



(REVIEW ARTICLE)



Cross agency data fusion and predictive intelligence to reduce fraud across U.S benefits programs

Omotoso Samuel Sunday ^{1,*} and Oluwabusayo Olufunke Awoyomi ²

¹ Department of Information Systems, University of Arkansas.

² The College of Saint Rose, Department of Computer Science.

World Journal of Advanced Research and Reviews, 2026, 30(02), 101-111

Publication history: Received on 22 March 2026; revised on 26 April 2026; accepted on 29 April 2026

Article DOI: <https://doi.org/10.30574/wjarr.2026.30.2.1167>

Abstract

Fraud and improper payments across U.S. federal benefits programs impose substantial fiscal and social costs, with the Government Accountability Office estimating annual fraud-related losses of \$233–521 billion and cumulative improper payments exceeding \$2.7 trillion since fiscal year 2003. Traditional "pay-and-chase" recovery models have proven insufficient, prompting growing interest in preventive strategies that leverage cross-agency data fusion and predictive intelligence. This paper presents a qualitative comparative analysis of four dominant institutional and technical models used to integrate data and apply predictive analytics for fraud reduction: centralized federal data hubs, federated and privacy-preserving architectures, vendor-managed identity and fraud detection systems, and state-level or program-specific predictive analytics. Each model is evaluated against five criteria derived from federal oversight guidance and academic literature: legal and institutional authority, data governance and privacy protection, technical architecture and scalability, analytic effectiveness, and transparency and accountability. The analysis demonstrates that no single model adequately balances these competing demands. Centralized hubs offer statutory grounding but limited analytic flexibility; federated approaches strengthen privacy at the cost of governance complexity; vendor systems raise accountability concerns; and program-specific models lack cross-jurisdictional reach. The paper proposes a hybrid framework combining centralized authoritative checks, federated analytics for sensitive data, and program-level predictive modeling under unified governance with human-in-the-loop decision-making. Policy recommendations address statutory authorization, oversight structures, model validation, vendor accountability, and workforce capacity to support responsible, equitable, and scalable fraud prevention.

Keywords: Cross-agency data fusion; Predictive intelligence; Improper payments; Federal benefits programs; Payment integrity; Privacy-preserving analytics

1. Introduction

Fraud and improper payments in federal benefits programs present a significant and enduring challenge for the United States government, eroding public trust, misallocating taxpayer dollars, and hampering the equitable delivery of services to eligible populations. Improper payments defined as payments that should not have been made or were made in incorrect amounts are pervasive across federal agencies, and while not all improper payments result from fraud, fraudulent activity constitutes a substantial subset of these financial losses (GAO, 2025; GAO, 2024). Between fiscal years 2018 and 2022, the U.S. Government Accountability Office (GAO) estimated that fraud alone cost federal programs between \$233 billion and \$521 billion annually, underscoring the magnitude of the problem and the urgent need for more effective prevention strategies (GAO, 2025). The persistence of fraud and improper payments is not a trivial administrative issue; it reflects structural deficiencies in how programs manage risk, verify eligibility, and coordinate oversight. Since fiscal year 2003, cumulative improper payment estimates for executive branch agencies have surpassed

* Corresponding author: Omotoso Samuel Sunday

\$2.7 trillion, reflecting both the scale of benefits distributed and systemic vulnerabilities in existing payment integrity frameworks (GAO, 2024). In the context of this challenge, the traditional “pay and chase” model in which funds are disbursed before fraud is detected and recovered is increasingly viewed as both inefficient and insufficiently proactive. This has led lawmakers and practitioners to explore preventive mechanisms that leverage data sharing, analytics, and cross-program coordination to detect fraud before it occurs (House Oversight Committee, 2025).

A central component of these efforts is cross-agency data fusion: the integration or harmonization of disparate datasets from multiple federal and state entities to create a comprehensive and unified analytic foundation for fraud detection. When appropriately implemented, data fusion can reveal patterns of behavior, duplicate claims, and anomalies that would be invisible within the siloed views maintained by individual programs. For example, linking benefits claims with death records can prevent payments to deceased recipients, a long-standing source of improper payments that agencies have struggled to address effectively in isolation (GAO, 2025). Complementing data fusion, predictive intelligence often enabled by machine learning (ML), artificial intelligence (AI), and advanced statistical modeling offers the potential to analyze large volumes of administrative and behavioral data to anticipate fraud risk, prioritize cases for human review, and adapt to evolving fraud tactics. Predictive models can identify patterns indicative of fraudulent claims by learning from historical data, extracting subtle correlations, and generating risk scores that inform program integrity workflows. These capabilities are increasingly attractive as agencies seek to move beyond simple rule-based checks toward more nuanced and dynamic fraud detection strategies.

However, the application of cross-agency data fusion and predictive intelligence in benefits program oversight presents complex legal, technical, and ethical challenges. On the legal front, data sharing across agencies must navigate a labyrinth of statutory authorities, privacy protections, and governance requirements. While some models of cross-agency collaboration such as the Department of the Treasury’s Do Not Pay working system have statutory authorization and broad interagency buy-in, other approaches require negotiating interagency memoranda of understanding and ensuring compliance with privacy laws and Office of Management and Budget (OMB) guidance (Congressional Research Service, 2025). Furthermore, the increased aggregation of personally identifiable information (PII) raises concerns related to the mosaic effect, where seemingly innocuous datasets, when combined, can reveal sensitive information and pose privacy risks (Mosaic Effect, 2025).

Technical challenges also abound. Predictive models are only as effective as the data they are trained on, and poor-quality or inconsistent data can degrade model performance and erode confidence in analytical outputs. The GAO has noted that while AI and advanced analytics offer opportunities for enhancing fraud detection, their reliability depends on high-quality, well-structured data, and agencies often lack the necessary data infrastructure and skilled workforce to manage and interpret complex analytical systems (GAO, 2025). Without sufficient expertise in data science, statistics, and AI governance, federal agencies risk deploying models that generate high false-positive rates, entrench bias, or lack transparency raising both operational and ethical concerns. In addition to technical requirements, effective cross-agency approaches must address governance and accountability. The sheer scale of federal benefit programs spanning healthcare, nutrition assistance, income support, and tax credits complicates efforts to standardize data practices and risk management. The Federal Data Strategy and related OMB guidance emphasize the importance of establishing clear governance frameworks, data stewardship roles, and transparent processes for data access and use across agencies. These frameworks are intended to balance the imperative of preventing fraud with the obligation to protect beneficiary privacy and ensure due process (OMB Federal Data Strategy, 2025). Governance challenges are amplified by the diversity of program administration. Some benefits, such as Medicaid and SNAP, are jointly managed by federal and state entities, requiring coordination not only across federal agencies but also across state boundaries. This complexity has motivated the development of state-level data systems capable of integrating federal eligibility data with state administrative records to support more effective fraud detection and eligibility verification. For instance, Utah’s eFind system aggregates data from multiple databases, enabling real-time updates and cross-program validation of eligibility criteria across SNAP, Temporary Assistance for Needy Families (TANF), and other benefits (Utah eFind, 2025). Beyond governance and technical integration, the rise of predictive analytics in government has prompted broader questions about fairness, transparency, and the potential for disparate impacts on vulnerable populations. While advanced tools can improve detection performance and efficiency, stakeholders and researchers have raised concerns about algorithmic bias, opaque decision logic, and inadequate mechanisms for redress when individuals are incorrectly flagged or denied benefits. These concerns suggest that technical solutions must be complemented by ethical safeguards and oversight structures that ensure accountability and protect beneficiaries’ rights.

Despite these obstacles, the potential gains from effective cross-agency data fusion and predictive modeling are substantial. Advanced analytics not only promises to reduce financial losses due to fraud and improper payments but also to improve program integrity processes in a way that conserves resources and enhances public confidence. During the COVID-19 pandemic, for example, analytics platforms used by the Pandemic Response Accountability Committee

(PRAC) enabled the analysis of over 150 million records, illustrating the scale and value of integrated data systems in fraud detection and prevention (Congressional Research Service, 2023). Federal agencies have also demonstrated measurable results from enhanced analytical efforts; in early 2025, Treasury reported that a pilot linking Social Security Administration (SSA) death data with the Do Not Pay system recovered \$31 million by identifying improper payments to deceased individuals (GAO, 2025). This paper undertakes a comparative analysis of leading models and frameworks for cross-agency data fusion and predictive intelligence applications in federal benefits programs. By systematically evaluating centralized hubs, federated data architectures, vendor-managed solutions, and state-level predictive systems, the analysis aims to elucidate the strengths and limitations of each approach across dimensions such as legal compliance, privacy risk, technical feasibility, and operational effectiveness. Through this comparison, the paper seeks to provide policymakers, program administrators, and researchers with insights that can guide the design and implementation of responsible, effective anti-fraud strategies that protect both government resources and beneficiary rights.

2. Background and Policy Context

2.1. Overview of Fraud and Improper Payments in U.S. Benefits Programs

Fraud and improper payments are long-standing and significant challenges across federal benefits programs, directly affecting both fiscal stewardship and public trust in government administration. Improper payments are defined as federal outlays that “*should not have been made or were made in incorrect amounts under statutory, contractual, administrative, or other legally applicable requirements*” (GAO, 2025). Fraud typically a subset of improper payments occurs when benefits are obtained or attempted to be obtained *through willful misrepresentation or deception* (GAO, 2025). These issues have persisted for decades and continue to impose substantial costs on the federal government. According to the U.S. Government Accountability Office (GAO), federal agencies have reported cumulative improper payment estimates of roughly \$2.8 trillion since fiscal year 2003, including hundreds of billions of dollars per year in more recent reporting periods (GAO, 2025; GAO, 2024). Additionally, the GAO estimates that government-wide fraud costs range from \$233 billion to \$521 billion annually, based on data from fiscal years 2018 through 2022, though the actual amounts could be higher due to reporting limitations (GAO, 2025). High-priority programs such as Medicare, Medicaid, Unemployment Insurance, Supplemental Security Income, and the Earned Income Tax Credit account for a disproportionate share of improper payments, reflecting both their large scale and inherent risk factors associated with eligibility and payment complexity (Congressional Research Service, 2024).

Improper payments often arise from errors in eligibility determination or payment processing, whereas fraud involves deliberate deception. For example, in Unemployment Insurance, a claimant may knowingly conceal employment or earnings to receive benefit payments to which they are not entitled a type of fraud that can vary in definition and enforcement practices across states (Wikipedia, 2025). Both categories of improper payments and fraud undermine fiscal integrity and degrade program performance outcomes, highlighting the need for more robust detection and prevention mechanisms.

2.2. Federal Oversight and Accountability Mechanisms

The U.S. government has established a multi-layered framework for oversight and accountability designed to monitor, report on, and reduce improper payments and fraud across federal programs. Key actors in this framework include the GAO, the Office of Management and Budget (OMB), and Inspectors General (IGs) across executive branch agencies. The Government Accountability Office (GAO) regularly conducts and publishes comprehensive reviews of improper payments and fraud, identifies root causes, and issues recommendations to both Congress and federal agencies. GAO’s work is frequently cited by policymakers and used as a basis for legislative and administrative reforms aimed at strengthening payment integrity (GAO, 2025). These reports emphasize the importance of prevention, data sharing, risk assessment, and the use of technology to improve fraud detection (GAO, 2025). The OMB plays a central coordinating role in establishing government-wide policy on improper payments and data governance. Under statutory authority and executive guidance, OMB issues directives and circulars that set expectations for agencies’ estimation, reporting, and mitigation of improper payments. OMB’s role extends to developing analytic standards, guiding corrective action plans, and overseeing agencies’ compliance with relevant statutes such as the Payment Integrity Information Act (PIIA). Through platforms such as PaymentAccuracy.gov, OMB also enhances transparency by publishing program-level performance scorecards and interventions to reduce improper payments (PaymentAccuracy.gov, 2025).

Inspectors General (IGs) supplement ongoing oversight by conducting independent audits and evaluations of agency compliance with improper payment and fraud reduction requirements. The IGs are mandated to assess their agencies’ adherence to statutory reporting and internal control standards, and to report findings to Congress. Where agencies are

found noncompliant, IGs must document corrective action plans and proposed steps to come into compliance with federal law.

2.3. Legislative and Regulatory Foundations

The contemporary federal approach to combating improper payments and fraud is grounded in several legislative mandates and policy frameworks that establish requirements for data **management, reporting, and evidence-based decision-making**.

2.3.1. Payment Integrity Information Act (PIIA)

Enacted in 2019, the Payment Integrity Information Act of 2019 (PIIA) represents the primary legislative foundation governing federal improper payment estimation, reporting, and accountability. PIIA consolidated and updated previous statutes designed to reduce improper payments and introduced new requirements for risk assessment, corrective actions, and transparency (Congressional Research Service, 2024). Specifically, PIIA mandates that executive agencies:

- *Assess the risk of significant improper payments* in programs exceeding statutory spending thresholds;
- *Publish estimates, corrective action plans, and reduction targets* in annual financial statements; and
- *Report on recovery audit programs* when cost-effective (Congressional Research Service, 2024).

Under PIIA, OMB is tasked with issuing guidance on risk assessment methodology and compliance criteria, and agency IGs must evaluate agency conformity with these requirements in annual reports. Noncompliance can trigger mandatory reporting of corrective action plans to congressional authorizing and appropriations committees.

2.3.2. Foundations for Evidence-Based Policymaking Act (Evidence Act)

The Foundations for Evidence-Based Policymaking Act of 2018 (commonly the *Evidence Act*) seeks to modernize federal data management and evidence-building practices across government. Its objectives include improving data interoperability, standardizing open data practices, and enhancing agencies' capacity to use data and analytics to inform policy and operational decisions (Wikipedia, 2025). The Evidence Act also reauthorized the Confidential Information Protection and Statistical Efficiency Act and incorporated the *OPEN Government Data Act* as Title II, which requires agencies to publish open, machine-readable data assets unless restricted by privacy or security concerns. Implementation of the Evidence Act has furthered cross-agency data initiatives and created a structural basis for agencies to consider analytic and evidence needs as part of program planning and evaluation. The Act's emphasis on integrating and leveraging data assets aligns with efforts to improve fraud detection and prevention through data sharing and evidence generation.

2.3.3. Federal Data Strategy and OMB Guidance

Complementing the Evidence Act, the Federal Data Strategy provides a ten-year framework of principles and practices designed to help federal agencies manage and use data as a strategic asset while protecting security, privacy, and confidentiality. Established in OMB Memorandum M-19-18 (2019), the strategy outlines cross-cutting principles such as ethical governance, conscious design, and a learning culture, and enumerates specific actions that agencies should adopt to improve data quality, coordination, and sharing across government (Federal Data Strategy, 2019). Key elements of the Federal Data Strategy relevant to cross-agency fraud detection include *coordinating federal data assets, sharing data between state, local, tribal, and federal levels, and ensuring appropriate data use practices* that maximize program outcomes while safeguarding privacy (Federal Data Strategy Practices, 2025). These practices provide a policy foundation for systematic data linking, analytics, and cross-program insights that are central to predictive fraud detection.

OMB continues to refine data management guidance to clarify expectations for agency data structures, metadata standards, and open data publication requirements. Recent guidance emphasizes that federal data should be "open by default," subject to privacy and security considerations, strengthening the policy basis for sharing and integrating data across programs to foster evidence-based oversight (OMB guidance 2025).

2.4. Rationale for Cross-Agency Collaboration

The scale and complexity of improper payments and fraud underscore the rationale for enhancing cross-agency collaboration. Many forms of fraud involve actors or patterns that transcend individual program boundaries, rendering siloed data and isolated analytic efforts insufficient. For example, fraudulent claim activity may involve coordinated exploitation of benefits across health care, income support, and tax credit programs signal patterns that become detectable only when data are consolidated across administrative domains. Cross-agency data fusion enables the

identification of such multi-program patterns, filling analytic gaps that single-agency systems cannot address on their own (GAO, 2025). Moreover, collaborative data sharing is supported by policy frameworks that emphasize the strategic value of federal data assets for mission outcomes and accountability. By promoting interoperable data infrastructures and standardized governance practices, federal policy frameworks such as the Federal Data Strategy and the Evidence Act help create conducive conditions for age

In addition to governance frameworks, there is growing recognition that preventing fraud before payments are made is both fiscally and socially preferable to after-the-fact recovery efforts. Predictive and preventive models depend on access to a wide range of data inputs, from eligibility determinations to cross-program identifiers, making collaboration a practical imperative to support more advanced analytic capabilities. Collectively, these policy and operational rationales form the foundational context for pursuing cross-agency data fusion and predictive intelligence as part of an integrated approach to reducing fraud and improving payment integrity across U.S. benefits programs.

3. Conceptual Framework

This section outlines the conceptual framework that underpins the comparative analysis of cross-agency data fusion and predictive intelligence in reducing fraud across U.S. benefits programs. The framework clarifies the meaning and scope of cross-agency data fusion, explains how predictive intelligence is applied in public administration, and situates both within broader governance, accountability, and ethical considerations. Establishing these concepts is essential for evaluating different institutional and technical models on consistent analytical grounds. Cross-agency data fusion refers to the systematic integration, linkage, or coordinated analysis of data originating from multiple governmental entities in order to generate insights that cannot be obtained from isolated datasets. In the administration of public benefits, this typically involves the use of data from federal agencies, state governments, and, in some cases, external partners to support eligibility verification, payment oversight, and fraud detection. Unlike basic data sharing, which often occurs on a limited or ad hoc basis, data fusion implies a sustained and structured process oriented toward analytical outcomes, such as identifying duplicate participation across programs or detecting coordinated fraudulent activity. Federal oversight bodies, including the Government Accountability Office, increasingly characterize data fusion as a strategic capability rather than a purely technical function, emphasizing its role in preventive program integrity efforts (GAO, 2025).

The scope of cross-agency data fusion varies significantly depending on legal authority, governance structures, and technical capacity. Some models rely on centralized architectures in which data from multiple agencies are aggregated into a single system for analysis, while others employ decentralized or federated approaches that allow agencies to retain control over their data while participating in shared analytical processes. These architectural differences are not merely technical choices; they reflect underlying policy decisions about privacy, accountability, and institutional trust. As a result, the degree and method of data fusion become critical dimensions for comparative evaluation.

Data fusion efforts in benefits programs typically draw on several categories of administrative data, including identity and demographic information, eligibility and enrollment records, payment and transaction data, and historical program integrity findings. Identity data enable record linkage across systems, while eligibility and enrollment data provide context for assessing compliance with program rules. Payment data support the detection of anomalous patterns, such as repeated disbursements to the same account across multiple programs. Program integrity data, including prior fraud determinations and recovery actions, are particularly valuable for training predictive models but are often inconsistently captured across agencies. The analytical value of combining these data sources is substantial, yet their integration raises significant challenges related to data quality, standardization, and privacy protection.

Within this data environment, predictive intelligence refers to the use of statistical and computational techniques to estimate fraud risk and support decision-making in program oversight. Predictive intelligence encompasses a range of methods, from traditional statistical models to more advanced machine learning approaches capable of identifying complex, non-linear relationships in large datasets. In contrast to rule-based systems, which rely on predefined thresholds and conditions, predictive models infer patterns from historical data and adapt as new information becomes available. Federal agencies have increasingly explored these tools as a means of moving from reactive “pay-and-chase” strategies toward proactive fraud prevention (OMB, 2024).

However, predictive intelligence in public administration is best understood as a complement to, rather than a replacement for, existing rules and controls. Most operational systems employ hybrid approaches in which statutory eligibility rules establish baseline compliance, while predictive models generate risk scores or alerts that guide investigative priorities. Predictive intelligence can be implemented using supervised learning techniques, which rely on labeled examples of known fraud, or unsupervised techniques, which identify anomalies without prior labels. Each

approach has advantages and limitations. Supervised models may replicate historical enforcement biases, while unsupervised models may flag legitimate but atypical behavior. Consequently, human judgment remains a critical component of any predictive fraud detection system. A central conceptual distinction in the public sector concerns the role of predictive intelligence in decision-making. Predictive systems may function as decision-support tools, providing information to human reviewers, or as automated decision systems that directly trigger enforcement actions such as payment suspension or benefit termination. Federal guidance and oversight bodies have consistently emphasized the importance of maintaining human-in-the-loop processes, particularly when decisions affect access to essential benefits. This distinction is essential for evaluating the appropriateness of different models, as automated enforcement raises heightened concerns related to due process, transparency, and accountability.

Governance and accountability considerations form a critical component of the conceptual framework. Cross-agency data fusion and predictive intelligence operate within a complex legal and institutional environment shaped by statutes such as the Payment Integrity Information Act and the Evidence-Based Policymaking Act, as well as guidance issued by the Office of Management and Budget. These frameworks require agencies to document the purpose and authority for data use, safeguard personally identifiable information, and demonstrate the effectiveness of analytic interventions. Governance structures also determine how responsibilities are allocated among agencies, how access to data is controlled, and how analytic outputs are reviewed and acted upon.

Transparency and explainability are particularly salient in the context of predictive intelligence. Oversight bodies such as Inspectors General and the GAO require agencies to be able to explain how fraud detection systems operate, how decisions are made, and how errors are identified and corrected. Explainable systems support auditing, enable legal review, and provide a basis for redress when individuals are adversely affected. Conversely, opaque or proprietary systems complicate oversight and undermine public trust, especially when they influence eligibility or payment decisions.

Finally, the framework recognizes the ethical and equity implications of fraud detection in benefits programs. Predictive systems operate within social contexts marked by economic vulnerability, and errors can have serious consequences for individuals who rely on public assistance. False positives may result in delayed payments, administrative burdens, or stigmatization, disproportionately affecting marginalized populations. As a result, considerations of fairness, bias mitigation, and procedural justice are integral to evaluating the design and deployment of cross-agency fraud detection systems. Taken together, these conceptual dimensions modes of data fusion, forms of predictive intelligence, and governance and ethical constraints provide the foundation for the comparative analysis that follows. By applying this framework, the study assesses not only the technical effectiveness of different models, but also their legal feasibility, institutional sustainability, and societal impact.

4. Methodology of Comparative Analysis

This study employs a qualitative comparative analysis to examine alternative models of cross-agency data fusion and predictive intelligence used to reduce fraud and improper payments in U.S. benefits programs. Rather than evaluating a single system or program, the analysis compares multiple institutional and technical approaches that have been implemented or proposed within the federal and state administrative landscape. This methodology is appropriate given the diversity of legal authorities, governance structures, and operational constraints that shape fraud detection efforts across programs. The comparative analysis is guided by a set of evaluative criteria derived from federal oversight guidance, academic literature on public-sector analytics, and documented challenges identified by the Government Accountability Office (GAO) and the Office of Management and Budget (OMB). These criteria are designed to capture not only technical effectiveness, but also legal feasibility, governance quality, and societal implications. The analysis therefore moves beyond narrow performance metrics to assess how different models align with public-sector values such as accountability, transparency, and equity.

Five primary criteria structure the comparison. First, legal and institutional authority assesses whether a model operates under explicit statutory authorization or relies on discretionary agreements such as memoranda of understanding. This criterion reflects the central role of law in shaping permissible data use and system design in government. Second, data governance and privacy protection evaluates how models manage personally identifiable information, limit access, and mitigate privacy risks, including risks associated with large-scale data aggregation. Third, technical architecture and scalability considers whether models rely on centralized, federated, or hybrid data infrastructures and evaluates their capacity to scale across agencies, programs, or jurisdictions. Fourth, analytic effectiveness examines the extent to which models support meaningful fraud detection, including their ability to identify cross-program patterns, adapt to evolving fraud schemes, and prioritize cases for review. Finally, transparency and

accountability evaluates whether analytic processes are explainable, auditable, and subject to meaningful oversight by Inspectors General, GAO, and other accountability institutions.

The analysis draws on publicly available government reports, oversight testimony, policy guidance, and documented program examples rather than proprietary or classified system evaluations. This approach ensures transparency and replicability while acknowledging limitations related to the availability of standardized performance data across programs. Importantly, the study does not seek to rank models as universally superior, but rather to illuminate trade-offs and contextual suitability across different administrative environments.

5. Comparative Models of Cross-Agency Data Fusion and Predictive Intelligence

Building on the methodological framework established in Section 4, this section compares four dominant models used to integrate data and apply predictive intelligence for fraud reduction in U.S. benefits programs. Each model reflects a distinct approach to balancing analytic capability, legal authority, and governance constraints.

5.1. Centralized Federal Data Hub Model

The centralized federal data hub model consolidates data from multiple agencies into a single platform that performs standardized eligibility checks and fraud-related screenings. The most prominent example is the U.S. Department of the Treasury's *Do Not Pay* system, which aggregates data such as death records, exclusion lists, and incarceration information to prevent improper payments before funds are disbursed (U.S. Treasury, 2023; GAO, 2019). From a legal and institutional perspective, centralized hubs benefit from explicit statutory authorization and government-wide policy support, which facilitates adoption across agencies (Moynihan, Herd, & Harvey, 2015). Governance structures are typically formalized, with defined roles for data providers, system operators, and oversight bodies, enhancing accountability and auditability (GAO, 2020).

However, centralized models face notable limitations. Their effectiveness is strongest for detecting known and well-defined improper payment scenarios, such as payments to deceased individuals, but weaker for identifying complex or adaptive fraud schemes that require behavioral or network analysis (Button & Brooks, 2016). Additionally, data centralization concentrates privacy and cybersecurity risks, increasing the consequences of potential breaches and raising concerns about the long-term accumulation of sensitive personal data (Kuner, Cate, Millard, & Svantesson, 2017).

5.2. Federated and Privacy-Preserving Data Models

Federated models represent an alternative approach in which agencies retain control over their data while participating in shared analytical processes. Rather than transferring raw data to a central repository, agencies exchange encrypted identifiers, tokens, or analytic outputs, or participate in secure multiparty computation frameworks (Hardy et al., 2017). This approach aligns closely with privacy-by-design principles and is often proposed when statutory or ethical constraints limit direct data sharing (Cavoukian, 2011).

From a governance standpoint, federated models offer strong privacy protections and reduce the risks associated with centralized data storage, making them particularly suitable for sensitive datasets such as health or disability records (Rieke et al., 2020). However, these models require substantial technical coordination, shared standards, and trust among participating agencies, which can slow deployment and complicate oversight (Janssen, Charalabidis, & Zuiderwijk, 2012). Analytically, federated models can support sophisticated pattern detection across programs, but their effectiveness depends heavily on interoperability and data quality. Transparency and auditability may also be more challenging, as oversight bodies must trace analytic decisions across distributed systems rather than a single platform (Veale & Borgesius, 2021).

5.3. Vendor-Managed Identity and Fraud Detection Systems

Vendor-managed systems rely on private-sector contractors to perform identity verification, fraud screening, or risk scoring on behalf of government agencies. These systems are often adopted to accelerate deployment and leverage specialized technical expertise, particularly in identity verification and biometric analysis (Whittaker et al., 2018). While vendor-managed systems can offer rapid scalability and high transaction throughput, they raise significant governance and accountability concerns.

Agencies may have limited visibility into proprietary algorithms, constraining explainability and oversight (Burrell, 2016). Public scrutiny and congressional investigations into certain vendor implementations have highlighted risks

related to transparency, due process, and equitable access to benefits (House Committee on Oversight and Reform, 2021). From a comparative perspective, vendor-managed models score relatively high on short-term operational efficiency but lower on public-sector accountability criteria. Their reliance on contractual rather than statutory authority further complicates long-term sustainability and interagency interoperability (Kroll, Huey, Barocas, Felten, Reidenberg, Robinson, & Yu, 2017).

5.4. State-Level and Program-Specific Predictive Analytics Models

State-level and program-specific models involve agencies developing their own predictive analytics systems tailored to specific benefits programs such as Medicaid, SNAP, or Unemployment Insurance. These systems typically integrate local administrative data and apply predictive models to identify high-risk claims or recipients (Eubanks, 2018). These models benefit from contextual knowledge and flexibility, allowing agencies to adapt analytics to program-specific rules and fraud patterns (Bannister & Connolly, 2020). They are often easier to pilot and iterate than federal-scale systems.

However, their fragmentation limits their ability to detect cross-program or cross-jurisdictional fraud, and governance practices vary widely across states (GAO, 2022). Concerns about fairness and algorithmic bias have been particularly salient in this model category, as some state implementations have been criticized for opaque decision-making and insufficient safeguards for beneficiaries (Citron & Pasquale, 2014). As a result, while program-specific models can be analytically effective, they pose challenges for consistency, equity, and federal oversight.

5.5. Comparative Table of Models

Criterion	Centralized Federal Hub	Federated / Privacy-Preserving	Vendor-Managed Systems	State / Program-Specific Models
Legal authority	Strong statutory basis	Context-dependent	Contractual	Program-specific
Privacy risk	Moderate-High	Low-Moderate	Variable-High	Variable
Scalability	High	Moderate	High	Low-Moderate
Analytic sophistication	Moderate	High (potential)	Moderate-High	Moderate
Transparency & auditability	High	Moderate	Low-Moderate	Variable
Equity safeguards	Strong (with oversight)	Strong (design-dependent)	Weak-Moderate	Inconsistent

5.6. Discussion of Comparative Findings

The comparative analysis reveals that no single model optimally satisfies all evaluative criteria. Centralized federal hubs perform well in terms of legal authority, transparency, and oversight, but are limited in analytic flexibility and raise concerns about data concentration. Federated models offer strong privacy protections and analytic potential but face implementation and governance challenges that may hinder widespread adoption. Vendor-managed systems demonstrate the tension between efficiency and accountability, illustrating how private-sector solutions can introduce risks that are difficult to reconcile with public-sector norms. State-level and program-specific models highlight the value of contextual expertise while underscoring the limitations of fragmented approaches to fraud detection. Taken together, these findings suggest that hybrid approaches combining centralized authoritative checks, federated analytics for sensitive data, and program-level contextual modeling may offer the most balanced path forward. Importantly, the success of any model depends not only on technical design, but also on governance, oversight, and sustained institutional capacity.

6. Toward a Hybrid Model and Recommendations

The comparative analysis presented in the preceding sections demonstrates that no single model of cross-agency data fusion and predictive intelligence fully satisfies the competing demands of legal compliance, analytic effectiveness, privacy protection, and public accountability. Centralized federal hubs offer strong statutory grounding and transparency but are analytically constrained, while federated models provide privacy advantages at the cost of technical and governance complexity. Vendor-managed systems raise concerns about oversight and equity, and state-

level predictive systems, though flexible and context-sensitive, are limited in their ability to detect cross-program fraud at scale. These findings suggest that a hybrid model, deliberately combining complementary elements from multiple approaches, offers the most promising path forward for reducing fraud across U.S. benefits programs.

A hybrid model recognizes that different categories of fraud and improper payments require different analytical and institutional responses. Certain high-frequency improper payment risks such as payments to deceased individuals or to ineligible incarcerated recipients are well suited to centralized, authoritative checks that can be applied uniformly across programs. In contrast, more complex or adaptive fraud schemes, including coordinated activity spanning multiple benefits programs or jurisdictions, require analytic flexibility and access to diverse data sources that centralized systems alone cannot provide. By integrating centralized screening with federated and program-level analytics, a hybrid approach can address both routine and sophisticated fraud risks while respecting legal and privacy constraints.

At the core of the proposed hybrid model is the continued use of centralized federal data hubs as authoritative sources for standardized eligibility and exclusion checks. Systems such as the Do Not Pay working system should serve as the first line of defense, providing consistent and auditable screening against nationally maintained datasets. These checks are particularly effective for preventing known categories of improper payments and offer a strong foundation for government-wide payment integrity efforts. Maintaining and expanding the quality and timeliness of these authoritative datasets should remain a policy priority.

Layered on top of centralized checks, the hybrid model incorporates federated or privacy-preserving analytic capabilities to enable cross-agency pattern detection without requiring full data centralization. Under this approach, agencies retain control over sensitive datasets while participating in shared analytical processes through secure linkage mechanisms, encrypted identifiers, or the exchange of analytic outputs such as risk scores. This design aligns with privacy-by-design principles and accommodates statutory limitations on data sharing, particularly for health, disability, and income data. Although federated analytics require greater technical coordination, their inclusion is essential for detecting cross-program fraud patterns that transcend individual agency boundaries.

The hybrid model also preserves a role for program-level predictive analytics, allowing agencies and states to tailor models to the specific risk profiles and administrative realities of individual benefits programs. These localized models can incorporate contextual knowledge, operational constraints, and program-specific indicators that may not be visible in federal datasets. However, to mitigate risks associated with inconsistency and bias, program-level analytics should operate within a common governance framework that establishes minimum standards for transparency, validation, and human review. Federal guidance and oversight can play a critical role in promoting alignment across decentralized implementations.

A defining feature of the hybrid model is its emphasis on human-in-the-loop decision-making. Predictive intelligence should function primarily as a decision-support tool rather than an automated enforcement mechanism. Risk scores and alerts generated by analytic systems should inform investigative prioritization and eligibility review, but final determinations particularly those affecting benefit access should remain subject to human judgment. This approach mitigates due process risks, supports explainability, and ensures that predictive tools enhance rather than replace professional discretion.

Based on this synthesis, several policy and operational recommendations emerge. First, policymakers should explicitly authorize and incentivize hybrid data architectures through statutory and budgetary mechanisms. Existing legal frameworks such as the Payment Integrity Information Act and the Evidence-Based Policymaking Act provide a foundation for such efforts, but clearer guidance on acceptable hybrid models would reduce uncertainty and facilitate adoption. Targeted funding for interoperability and privacy-preserving technologies would further support implementation.

Second, federal agencies should strengthen data governance and oversight structures to support cross-agency collaboration. This includes formalizing roles for chief data officers, privacy officials, and program integrity leads in overseeing analytic systems, as well as ensuring that Inspectors General and the Government Accountability Office have sufficient access to evaluate system design and performance. Governance frameworks should require documentation of data sources, model logic, and decision workflows to support auditability and accountability.

Third, agencies should adopt standardized evaluation and validation practices for predictive intelligence systems. These practices should include regular assessments of model accuracy, false-positive rates, and potential disparate impacts, as well as mechanisms for updating models in response to changing fraud patterns. Transparency in evaluation results

can enhance institutional learning and public trust, while independent validation can reduce the risk of unintended harm.

Fourth, the use of vendor-managed systems within a hybrid model should be carefully constrained and governed. Where private-sector tools are employed, agencies should require contractual provisions that ensure transparency, audit rights, data ownership clarity, and compliance with public-sector equity standards. Vendor systems should supplement, rather than substitute for, government-controlled analytic capabilities.

Finally, sustained investment in workforce capacity is essential to the success of any hybrid approach. Cross-agency data fusion and predictive intelligence require personnel with expertise in data science, statistics, program operations, privacy law, and ethics. Training and recruitment strategies should be aligned with long-term payment integrity goals, recognizing that technical systems alone cannot resolve institutional and governance challenges.

a hybrid model of cross-agency data fusion and predictive intelligence offers a balanced and pragmatic approach to reducing fraud in U.S. benefits programs. By combining centralized authority, federated analytics, and program-level expertise within a robust governance framework, such a model can enhance fraud detection while safeguarding privacy, equity, and public accountability. The recommendations outlined in this section provide a roadmap for translating comparative insights into actionable policy and administrative reforms.

7. Conclusion

Fraud and improper payments across U.S. benefits programs reflect long-standing structural challenges rooted in fragmented data systems, complex eligibility requirements, and limited cross-agency coordination. As this study demonstrates, cross-agency data fusion and predictive intelligence offer meaningful opportunities to improve fraud detection and prevention, but their effectiveness depends on how they are governed, implemented, and integrated into public administration processes. No single model centralized, federated, vendor-managed, or program-specific adequately balances analytic capability, legal authority, privacy protection, and accountability on its own. The comparative analysis supports the adoption of a hybrid approach that combines centralized authoritative checks, federated analytics for sensitive data, and program-level predictive modeling within a robust governance framework. Such an approach enables agencies to address both routine and complex fraud risks while preserving transparency, due process, and public trust. Ultimately, the success of cross-agency predictive systems will depend not only on technological innovation, but on sustained institutional commitment to oversight, equity, and responsible data use.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

- [1] Bannister, F., & Connolly, R. (2020). *The trouble with transparency: A critical review of openness in e-government*. *Policy & Internet*, 12(3), 343–367.
- [2] Button, M., & Brooks, G. (2016). *The changing nature of fraud: Fraudsters adapt*. *European Journal of Criminal Policy and Research*, 22(4), 491–505.
- [3] Burrell, J. (2016). *How the machine “thinks”: Understanding opacity in machine learning algorithms*. *Big Data & Society*, 3(1).
- [4] Cavoukian, A. (2011). *Privacy by design: The 7 foundational principles*. Information and Privacy Commissioner of Ontario.
- [5] Citron, D. K., & Pasquale, F. (2014). *The scored society: Due process for automated predictions*. *Washington Law Review*, 89, 1–33.
- [6] Congressional Research Service. (2023). *Preventing improper payments: Lessons from using data matching (IF12334)*. U.S. Congress.
- [7] Congressional Research Service. (2024). *The Payment Integrity Information Act of 2019: Overview and implementation issues (R48296)*. U.S. Congress.

- [8] Congressional Research Service. (2025). *Using data to reduce improper payments: Overview of the Do Not Pay working system* (IF12936). U.S. Congress.
- [9] Data Foundation. (2024). *Open by default: OMB's data management guidance and the Evidence Act*.
- [10] Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- [11] Federal Data Strategy. (2019). *Federal data strategy: A framework for leveraging data as a strategic asset*. Office of Management and Budget.
- [12] Government Accountability Office. (2019). *Improper payments: Opportunities remain to reduce government-wide improper payments*.
- [13] Government Accountability Office. (2020). *Data governance: Federal efforts could benefit from better coordination*.
- [14] Government Accountability Office. (2022). *Artificial intelligence: Emerging opportunities, challenges, and implications*.
- [15] Government Accountability Office. (2024). *Payment integrity: Significant improvements are needed to address improper payments and fraud*.
- [16] Government Accountability Office. (2025). *Fraud and improper payments: Data quality and a skilled workforce are essential*.
- [17] Government Accountability Office. (2025). *Fraud risk management: Federal efforts and remaining challenges*.
- [18] Hardy, S., Henecka, W., Ivey-Law, H., Nock, R., Patrini, G., Smith, G., & Thorne, B. (2017). *Private federated learning on vertically partitioned data via entity resolution and additive secret sharing*.
- [19] House Committee on Oversight and Accountability. (2021). *Algorithmic accountability and transparency in federal agencies*.
- [20] House Committee on Oversight and Reform. (2025). *Preventing fraud in government payment systems through bipartisan solutions*.
- [21] Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). *Benefits, adoption barriers and myths of open data and open government*. *Information Systems Management*, 29(4), 258–268.
- [22] Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). *Accountable algorithms*. *University of Pennsylvania Law Review*, 165, 633–705.
- [23] Kuner, C., Cate, F. H., Millard, C., & Svantesson, D. J. B. (2017). *Machine learning with personal data*. Oxford University Press.
- [24] Moynihan, D. P., Herd, P., & Harvey, H. (2015). *Administrative burden: Learning, psychological, and compliance costs in citizen-state interactions*. *Journal of Public Administration Research and Theory*, 25(1), 43–69.
- [25] Office of Management and Budget. (2019). *Memorandum M-19-18: Federal data strategy*.
- [26] Office of Management and Budget. (2023). *Evidence Act implementation guidance and learning agendas*.
- [27] Office of Management and Budget. (2024). *Payment integrity and improper payment reduction guidance*.
- [28] Office of Management and Budget. (2025). *Data governance and interoperability guidance under the Evidence Act*.
- [29] Pandemic Response Accountability Committee. (2023). *Data analytics and oversight of COVID-19 relief programs*.
- [30] Rieke, N., et al. (2020). *The future of digital health with federated learning*. *npj Digital Medicine*, 3(1).
- [31] U.S. Department of the Treasury. (2024). *Do Not Pay working system: Program overview and data sources*.
- [32] Utah Department of Workforce Services. (2025). *Utah eFind eligibility integration system*.
- [33] Veale, M., & Borgesius, F. Z. (2021). *Demystifying the draft EU Artificial Intelligence Act*. *Computer Law Review International*, 22(4), 97–112.
- [34] Wikipedia contributors. (2025). *Foundations for Evidence-Based Policymaking Act*.
- [35] Wikipedia contributors. (2025). *Mosaic effect*.
- [36] Wikipedia contributors. (2025). *Unemployment insurance in the United States*.