properties of the socio-technical system — reasons that can be examined, records that cannot be silently altered, and human decisions that are preserved in context.

6.2 Trustworthy-AI governance as an asset, not a tax

It is tempting to view M-24-10 and the AI RMF as compliance overhead; GRACE's design shows the opposite for this domain [3] [5]. The governance artifacts — impact assessments, drift monitoring, explanation validation, human-oversight records — are precisely the artifacts that make anomaly-based oversight *admissible and defensible* when a subrecipient contests a finding or when an oversight body reviews the pass-through entity itself. A monitoring system that cannot explain itself is an operational and legal liability in an environment where flagged findings have due-process consequences.

6.3 Equity and error-distribution considerations

Risk-scoring systems inherit the biases of their training distributions. Subrecipients with smaller administrative capacity, first-time awardees, or rural organizations may exhibit feature distributions that score as anomalous without any wrongdoing — the classic false-positive inequity of anomaly-based systems [12][16]. GRACE mitigates but does not eliminate this: rule components encode documented patterns rather than learned stereotypes; the taxonomy separates context-sensitive features (scored against matched peer groups) from global ones; and governance requires periodic bias/drift reporting to oversight. Empirical measurement of error distribution across subrecipient classes is specified as a mandatory pre-deployment activity in the AI RMF "Measure" function [5].

6.4 Adoption path for pass-through entities

The framework is intentionally implementable in stages: (1) ledger-first (stand up tamper-evident logging of existing monitoring actions — immediate audit benefit, no model required); (2) taxonomy-first (rule-based scoring over the Table 2 features with explanations); (3) model-first (add the unsupervised component and drift monitors); (4) governance-first throughout. Each stage yields standalone value, allowing small agencies to begin without a data-science capability, and the staged path maps cleanly onto the AI RMF's Govern-Map-Measure-Manage lifecycle [5].

7. Limitations and Threats to Validity

Construct validity. The regulatory mapping (Table 1) was constructed by document analysis of the 2024 corpus; interpretations of 2 CFR 200.332's monitoring provisions and M-24-10's applicability to state-administered systems rest on the authors' reading and should be confirmed with counsel for specific deployments. **External validity.** The evaluation is analytical (mapping) plus synthetic (S-114); no field deployment, no real subrecipient data, and no measured false-positive rates are reported. The taxonomy's features are computable in principle; their actual availability, quality, and linkage across heterogeneous state systems is the dominant practical risk and varies by entity. **Internal validity.** The synthetic demonstration was constructed by the same authors who designed the framework and therefore cannot reveal design blind spots; the seeded patterns were chosen to be representable, not adversarial. Explanation fidelity is asserted through the validation protocol, not demonstrated on learned models. **Reliability.** Threshold values (30/70), auto-close percentages (75–90%), and cadence timelines are design targets and governance parameters, not empirical findings. **Temporal scope.** The analysis is frozen at the 2024 regulatory corpus; the Uniform Guidance is subject to further revision, and OMB AI policy has been, and should be expected to be, revised across administrations — the framework therefore treats regulatory anchors as configuration, not constants.

8. Conclusion

The stewardship of public funds is one of the oldest forms of institutional trust, and its digital transformation has outpaced its oversight. The 2024 regulatory environment — a Uniform Guidance that makes risk-based, continuous subrecipient monitoring the expectation, and an AI governance regime that makes explainability and human oversight mandatory — defines the terms on which that gap can be closed responsibly. This paper contributed GRACE, an explainable AI framework that fuses cross-system evidence

into continuously scored, feature-explained compliance risk, preserved on a tamper-evident evidence ledger that auditors and oversight bodies can independently verify. Its synthetic demonstration shows the intended analyst experience end-to-end; its regulatory mapping shows that every component is traceable to a 2024 obligation; and its stated limitations define the empirical agenda — operational validation on real pass-through data — that must precede deployment. Continuous, explainable, and tamper-evidently documented oversight is not merely a technology upgrade; it is the foundation on which digital trust in public funds can be continuously earned rather than episodically asserted.

Declarations

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRediT contribution statement. Ayesha Arobee: conceptualization, methodology, investigation, framework design, writing – original draft, writing – review & editing.

Competing interests. The author is employed by the Kansas Department for Aging and Disability Services (KDADS), a state agency that administers federal grant funding. No KDADS data, systems, or personnel were used in this research; the framework presented is generic and does not reflect or disclose any operational information of the agency.

Ethics and dual-use. The paper does not disclose live credentials, personal data, or exploit material. The case demonstration uses exclusively synthetic data. Because compliance-scoring systems can affect subrecipients, the design deliberately excludes automated adverse actions (Section 4.2).

Data and code availability. No real-world datasets were used. The synthetic case parameters for S-114 and the red-flag feature definitions (Table 2) are fully specified in the paper; generation code is available from the corresponding author on reasonable request.

AI assistance declaration. The author used AI-based tools to assist with literature discovery, regulatory source verification, figure preparation, and manuscript drafting; all content was reviewed, verified against primary sources, and approved by the author, who takes full responsibility for the work.

Disclaimer. The views expressed are those of the author and do not necessarily represent the views of KDADS or the State of Kansas.

References

- [1] Office of Management and Budget, "Guidance for Federal Financial Assistance," 2 CFR Part 200 (2024 Revision), 89 FR 30046, Apr. 22, 2024 (effective Oct. 1, 2024). [Online]. Available: <https://www.federalregister.gov/documents/2024/04/22/2024-07496/guidance-for-federal-financial-assistance>
- [2] Office of Management and Budget, "Modernizing Federal Grant Guidance to Strengthen Agency Operations and Improve Outcomes," Executive Office of the President, Apr. 2024. [Online]. Available: <https://www.whitehouse.gov/omb/>
- [3] Office of Management and Budget, "Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence," Memorandum M-24-10 for the Heads of Executive Departments and Agencies, Mar. 28, 2024. [Online]. Available: <https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf>
- [4] Executive Order 14110, "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," 88 FR 75191, Nov. 1, 2023 (signed Oct. 30, 2023).
- [5] National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," NIST AI 100-1, Jan. 2023. doi: 10.6028/NIST.AI.100-1.
- [6] U.S. Government Accountability Office, "A Framework for Managing Fraud Risks in Federal Programs," GAO-15-593SP, Jul. 2015. [Online]. Available: <https://www.gao.gov/products/gao-15-593sp>
- [7] U.S. Government Accountability Office, "High-Risk Series: Efforts Made to Achieve Progress Need to Be Maintained and Expanded to Fully Address All Areas," GAO-23-106203, Apr. 2023. [Online]. Available: <https://www.gao.gov/products/gao-23-106203>
- [8] Association of Certified Fraud Examiners, "Occupational Fraud 2024: A Report to the Nations," ACFE, 2024. [Online]. Available: <https://legacy.acfe.com/report-to-the-nations/2024/>
- [9] D. R. Cressey, *Other People's Money: A Study in the Social Psychology of Embezzlement*. Glencoe, IL: Free Press, 1953.
- [10] M. Fazekas and G. Kocsis, "Uncovering high-level corruption: cross-national objective corruption risk indicators using public procurement data," *British Journal of Political Science*, vol. 50, no. 1, pp. 155-164, 2020. doi: 10.1017/S0007123417000461.
- [11] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. on Data Mining (ICDM 2008)*, 2008. doi: 10.1109/ICDM.2008.17.
- [12] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: a survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1-58, 2009. doi: 10.1145/1541880.1541882.
- [13] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proc. NAACL-HLT 2016*, 2016. doi: 10.18653/v1/n16-3020.
- [14] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, 2017. [Online]. Available: <https://arxiv.org/abs/1705.07874>
- [15] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual explanations without opening the black box: automated decisions and the GDPR," *Harvard Journal of Law & Technology*, vol. 31, no. 2, 2017. doi: 10.2139/ssrn.3063289.

- [16] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, pp. 206-215, 2019. doi: 10.1038/s42256-019-0048-x.
- [17] National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile," NIST AI 600-1, Jul. 2024. doi: 10.6028/NIST.AI.600-1.
- [18] ISO/IEC 42001:2023, Information technology - Artificial intelligence - Management system. International Organization for Standardization, 2023.
- [19] J. A. Kroll, J. Huey, S. Barocas, E. W. Felten, J. R. Reidenberg, D. G. Robinson, and H. Yu, "Accountable algorithms," *University of Pennsylvania Law Review*, vol. 165, pp. 633-705, 2017.
- [20] R. C. Merkle, "A digital signature based on a conventional encryption function," in *Advances in Cryptology - CRYPTO '87*, LNCS 293, Springer, 1987. doi: 10.1007/3-540-48184-2_32.
- [21] B. Laurie, A. Langley, E. Kasper, E. Messeri, and R. Stradling, "Certificate Transparency Version 2.0," IETF RFC 9162, Dec. 2021. doi: 10.17487/RFC9162.
- [22] K. Kent and M. Souppaya, "Guide to Computer Security Log Management," NIST SP 800-92, Sep. 2006. doi: 10.6028/NIST.SP.800-92.
- [23] National Institute of Standards and Technology, "Security and Privacy Controls for Information Systems and Organizations," NIST SP 800-53 Rev. 5, Sep. 2020. doi: 10.6028/NIST.SP.800-53r5.